



Étude sur l'écologie et l'économie des praxéologies de la recherche d'information sur Internet. Une contribution à la didactique de l'enquête codisciplinaire

Caroline Ladage

► To cite this version:

Caroline Ladage. Étude sur l'écologie et l'économie des praxéologies de la recherche d'information sur Internet. Une contribution à la didactique de l'enquête codisciplinaire. Education. AIX-MARSEILLE UNIVERSITÉ, 2008. Français. NNT: . tel-01964613

HAL Id: tel-01964613

<https://hal.science/tel-01964613>

Submitted on 22 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITÉ AIX-MARSEILLE 1 – UNIVERSITÉ DE PROVENCE
AIX-MARSEILLE UNIVERSITÉ**

■■■■■■■■■■

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ D'AIX-MARSEILLE UNIVERSITÉ

Formation doctorale :

Sciences de l'éducation

présentée et soutenue publiquement

par

Caroline LADAGE

le 4 décembre 2008

TITRE :

ÉTUDE SUR L'ÉCOLOGIE ET L'ÉCONOMIE DES PRAXÉOLOGIES

DE LA RECHERCHE D'INFORMATION SUR INTERNET

Une contribution à la didactique de l'enquête codisciplinaire

Directeur de thèse :

Yves CHEVALLARD

JURY

Georges-Louis BARON, professeur, Université René Descartes Paris 5, *examineur*

Marianna BOSCH, *profesora titular*, Université Ramon Llull (Barcelone), *examinatrice*

Éric BRUILLARD, professeur, Université de Paris 12, *rapporteur*

Yves CHEVALLARD, professeur, Université de Provence Aix-Marseille 1, *examineur*

Jean RAVESTEIN, professeur, Université de Provence Aix-Marseille 1, *président*

Gérard SENSEVY, professeur, Université de Rennes 2, *rapporteur*

Laboratoire où la thèse a été préparée : UMR P3 ADEF (Aix-Marseille Université, INRP)

Résumé en français. Cette recherche se situe dans le cadre de la *théorie anthropologique du didactique* (TAD) dont elle utilise la notion clé de *praxéologie*. Le problème étudié est celui des conditions et contraintes déterminant *les équipements praxéologiques* institutionnels et personnels en matière de *recherche d'information sur Internet* (RII). L'étude se scinde en trois questions non indépendantes. Tout d'abord, quelles praxéologies peuvent être regardées comme *utiles* à la RII ? Ensuite, quelle est *l'offre praxéologique* en la matière ? Enfin, quels sont les obstacles à un accès large à une offre praxéologique idoine ? Ce travail, qui mobilise la notion de *culture* d'un domaine d'activité, développe les notions de *frontière praxéologique* et de *lecture questionnante* et met en œuvre une technique originale d'exploration d'un ensemble praxéologique. Ses résultats tendent vers la constitution, à l'intersection de la *didactique de l'enquête codisciplinaire* et de la *didactique du Web*, d'une *didactique de la RII*.

Mots clés : conditions et contraintes, culture (de la RII), didactique de l'enquête codisciplinaire, équipement praxéologique, frontière praxéologique, lecture questionnante, offre praxéologique, recherche d'information sur Internet, théorie anthropologique du didactique, Web.

Titre en anglais : *Study on the ecology and economy of the praxeologies of Web search. A contribution to the didactics of codisciplinary inquiry*

Résumé en anglais. The framework of this research is *the anthropological theory of the didactic* (ATD), of which it uses the key concept of *praxeology*. It studies the conditions and constraints determining institutional and personal Web search praxeological equipments. The study is divided into three closely related questions. Firstly, which praxeologies can be regarded as useful to Web search? Secondly, what is the offer of praxeologies in this field? Finally, what are the obstacles which limit access to a large supply of appropriate praxeologies? This work, which mobilizes the concept of a domain's *culture*, develops the notions of *praxeological border* and *question-led reading* and implements an original technique to explore a given set of praxeologies. Its results tend to the constitution, at the intersection of the *didactics of codisciplinary inquiry* and of *Web didactics*, of a *Web search didactics*.

Keywords: conditions and constraints, culture (of Web search), didactics of codisciplinary inquiry, praxeological equipment, praxeological border, question-led reading, praxeological offer, Web search, anthropological theory of the didactic, Web.

Remerciements

Mes remerciements vont d'abord...

... à *Yves Chevallard*, qui m'a permis de porter un regard nouveau sur une activité que je pensais pourtant familière, pour sa direction sans faille tout au long des différentes étapes de ce travail, pour ses compétences scientifiques, son érudition et sa curiosité sans frontières, pour sa patience, ses conseils scientifiques et linguistiques, pour les séances de travail d'une valeur inestimable, pour les pistes et les projets de recherche qu'il m'a suggérés,

... à *Jean Ravestein*, grâce à qui ce travail a pris son essor,

... à *Gérard Sensevy*, *Georges-Louis Baron*, *Éric Bruillard* et *Marianna Bosch*, qui ont accepté d'être les rapporteurs et les membres du jury de ce travail.

Je tiens aussi à remercier vivement...

... *Michèle Artaud*, *Joël Denisot*, *Michel Jullien*, *Jean-Pierre Lapébie*, *Christian Reymonet* et *Odile Schneider*, pour l'aide précieuse qu'ils m'ont apportée,

... ainsi que les enseignants, étudiants, élèves et toutes les personnes qui ont accepté de participer aux enquêtes, questionnaires, interviews, en particulier *Corinne*, *Perrine* et *Yann Gaël*.

Je remercie particulièrement *Sergine Saiz-Oliver* et ses élèves pour m'avoir permis d'observer leur activité en classe depuis de nombreuses années, ainsi qu'*Odile Chenevez* grâce à qui j'ai pu observer le travail de formation de professeurs stagiaires. Ces moments ont été très riches en éclairages.

Mes remerciements vont encore...

... à mes anciens collaborateurs professionnels du Web, *Bernard Fernandez* et *Stefan Polo*, qui m'ont maintes fois accordé leur temps pour soupeser mes questions parfois inattendues,

... aux participants au séminaire TAD/IDD animé par Yves Chevallard, auprès de qui les thèmes de cette recherche ont pu rencontrer un premier public,

... au directeur de l'UMR P3 ADEF, *Alain Mercier*, ainsi qu'aux chercheurs de ce laboratoire et à son personnel, en particulier *Annie Rombi*, pour m'avoir accueillie en m'offrant un lieu de travail, de présentation et de mise en débat du travail engagé.

Je tiens bien sûr à remercier chaleureusement...

... l'ensemble des enseignants chercheurs et du personnel du département des sciences de l'éducation et de l'université de Provence qui, d'une manière ou d'une autre, m'ont aidée et encouragée, tout particulièrement *Christian Roux, Yvan Abernot, Bernard Donnadiou, Jeanne Mallet, Chantal Eymard, Odile Thuilier, Stéphane Simonian, Jacques Audran, Patrick Lascabettes, Christine Vallier, François Michel, Danièle Picard, Odile Sylvain-Cifre, Leila Temli, Viviane Dho et Françoise Fabre,*

... sans oublier les apports d'autres universitaires, et plus particulièrement de *Valérie Campillo* et de *Juliette Simont*.

Enfin, je remercie mes enfants, *Laurent* et *Aurélie* (ainsi que leurs tribus), et mes parents, dont les récits de leurs quêtes sur le Web ont été pour moi des sources précieuses de réflexion et de questionnement.

Caroline Ladage

Table des matières

*It was a process of accretion, not the linear solving
of one well-defined problem after another.*

Tim Berners-Lee, 2000

Résumés (p. 3)

Remerciements (p. 5)

INTRODUCTION GÉNÉRALE. LE « MONDE INTERNET » ET LA THÉORIE ANTHROPOLOGIQUE DU DIDACTIQUE. ÉLÉMENTS POUR UNE RECHERCHE

Du référencement à la RII (p. 15)

La théorie anthropologique du didactique (p. 19)

Conditions et contraintes (p. 21)

Une contrainte sensible : la péjoration du monde Internet (p. 23)

Personnes, institutions, rapports (p. 28)

La notion de praxéologie (p. 32)

Didactique de l'enquête codisciplinaire (p. 40)

Problématiques de recherche connexes (p. 48)

Problématique de la recherche (p. 59)

Aspects méthodologiques (p. 65)

Note sur le « style » (p. 72)

ANNEXE 1. La péjoration de l'Internet à l'épreuve du différenciateur sémantique (p. 74)

PREMIÈRE PARTIE. ENTRER DANS LE MONDE INTERNET ? ANALYSE D'UNE ENQUÊTE CODISCIPLINAIRE

Introduction (p. 87)

Chapitre 1. L'enquêtrice, la question, les semblables, les savoirs (p. 90)

La question inaugurale : comment vivre (au) mieux avec Internet et Google ? (p. 91)

Le point de départ : une "vanity search" expéditive (p. 92)

Le trauma et le collègue invisible (p. 93)

Savoirs indigènes, savoirs allogènes (p. 97)

Non pas organiser, mais faire système (p. 100)

Pan et holon (p. 102)

Chapitre 2. Ratés de l'enquête et contre-enquête avec Internet (p. 105)

De "In God We Trust" à "Our mission" (p. 105)

L'expression "Our mission is" vue à travers Google (p. 107)

"Our mission" en français (p. 110)

Que disent les dictionnaires ? La notion de "mission statement" (p. 111)

Une autre recherche avec Google : "mission statement" (p. 113)

Des outils d'étude pertinents (p. 122)

Complément d'enquête : déclarer sa mission est-il obligatoire ? (123)

Chapitre 3. Égarements et approximations de l'enquête (p. 131)

Organiser, dites-vous ? (p. 131)

Élargir l'enquête... (p. 138)

"All the information in the world" ou les aléas d'une enquête (p. 140)

Chapitre 4. Le cas de PageRank (p. 148)

Google : une image ambivalente (p. 149)

PageRank, mathématiques et bibliométrie (p. 150)

PageRank contre la démocratie et la culture ? (p. 154)

Méprises et malentendus (p. 157)

Chapitre 5. Scilicet ou Qu'est-il permis de savoir ? (p. 162)

Savoirs et folklore de l'Internet (p. 163)

Politique et savoirs de l'Internet (p. 167)

De l'Internet au Web : invasion technique et déroute de la culture (p. 172)

Le recours aux savoirs « indigènes » : à propos de doxa (p. 176)

Traduction, trahison ? Une autre contre-enquête (p. 177)

Un moteur de recherche européen ? (p. 187)

Conclusion (p. 188)

ANNEXE 2. Les « aides à l'étude » : une somme de compétences et d'expériences
(p. 195)

DEUXIÈME PARTIE. QUE PEUT-ON SAVOIR ? EXPLORATIONS PRAXÉOLOGIQUES

Introduction (p. 213)

Chapitre 6. Ordre et désordre dans la connaissance (p. 218)

Des ordres tout faits à interroger (p. 218)

Un symptôme : Shmoogle (p. 219)

L'expertise supposée ne suffit pas (p. 223)

Chapitre 7. Deux analyses praxéologiques (p. 225)

La RII, un thème presque absent (p. 225)

Le Web et ses liens (p. 227)

Adresses Web et onglets (p. 232)

À la recherche des moteurs de recherche (p. 234)

“Basic Computer Skills” (p. 237)

La syntaxe des requêtes (p. 239)

Chapitre 8. Un peu au-delà des rudiments (p. 241)

Préparer une requête (p. 242)

Écrire une requête : les opérateurs booléens (p. 244)

La recherche par champ (p. 247)

Outils de la recherche : les annuaires (p. 248)

Les moteurs de recherche face aux annuaires (p. 252)

Méta-moteurs et « moteurs de moteurs » (p. 254)

Bases de données : errances et sérendipité ? (p. 255)

Manuel d’instruction ou programme d’études ? (p. 258)

Du français à l’anglais et retour (p. 263)

Conclusion (p. 268)

TROISIÈME PARTIE. DES SAVOIRS UTILES À LA RII ? UN REPÉRAGE PRAXÉOLOGIQUE

Introduction (p. 279)

Chapitre 9. Premiers repérages : histoire, économie, politique, documentation, droit (p. 283)

Domaines praxéologiques et RII (p. 284)

Personnes, institutions, dispositifs (p. 285)

L’irruption de l’histoire (p. 286)

L’enracinement économique (p. 289)

La politique ne se fait jamais oublier (p. 292)

Une vieille dame saisie par la modernité : la documentation (p. 293)

Le droit, outil et modèle (p. 298)

Chapitre 10. Mathématiques, physique et technologies numériques (p. 314)

<i>Des mathématiques dans la RII ?</i>	(p. 315)
<i>Les bases physiques du numérique</i>	(p. 317)
<i>Le cas de l'électricité</i>	(p. 318)
<i>L'induction électromagnétique</i>	(p. 320)
<i>À propos de disques durs</i>	(p. 323)
Chapitre 11. L'informatique	(p. 326)
<i>Théorie, pratique : une tension indépassable ?</i>	(p. 327)
<i>Matériel et logiciel</i>	(p. 329)
<i>Protocoles</i>	(p. 331)
<i>Difficulté de l'élémentation : l'exemple du BIOS</i>	(p. 333)
Chapitre 12. La fracture de l'anglais	(p. 340)
<i>Un faux problème ?</i>	(p. 340)
<i>Quand la langue s'opacifie...</i>	(p. 342)
<i>HTML : le préfixe hyper</i>	(p. 345)
<i>HTML : le markup</i>	(p. 348)
<i>Comment traduire "A (Web) search" ?</i>	(p. 352)
Chapitre 13. L'Internet	(p. 355)
<i>L'union du téléphone et de l'ordinateur : le cas de l'ADSL</i>	(p. 355)
<i>Modem, modulation, etc.</i>	(p. 358)
<i>Questions de débit : les liaisons physiques</i>	(p. 359)
<i>Des connaissances à diffuser ?</i>	(p. 360)
<i>Numéros IP</i>	(p. 362)
<i>Réseaux et protocoles</i>	(p. 364)
<i>Le modèle client/serveur et le DNS</i>	(p. 366)
<i>Le cas des noms de domaine</i>	(p. 370)
<i>Le modèle TCP-IP</i>	(p. 372)
<i>Le contrôle de la transmission des données</i>	(p. 375)
<i>Adresses IP et routage des données</i>	(p. 376)
<i>La couche liaison de données</i>	(p. 378)
<i>Une culture appropriée est une culture large</i>	(p. 379)
Chapitre 14. Le World Wide Web	(p. 382)
<i>Le Web et le modèle client-serveur</i>	(p. 383)
<i>Le Web en lui-même</i>	(p. 384)
<i>La genèse du Web : concepts et outils</i>	(p. 387)
<i>Un glossaire du Web</i>	(p. 389)

Les RFC et la notion de ressource (p. 392)
Une notion subtilement complexe : la notion d'URL (p. 393)
Hyperliens et HTML (p. 398)
Le cas du hotlinking (p. 402)
Les pages de cadres comme problème (p. 405)
Quels contenus ? Quels pièges ? (p. 412)
Les cookies et la liberté de l'internaute (p. 422)
Conclusion (p. 435)
ANNEXE 3. Sur la notion d'éléments d'une science (p. 440)

QUATRIÈME PARTIE. DES SAVOIRS SPÉCIFIQUES DE LA RII ? BESOINS ET OFFRES PRAXÉOLOGIQUES

Introduction (p. 457)
Chapitre 15. La recherche d'information sur Internet : un exposé encore (p. 464)
Un exposé tout simple (p. 464)
Annuaire, moteurs, métamoteurs (p. 465)
Google, toujours lui ! (p. 469)
Rechercher une information (p. 470)
Au-delà des apparences : ordre des résultats et référencement (p. 471)
Questions d'URL (p. 474)
Contrôler l'information : au-delà de l'URL (p. 479)
Médias en ligne : une autre frontière (p. 481)
Chapitre 16. Questions de RII (p. 484)
Quelles questions se poser ? (p. 485)
"How to search" (p. 485)
"Searching Text" (p. 491)
Un vaste panorama praxéologique (p. 500)
Points de vue de formateurs (p. 509)
Synonymie ou idionomie ? Le cas des booléens (p. 515)
Le moment socratique et son évitement (p. 518)
Questions de RII (p. 520)
Chapitre 17. Vers des réponses ? (p. 524)
Témoigner de l'offre praxéologique ? (p. 528)
Une enquête auprès de « témoins » (p. 529)

Offre et besoins praxéologiques : une rencontre délicate (p. 537)

Une offre praxéologique accessible ? (p. 539)

Construire des réponses R^\forall ? (p. 547)

Conclusion (p. 551)

ANNEXE 4. “*How to search*” : *un exposé en questions* (p. 555)

ANNEXE 5. *Des questions de RII extraites de Notess, 2006* (p. 584)

ANNEXE 6. *Présentation du questionnaire en ligne* (p. 589)

CONCLUSION GÉNÉRALE. VERS UNE DIDACTIQUE DU WEB. RÉSULTATS ET PERSPECTIVES

Praxéologies utiles à la RII (p. 598)

L’offre praxéologique en RII (p. 600)

Un triple problème épistémologique, didactique, éthique (p. 601)

L’accès à l’offre praxéologique en RII (p. 605)

Culture de la RII et codisciplinarité (p. 607)

Recherches à venir (p. 609)

Bibliographie (p. 613)

INTRODUCTION GÉNÉRALE

**LE « MONDE INTERNET »
ET LA THÉORIE ANTHROPOLOGIQUE DU DIDACTIQUE
ÉLÉMENTS POUR UNE RECHERCHE**

INTRODUCTION GÉNÉRALE

Le « monde Internet » et la théorie anthropologique du didactique Éléments pour une recherche

*Une œuvre où il y a des théories
est comme un objet sur lequel on laisse la marque du prix.*

Marcel Proust, 1927

Le travail que nous présentons est le produit de la rencontre entre deux dynamiques de développement. La première, qui donne à ce travail sa spécificité conceptuelle et technique, est celle de la *théorie anthropologique du didactique* (TAD), au sein de laquelle, à partir de l'année 2000 au moins, s'est approfondi le souci d'une *didactique de l'enquête codisciplinaire*. La seconde, qui fournit à notre recherche son objet, est celle du *World Wide Web* lui-même, dont l'impressionnante croissance depuis le milieu des années 1990 n'est évidemment pas sans effet sur la pratique et la théorie de l'enquête codisciplinaire telle que l'étudie la TAD : dès 2000, les savoirs « dialectiques » que dégage et commente Yves Chevallard à propos des TPE, les *travaux personnels encadrés*, introduits alors dans les classes de première de l'enseignement général, sont reliés explicitement à l'utilisation des richesses documentaires du Web, par contraste avec la (relative) rareté documentaire traditionnelle dans l'éducation scolaire (Chevallard, 2001a ; et aussi Chevallard & Matheron, 2002). Ce sont les tenants et les aboutissants de cette rencontre que, du point de vue de ce travail, nous nous efforcerons de préciser dans cette *Introduction générale*.

Du référencement à la RII

Nous dirons ici rapidement comment nous avons été conduite à nous situer, personnellement, dans la configuration objective évoquée dans ce qui précède. Après des études parallèles de philosophie (1989) et de sciences politiques (1990) à l'Université Libre de Bruxelles,

complétées ensuite par des études de gestion des entreprises à l'Institut d'Administration des Entreprises (IAE) de l'Université Paul-Cézanne (1995), nous avons eu depuis le début des années 1990 une activité professionnelle, dont l'importance s'est rapidement accrue, au sein de ce que nous appellerons dans tout ce travail le *monde Informatique & Internet*. Cette activité, soit comme consultant indépendant, soit dans le cadre de SARL à la création desquelles nous avons été associée, nous a amenée à acquérir des compétences approfondies en matière de création et de référencement de sites Web ainsi que de gestion informatisée de données, activités que nous avons développées pour des entreprises de tous secteurs. En même temps, dans le prolongement de ces travaux, et en raison de la labilité intrinsèque de leurs produits, nous avons eu à répondre à des demandes de formation à la création de sites et à leur référencement, en français et en anglais, émanant d'entreprises désireuses d'accroître les compétences de leurs dirigeants et de certains de leurs salariés. Ce fut là une expérience cruciale pour le cheminement qui a été le nôtre.

La labilité des produits est une chose ; mais les actions de formation que, à l'époque, nous avons eu à concevoir et à réaliser nous montraient chaque fois le caractère au moins partiellement labile du *contenu même* de ces formations. Une telle observation nous conduisit alors à nous interroger, au-delà de ce contenu spécifique, sur les savoirs existant en matière de formation, ce qui nous incita à nous tourner vers le département de sciences de l'éducation de l'université de Provence, avec le désir d'y acquérir une formation systématique à la formation d'adultes engagés dans la vie professionnelle et, plus précisément, aux prises avec le monde Informatique & Internet, lequel nous apparaissait alors comme la terre d'élection des « savoirs labiles ». Guidée par Jean Ravestein, nous avons d'abord suivi les enseignements du DESS « Ingénierie de la formation multimédia »¹. Dans ce cadre, notre mémoire, intitulé *Utilisation des nouvelles technologies dans un dispositif de formation d'un savoir labile* (Ladage, 2003), portait sur la conception d'une plate-forme hybride d'apprentissage du référencement (le public visé étant celui des clients des SARL où nous avons exercé notre activité professionnelle). L'année suivante, dans le cadre du DEA « Systèmes d'apprentissage, systèmes d'évaluation », nous avons consacré notre mémoire, intitulé *Approche didactique de l'enseignement d'un savoir labile en technologies de l'information et de la communication* (Ladage, 2004), à la question de l'apprentissage du référencement. Le jury de soutenance comprenait, outre Jean Ravestein, Samuel Johsua et Yves Chevallard – dont, sur les conseils de Jean Ravestein, nous avons commencé de suivre avec attention et intérêt le séminaire. Le « tournant didactique » était pris.

Le travail que nous présentons est le fruit de ce cheminement. Mais plusieurs éléments doivent encore être précisés. Dans des contextes institutionnels divers, nous avons, dès 1997,

été amenée à enseigner la recherche d'information sur Internet – que nous désignerons désormais, dans ce travail, par le sigle RII². Parallèlement, nous observions les comportements de recherche sur Internet, ainsi que les manières d'en parler, tant dans des situations de la vie courante ou dans des situations de formation à l'université que, de façon plus systématique, dans des situations de classe à l'école primaire, où il est traditionnel de demander aux élèves de « chercher des mots », mais où l'accès croissant au Web conduit aujourd'hui à leur demander aussi de rechercher de l'information pour répondre à des questions plus ou moins précises (Ladage, 2007a). Dans cette perspective, notre article avec Jean Ravestein et Samuel Johsua (2007) portait ainsi sur le bruit, le silence, les « incongruités » observables dans les résultats affichés en réponse à certaines requêtes adressées aux moteurs de recherche Google, Yahoo! et MSN. Dans le même temps, et de notre point de vue de référenceur, il devenait clair que nombre de savoirs qui nous auraient semblé utiles à la conduite des recherches évoquées ici paraissaient étrangement absents, quel que soit le type d'utilisateur observé. En particulier, il en était ainsi des savoirs du référencement, qui fournissaient de cette absence un exemple à plusieurs égards paradigmatiques (Ladage & Ravestein, 2005).

En français, le mot de référencement (appliqué à un site ou à une page Web) est en fait polysémique. Le sens restreint du mot désigne simplement le fait qu'un site Web est indexé par tel moteur de recherche, c'est-à-dire est « présent » dans sa base de données – y ait « référencé ». Mais le fait qu'il soit ainsi « connu » du moteur de recherche ne signifie pas pour autant qu'il occupera une position avantageuse (sur la première page de résultats...) dans les réponses apportées par le moteur aux requêtes contenant certains mots clés déterminés. Et cela n'est pas tout : comme le souligne un auteur reconnu en la matière (Olivier Andrieu, 2008, p. 3), il ne servirait à rien « d'être bien positionné sur des mots clés que personne ne saisit ou sur des moteurs que personne n'utilise... ». En anglais, on parle plus globalement de search engine optimization (SEO), c'est-à-dire d'optimisation (d'un site ou d'une page) pour les moteurs de recherche, les référenceurs étant désignés comme search engine optimizers (l'acronyme SEO ayant ainsi, en anglais, deux significations). Depuis le début de la jeune histoire du référencement, il existe en quelque sorte un « jeu » – au sens de la théorie des jeux (game theory) – entre référenceurs et (gestionnaires de) moteurs de recherche. Classiquement, par exemple, ces derniers sont portés à « punir » (en les bannissant de leurs index) les sites qui tentent de les « tromper » quant à leur contenu réel (celui qui sera véritablement accessible à l'utilisateur) ou quant à leur popularité (définie en termes

de liens pointant vers le site). À l'inverse, on reconnaîtra à certains sites ou à certaines pages la vertu d'être "search engine friendly".

L'idée s'imposa à nous, ainsi, de l'utilité de travailler à une « didactique de l'Internet » ou « du Web » (Ravestein & Ladage, 2006). La pratique du référencement conduit en vérité à entrer dans un jeu à *trois* joueurs, car référenceurs et moteurs de recherche doivent les uns comme les autres tenir compte, dans leurs choix stratégiques, d'un troisième joueur : *l'utilisateur*. Inversement, on pouvait se demander – et c'est là l'origine de notre travail – si, au lieu de se représenter son action comme si elle relevait d'un jeu à *un seul joueur* ou, comme l'on dit encore de façon très suggestive, d'un jeu « contre la Nature » – *a game against Nature* –, le praticien ordinaire de la RII ne gagnerait pas sensiblement à assumer sa démarche comme participant d'un jeu à (au moins) *trois* joueurs, incluant, outre lui-même, *Webmasters* et référenceurs d'un côté, moteurs de recherche de l'autre – ce qui, bien entendu, supposerait de sa part la reconnaissance et une certaine connaissance de quelques-uns des savoirs portant sur les moteurs de recherche et sur le référencement.

Notre expérience du référencement de sites a, parallèlement, joué un rôle clé dans les réflexions didactiques qui ont accompagné notre questionnement. Deux éléments peuvent à cet égard être notés. Le premier paraîtra sans doute trivial et nous ne le citons ici que pour le mettre en relation avec un fait massif que nous expliciterons plus loin : il s'agit du temps consacré et de l'effort développé, au fil de plusieurs années d'un intense labeur fait d'allers-retours incessants entre pratique et théorie, à notre propre formation au monde Informatique & Internet, afin de nous donner les compétences professionnellement indispensables à notre activité de création et de référencement de sites. Le second élément, en vérité solidaire du premier, a constitué pour nous une expérience didactique essentielle. Loin d'être limitée à une « auto-formation » menée dans un face-à-face solitaire avec une sèche littérature spécialisée toujours plus volumineuse, notre formation à la création de sites, aux moteurs de recherche et au référencement s'est en effet appuyée de façon essentielle sur le travail collectif qui se menait alors dans divers forums en ligne dédiés aux questions que pouvaient se poser à l'époque les référenceurs, acteurs d'un champ de spécialisation encore à l'état naissant : de cette expérience subjectivement marquante, on trouvera une description et une analyse objectivées dans notre article sur *Les communautés virtuelles sur Internet* (Ladage, 2007b ; voir aussi Ladage, 2007c). Par contraste, nous pouvions observer autour de nous, hors du champ professionnel où nous opérons, une attitude souvent fort éloignée du climat d'engagement à la fois passionné et « hyperréaliste » qui prévalait sur la scène du référencement. D'un côté, l'utilisateur ordinaire, éternel débutant, semblait ne guère vouloir

s'instruire en matière de RII, qu'il se satisfasse de son niveau de pratique, qu'il se cache ses propres échecs ou qu'il les impute, plus communément, à une nature malveillante ou « mal faite » – celle « de l'Internet ». D'un autre côté, et de façon consonante, les « échanges de savoir » à propos de la RII, quand ils n'étaient pas entièrement absents, semblaient ne pouvoir s'établir que de façon allusive, hâtive, jamais approfondie, comme s'il y avait là un champ de problèmes qu'on ne pouvait pas reconnaître, sans parler de l'explorer méthodiquement. Pour mieux comprendre ces énigmes ordinaires, nous nous sommes alors tournée vers la TAD, qui nous paraissait à même, par son abord large, non sélectif, des faits didactiques, de nous permettre d'analyser ce que nous avons pu observer jusque-là.

La théorie anthropologique du didactique

La TAD repose sur un petit nombre de principes que Yves Chevallard a résumés à l'occasion du premier congrès international sur la théorie anthropologique du didactique tenu à Baeza (Espagne) en 2005. Le premier principe situe la didactique dans son *genre prochain*, celui des sciences en général : comme toute science (y compris les sciences « dures » : mathématiques, physique, etc.), la didactique « se définit en se donnant pour objet d'étude un certain ensemble *de conditions et de contraintes de la vie des sociétés humaines* » (Chevallard, 2005, p. 718). Le deuxième principe désigne la *différence spécifique* qui donne à la didactique son caractère distinctif parmi les autres sciences : la didactique, précise le même texte (p. 719), « se voue à étudier les conditions et contraintes sous lesquelles les praxéologies se mettent à vivre, à migrer, à changer, à opérer, à dépérir, à disparaître, à renaître, etc., au sein des institutions humaines ». La notion de *praxéologie* utilisée ici marque l'évolution entre les formulations de la *théorie de la transposition didactique* (Chevallard, 1991) et son extension ultérieure sous le nom de théorie anthropologique du didactique : le concept de praxéologie a en effet été engendré par le besoin de modéliser « la connaissance » en s'affranchissant des limitations imposées de façon plus ou moins subreptice par les notions jusque-là seules disponibles dans la culture courante, celle de *savoir* ou de *savoir-faire* notamment. Si composer un concerto suppose « du savoir », se moucher n'en supposerait pas, selon la culture dominante. Or là où celle-ci ne décèle aucun savoir digne de ce nom, la TAD voit des praxéologies qui, par leur nature, ne sont pas ontologiquement – sinon axiologiquement ! – différentes de celles mises en œuvre dans la composition musicale. Les « cotes » culturelles, quel qu'en soit l'objet et la nature, deviennent ainsi – le cas échéant – des réalités à étudier, à expliquer comme effet de certaines conditions et contraintes, et cessent d'être des passagers clandestins embarqués dans le travail scientifique.

Il y a là, croyons-nous, une raison clé du choix de l'adjectif *anthropologique* pour marquer l'apport propre de la TAD : au lieu de reprendre étroitement, de façon implicite ou parfois explicite (et même revendiquée), les points de vue et les valeurs de telle ou telle institution établie, la didactique – comme toute science, là encore – doit s'efforcer de bâtir ses praxéologies en fonction de ses besoins, sans s'embarrasser du « praxéologiquement correct » défini par telle ou telle institution existante, fût-elle au cœur de son domaine d'étude. Ainsi n'y a-t-il plus intrinsèquement de savoirs « dignes de ce nom » et de savoirs « indignes », de savoirs « nobles » et de savoirs sans noblesse, voire ignobles. Il y a des praxéologies, qui naissent, migrent, réussissent ou déperissent, sont déconstruites et reconstruites, s'installent durablement en tel paysage institutionnel ou disparaissent sans retour, etc. Cette puissance d'émancipation nous permettait en particulier de parler des praxéologies de la RII comme on parle en TAD des praxéologies de l'algèbre ou de la statistique, c'est-à-dire sans les dénigrer ou les louer, mais en les envisageant comme réalité à observer, à analyser et à suivre en ses migrations et transformations institutionnelles.

Cela nous permettait en particulier de reformuler et donc de penser autrement les constats auxquels nous étions arrivés jusque-là. Tout d'abord, pourquoi la culture praxéologique en matière de RII observable autour de nous semblait-elle tellement figée en un état rudimentaire, ce que confirmaient au reste plusieurs enquêtes internationales, dont l'une³ révélait par exemple, entre autres constats, la non-utilisation presque totale de l'opérateur booléen AND NOT ? Ensuite, la rareté et la pauvreté des « échanges de savoirs » en matière de RII pouvaient maintenant s'exprimer à l'aide de ce que la TAD nomme *le* didactique, en le définissant ainsi (Chevallard, 2005, p. 717).

... il y a du didactique lorsqu'une personne ou une institution, se faisant par là *agent didactique*, fait quelque chose *afin* qu'une institution ou une personne « apprenne » un certain ensemble praxéologique, c'est-à-dire afin que cet ensemble arrive jusqu'à cette institution ou cette personne et finisse par s'intégrer à son équipement praxéologique.

Là encore, nous disposions ainsi d'une notion très compréhensive, qui permettait de pointer tout à la fois l'apparente faible densité institutionnelle (ou interpersonnelle) du didactique à propos de RII et le caractère sommaire, non élaboré de ses formes – comme s'il n'y avait quasiment rien à apprendre et si ce peu-là pouvait être acquis à simplement l'évoquer à la volée. La question, bien entendu, était de comprendre *pourquoi* il en était ainsi, c'est-à-dire d'analyser les conditions et les contraintes qui maintenaient ainsi le didactique en matière de RII dans cet état de sous-développement apparent, au double plan quantitatif et qualitatif.

Pour cela, il convient maintenant d'entrer plus avant dans ce qui est au cœur de la définition de la didactique proposée plus haut, celle de *science des conditions et des contraintes de la diffusion des praxéologies dans la société*.

Conditions et contraintes

On peut tenter de résumer cette définition de la didactique par une autre, plus lapidaire, et formellement plus restrictive (Chevallard, 2005, p. 717) : « *La didactique est la science qui étudie le didactique.* » C'est supposer que la diffusion et la non-diffusion praxéologiques s'expliquent essentiellement par le didactique, ou encore qu'il n'y a pas d'apprentissage sans quelque intention didactique – ce qui est un postulat limite. Mais il n'en est pas moins vrai, nous semble-t-il, que la plupart des apprentissages sont en quelque sorte soutenus, impulsés, sinon tout entiers engendrés, par du didactique (y compris bien sûr de l'auto-didactique), faute de quoi ils demeurent en quelque sorte en suspens. Ce qui importe alors, c'est d'étudier ce que, en TAD, on nomme classiquement *l'écologie et l'économie* du didactique, notions étroitement liées à celles de conditions et de contraintes⁴. La distinction entre ces deux notions est simple dans son principe. *L'écologie* d'un objet peut être définie comme l'ensemble des conditions actuelles pesant sur la vie de cet objet. Les contraintes sont alors des conditions d'un type particulier : ce sont ces conditions qui, regardées depuis une certaine position institutionnelle au sein de la société, apparaissent (à un moment donné) comme *non modifiables* et sous lesquelles on devra alors, depuis cette position, organiser *l'économie* de l'objet, c'est-à-dire le remaniement ou la création des conditions tenues pour modifiables. Les situations didactiques – au sens de la théorie des situations didactiques (Brousseau, 1998) – qu'un enseignant peut envisager de créer dans une classe participent pour lui de *l'économie* de ces objets (mathématiques, grammaticaux, informatiques, etc.) regardés comme enjeux didactiques. En revanche, ce même enseignant regardera généralement comme une contrainte intouchable le fait d'assurer aux élèves des séances de travail d'une durée imposée, au lieu par exemple de « scander » le travail collectif – à la façon dont le psychanalyste Jacques Lacan conduisait ses séances – en découpant le temps d'horloge disponible en sessions pour les unes très courtes (d'une à deux minutes par exemple), pour d'autres beaucoup plus longues (alignant les heures au cours d'une même journée), le critère supposé en tout cela étant celui des conditions temporelles jugées optimales du point de vue didactique, c'est-à-dire du point de vue de la gestion des apprentissages.

Les distinctions entre condition (modifiable) et contrainte (non modifiable), d'une part, et économie et écologie, d'autre part, ne se superposent qu'imparfaitement. C'est ainsi

que l'innovation pédagogique consiste souvent à regarder certaines contraintes traditionnelles comme de simples conditions que, sans changer fondamentalement de position institutionnelle, on pourrait modifier à sa guise – pourvu toutefois que d'autres conditions et contraintes réputées favorables à la modification recherchée soient vérifiées. La volonté d'innover ou, plus souvent encore, le simple désir de croire à ce que l'on est contraint de faire conduisent en règle générale à *oublier*, voire à *nier*, un grand nombre de contraintes qui n'en déterminent pas moins l'*écologie* du didactique et ruinent parfois les projets les plus séduisants : l'utopie didactique rêvée se change alors en dystopie vécue ⁵. Si nous adoptons pour position institutionnelle celle du chercheur, l'oubli des contraintes comme, aussi bien, celui des conditions (modifiables) constituent une erreur lourde, contre laquelle la TAD tente de nous prémunir en proposant une échelle à plusieurs niveaux qui rappelle que les « contraintes » avec lesquelles il faut compter ne sont pas seulement les conditions « presque » modifiables que l'on renonce à modifier. Nous ne retiendrons ici que l'échelle partielle suivante, qui appelle un bref commentaire.

Civilisation
↓↑
Société
↓↑
École
↓↑
Pédagogie
↓↑
Discipline

L'échelon inférieur est, ici, celui de la « discipline », terme qui renvoie simplement à l'ensemble praxéologique dont on examine la diffusion. (Il peut évidemment s'agir de praxéologies de nature quelconque : mathématique, ou poétique, ou concernant la courtoisie dans la vie quotidienne, ou propres à la RII, ou relevant de la plomberie, etc. On retrouve là le caractère ouvert de la TAD.) Sur ce niveau pèsent des contraintes et opèrent des conditions ayant leur « siège » dans les niveaux supérieurs. Inversement, ce niveau impose des contraintes et crée des conditions propres à la discipline, qui « jouent » sur les niveaux supérieurs. À chaque utilisation, plus généralement, le schéma précédent doit être bien sûr spécifié : le niveau de l'école impose ainsi des contraintes et crée des conditions qui varient selon le type d'école considéré ; notons seulement, à titre d'illustration, qu'il peut s'agir d'une

école créée spécialement pour y enseigner la discipline considérée ou d'une école déjà existante, dans laquelle on vient loger un tel enseignement. Les mêmes remarques valent pour le niveau de la pédagogie : comme au niveau de l'école, et par contraste avec les niveaux supérieurs, il s'y déploie des conditions et des contraintes censées servir – d'une certaine manière – le projet de diffuser aux « élèves » de l'école la discipline praxéologique considérée. Même si les conditions et les contraintes issues des niveaux de l'école et de la pédagogie doivent être considérées, du point de vue de l'analyse et de l'ingénierie didactiques, comme des conditions et contraintes *comme les autres* (en dépit de leur finalité affichée), celles qui règnent au niveau de la société ou de la civilisation ne procèdent pas, sauf exception, d'une volonté de faciliter spécifiquement la diffusion de tel ou tel système de praxéologies. Pour le créateur d'un enseignement, ainsi, elles apparaîtront souvent comme des contraintes dont on peut tirer profit en certains cas, et dont on peut, en d'autres, atténuer les effets jugés nuisibles à la diffusion praxéologique visée par un aménagement local de conditions appropriées. Il y a là bien sûr le point de départ de l'illusion que le local – celui de l'enseignant s'affairant à préparer son « cours » par exemple – peut ignorer le global, alors que, bien souvent, le global pénètre avec les élèves eux-mêmes dans la salle de classe, dont « l'isolation didactique » n'est ainsi jamais assurée.

Ajoutons encore que le niveau de la *civilisation* a été introduit dans l'échelle des conditions et contraintes pour attirer l'attention sur le fait que la désignation d'une condition (modifiable ou non) comme relevant d'une *société* déterminée, qui est souvent un fait d'auto-désignation, masque par son localisme (voire son hyperlocalisme) l'extension du phénomène constaté à une zone parfois *beaucoup plus vaste* : la TAD retrouve ici le constat que faisait déjà Marcel Mauss (1873-1950) en proposant de réserver le nom de « phénomènes de civilisation » à ces phénomènes « communs à un nombre plus ou moins grand de sociétés et à un passé plus ou moins long de ces sociétés » (Mauss, 1930, p. 85). Notons comme probable que, en règle générale, une contrainte « de civilisation » tout à la fois passe davantage inaperçue (elle paraît « naturalisée », elle est vécue comme un fait « de nature ») et, toutes choses égales par ailleurs, est d'autant plus prégnante qu'elle est ainsi plus subreptice.

Une contrainte sensible : la péjoration du monde Internet

Une contrainte (ou une condition) peut favoriser la survenue d'un phénomène ou, au contraire, en gêner l'apparition. Dans les deux cas, on dira que cette contrainte est *pertinente* ou *sensible* (par rapport au phénomène), favorable dans le premier cas, défavorable dans le second. (Dans le cas contraire, on dira que la contrainte est *insensible* ou *non pertinente*.) À

titre d'illustration, nous nous arrêterons d'abord un bref instant sur une contrainte qui, sans y être centrale, n'est sans doute pas entièrement dénuée de pertinence en matière de RII. Voici la chose : on peut imaginer deux types de sociétés (ou de civilisations) où prévalent des attitudes opposées à l'endroit des produits de la robotique ; dans les premières, l'idée même de robot évoquera de sourdes menaces pour l'humanité et l'on y raillera avec un soulagement certain les déficiences éventuelles et l'inachèvement parfois évident de ces créations jugées inquiétantes ; dans les secondes, tout au contraire, on supposera que le robot est regardé comme un *ami de l'homme*, que sa mise au point est encouragée et que ses imperfections, supposées provisoires, sont regardées avec bienveillance. L'opposition est classique : la société américaine relève, nous dit-on, du premier cas (comme la société française, d'ailleurs), tandis que la société japonaise participe largement du second paradigme (Hornyak, 2006a). Ainsi l'écologie du rapport ordinaire aux robots ne sera-t-elle pas la même dans l'un et l'autre types de sociétés : difficile dans les premières, elle se révélera beaucoup plus facile dans les secondes. Se former à la robotique, par exemple, sera, selon le cas, un choix à assumer *contre* le regard social ordinaire ou, tout au contraire, un choix porté par le regard majoritaire ; et de même pour l'utilisation de robots dans les différents types de tâches qu'ils sont ou seront capables d'accomplir ⁶.

Dans l'abord de la RII qui est le nôtre ici, on ne peut manquer de rencontrer – nous aurons mainte occasion de nous en assurer au fil de ce travail – une contrainte analogue, celle matérialisée dans une attitude très répandue de suspicion et de péjoration à l'endroit de l'Internet et de ce qui y vit. En voici un exemple typique : dans un petit livre limpide et tonique, intitulé *L'Esprit des Lumières*, l'auteur, lui-même esprit éclairé, à la fin d'un développement questionnant le recours aux experts et à la science dans la détermination du bien, et avant une protestation de sens inverse contre l'emprise de la morale sur la recherche de la vérité, écrit tout à coup ceci (Todorov, 2006, p. 84).

La croissance vertigineuse des moyens de stockage et de transmission de l'information a révélé un danger nouveau : trop d'information tue l'information. Il suffit de poser une question sur Internet pour recevoir aussitôt cent mille réponses ; comment savoir laquelle est la plus digne de confiance et la plus éclairante ? Une encyclopédie composée librement par ses utilisateurs (la « Wikipédia ») est-elle préférable à celle rédigée par des scientifiques compétents ? Seulement si l'on efface la limite entre vouloir et savoir.

La thématique est classique : nous y reviendrons. Tout se passe comme s'il y avait là une réaction très large à ce qui est vécu par beaucoup comme une agression culturelle contre le monde d'hier. Plusieurs clarifications sont, à ce propos, utiles.

Notons d'abord que, contrairement à l'hypothèse évoquée à propos de l'exemple précédent, une telle réaction, qui semble chez beaucoup de personnes procéder d'une culpabilité imaginaire devant l'effacement anticipé du monde d'hier, peut être en même temps le fait de personnes fortement utilisatrices du Web, voire de spécialistes (ou d'ex-spécialistes) du monde Internet. Ainsi en va-t-il d'Andrew Keen, l'auteur d'un ouvrage à succès paru en 2007 sous le titre *The cult of the amateur*, dont le sous-titre est parfaitement explicite : *How today's internet is killing our culture*. Plus récemment, dans le numéro de juillet-août du magazine *The Atlantic*, l'auteur d'un ouvrage récent intitulé *The Big Switch: Rewiring the World, from Edison to Google* (Carr, 2008a), Nicholas Carr, a publié un article (Carr, 2008b) dont le titre parle de lui-même : *Is Google Making Us Stupid?* Cet article a suscité aux États-Unis un large débat dont on trouvera des échos dans un post d'Andrew Keen daté du 27 juillet, au titre catastrophiste (Keen, 2008) : *Is the Internet killing the American reader?* Mais ce climat d'appréhension n'est pas réservé au Nouveau Monde. Notons en outre qu'il semble baigner l'ensemble de la société, y compris les générations montantes, comme en témoigne cet autre post, qui semble faire écho jusque dans son titre – *Google nous rend-il cons ?* – à l'article de N. Carr, dû à un jeune homme de 19 ans qui conclut par ces mots (Williamson, 2008) : « Je suis dingue d'Internet, mais plus j'y réfléchis, plus Internet me fait peur. »

La péjoration de l'Internet, on le voit, peut accompagner une fréquentation assidue, voire passionnée du Web : l'objet péjoré, même devenu familier, même vécu comme indispensable, n'en reste pas moins un « mauvais objet ». Nous reviendrons un peu plus loin sur cette attitude dépréciative, voire hostile, face à l'Internet. Mais pour illustrer encore un peu le « jeu » des contraintes de niveau supérieur – c'est-à-dire qui ne tiennent pas *d'abord* à l'école ou à sa pédagogie, même si, bien sûr, elles sont sensibles à la discipline à diffuser –, nous nous arrêterons un instant sur une dissonance plus localisée, que nous retrouverons, elle aussi, tout au long de notre travail : celle entre le monde ancien et vénérable des *bibliothèques* et de la *documentation*, d'une part, et le monde Internet, d'autre part. La création du World Wide Web ne s'est pas faite dans un univers vierge en matière documentaire. Or le créateur du Web lui-même, Tim Berners-Lee, l'avoue, fût-ce implicitement, dans le passage suivant de son livre *Weaving the Web* : le mouvement de pensée qui a donné vie au Web porte en lui un affrontement presque inévitable avec le monde ancien de la gestion des « ressources » (Berners-Lee, 2000, p. 2).

The vision I have for the Web is about anything being potentially connected with anything. It is a vision that provides us with new freedom, and allows us to grow faster than we ever could when we were fettered by the hierarchical classification systems into which we bound ourselves. It leaves the entirety of our previous ways of working as just one tool among many.

Les auteurs d'un ouvrage intitulé *Web Dragons. Inside the Myths of Search Engine Technology*, qui nous accompagnera durant ce travail, consacrent leur deuxième chapitre – qu'ils intitulent "Literature and the Web" – à examiner le tableau historique sur le fond duquel naît le Web. Ils y insistent dans les termes suivants sur le choc culturel (et professionnel) qu'ont pu ressentir les milieux de la documentation et des bibliothèques devant l'irruption brutale de moyens d'information procédant, on l'a vu, d'une problématique radicalement nouvelle (Witten, Gori, & Numerico, 2007, p. 30).

How would you react if a technology juggernaut like Google suddenly declared that its mission was something that you thought society had entrusted to you alone, which you had been doing well for centuries: "to organize the world's information and make it universally accessible and useful"? How do librarians feel when people turn from their doors and instead visit web bookstores to check their references, to ferret out related books, or to see what others recommend—even to check out facts in standard reference works? Like everyone else, they are acutely aware that the world is changing, and they know their job descriptions and institutions must follow suit. Their clients now work online, and part of the contemporary librarian's job is to create websites, provide web interfaces to the catalogue, link to full text where available, and tend the e-mail help desk.

La situation est sans doute d'autant plus difficile à vivre que la doctrine de la profession ainsi remise en cause était depuis longtemps fortement affirmée, ce qu'un *outsider* – que nous avouons être en grande partie à cet égard⁷ – peut facilement méconnaître. Les mêmes auteurs signalent à cet égard une profession de foi aux accents imposants, celle de Richard Rogers Bowker (1848-1933), qui contribua fortement à la création en 1876 de l'*American Library Association* (ALA) et créa le *Library Journal* qu'il devait diriger pendant plus de cinquante ans. Dans une communication intitulée emphatiquement "The work of the nineteenth-century librarian for the librarian of the twentieth", présentée en 1883 à l'occasion du sixième congrès de l'ALA, Bowker écrit notamment ceci (Bowker, 1883, p. 85).

While the scientist is translating the book of nature and rewriting its palimpsests, it is for him [the librarian] to classify and catalogue the records of ascertained knowledge, the literature of the whole

past, and so to bring the books to readers and readers to books. He is the merchant, the middle-man, of thought, and performs in this field the function which political economy recognizes as so important, of bringing goods to the place where they are wanted, and so, also, creating demand. In this busy generation, when the hurried man grumbles that “all the time there is” is not enough for him, the librarian makes time for his fellow-mortals by saving it; for a minute saved is a minute added. And this function of organizing, of indexing, of time-saving and of thought-saving, is associated peculiarly with the librarian of the nineteenth century.

C’est à cette mission si hautement assumée que, en apparence au moins, le projet de Tim Bernes-Lee venait se heurter. Il n’y a pas si longtemps encore, un responsable d’une bibliothèque universitaire américaine (Campbell, 2003) pouvait suggérer “that the Association of Research Libraries should seriously pursue the feasibility of developing a ‘library.org’ web presence”, entreprise chimérique visant à promouvoir “the development of and provide access to the highest quality content on the Web”, ce qui se concrétiserait par la création d’un *scholars portal* à propos duquel l’auteur concluait : “the need within the academic community for a *scholars portal* is high, and the capacity to develop it is resident in ARL [Association of Research Libraries], OCLC [Online Computer Library Center], the Library of Congress, and other agencies.”

Deux raisons solidaires justifient à nos yeux que nous ayons choisi de mettre ainsi en exergue les contraintes exprimant une attitude toujours actuelle de défiance et de péjoration à l’encontre de l’Internet. La première, c’est tout simplement qu’il s’agit là d’une contrainte qui pèse sans doute lourdement sur l’élaboration et la diffusion sociales de praxéologies de la RII : qu’on minimise son investissement didactique dans ce que, somme toute, on tient en piètre estime – le Web et ses usages – ne paraît pas, en effet, déraisonnable ; qu’on justifie la chose, à ses propres yeux, en avançant que la RII, quant à elle, ne pose guère de problème particulier dont la résolution devrait être le fait de l’usager ne le semble pas davantage. Ainsi retrouve-t-on les constats mis au point de départ de notre travail. Bien entendu, ce travail montrera que bien d’autres contraintes existent, qui vont au delà d’une « simple » *attitude*. Mais c’est précisément l’effort engagé pour mettre au jour de telles contraintes qui nous apporte une seconde raison de souligner d’emblée ces pesantes contraintes : bien que n’ayant pas toujours manifesté le tranquille détachement qui nous aurait permis, à cet égard, une totale lucidité (ce qui n’est guère original dans la situation qui était la nôtre), nous avons cru constater tout au long de notre recherche combien celle-ci, en règle générale, hors du cercle nécessairement étroit des observateurs ou accompagnateurs fortement impliqués, donnait lieu à des malentendus, selon un mouvement logique que nous ne croyons pas caricaturer en le

schématisant ainsi : l'Internet étant un objet à l'évidence pour le moins « terne », sinon tout à fait inintéressant, sans attractivité « scientifique », cette recherche ne devait certainement d'exister qu'à se centrer sur quelques aspects piquants, sur lequel on attendait alors mille révélations. Sans digresser plus que nécessaire sur ce point, soulignons qu'il s'agit là d'une difficulté de notre recherche qui tient, non à son objet lui-même, mais au *rapport* à cet objet apparemment le plus répandu, encore aujourd'hui, ce qui rend délicate l'observance pourtant indispensable du précepte fameux, formulé par Spinoza (1632-1677) dans l'introduction de son *Traité politique* (1677), selon lequel il nous faut « ne pas railler, ne pas déplorer ni maudire, mais comprendre » – *non ridere, non lugere, neque detestari, sed intelligere*.

Personnes, institutions, rapports

La manière dont les conditions et contraintes évoquées jusqu'ici s'imposent dans un milieu humain doit encore être précisée. La TAD s'est construite, nous semble-t-il, par le développement, à partir du socle initial de la théorie de la transposition didactique, de plusieurs théories « sectorielles » successives et articulées entre elles. C'est à partir de l'élaboration de la théorie des *personnes*, des *institutions* et des *rapports* personnels et institutionnels (Chevallard, 1989) que le qualificatif *anthropologique* a, semble-t-il, commencé d'être usité (Chevallard, 1992). Comme précédemment, nous n'en retiendrons ici que le minimum qui nous a paru indispensable pour notre travail, tout en nous appuyant sur un certain nombre de textes fondamentaux (par exemple Chevallard, 1995, 2003).

Le point de départ se trouve ici dans une *dialectique* unissant deux termes, ceux de personne et d'institution : les personnes et les institutions se définissent les unes par les autres. Le mot d'institution est pris en TAD dans une acception des plus larges : est une institution tout ce qui peut être... institué. Une école, une classe, une séance en classe même sont des institutions (qui, en l'espèce, forment système). Une famille est une institution. Le déjeuner dominical est une institution – pourvu qu'il ait été institué, c'est-à-dire, nous le verrons un peu plus loin, qu'il soit marqué par certains rapports à peu près invariants à des objets emblématiques. La notion ainsi assumée par la TAD intègre même les emplois familiers du mot, dont le *Dictionnaire culturel en langue française* (Rey et al., 2005) propose l'exemple suivant : « La fraude est dans ce pays une véritable institution ! »

Les institutions sont des créations humaines ; elles ont des *acteurs*, qui en sont – plus ou moins – les *sujets* ; qui leur sont ainsi *assujettis* – plus ou moins strictement. L'assujettissement d'une personne à une institution peut survivre (en se transformant plus ou moins) au-delà de la période où l'institution regarde cette personne comme son sujet – ce

qu'exprime par exemple le passage du statut d'élève à celui d'*ancien* élève. Une personne apparaît de façon générale comme la résultante d'un système évolutif d'assujettissements institutionnels : l'*infans*, défini comme l'enfant qui n'a pas encore acquis le langage (Sándor Ferenczi), est ainsi d'emblée assujetti au langage, et à telle langue plutôt qu'à telle autre ; il est assujetti aussi à la pesanteur, ou plutôt à la manière dont son groupe humain a institué le traitement des effets de la pesanteur, lequel varie avec l'âge, etc. Et on pourrait faire des remarques analogues pour l'ensemble des contraintes « naturelles » – physiques, biologiques, etc. –, dont le traitement donne lieu, en un groupe humain donné, à des créations institutionnelles spécifiques. Ainsi les institutions « font »-elles les personnes, qui les font et les feront, ou les déferont avant de les refaire.

Selon une formule employée par Y. Chevallard, une personne est, à chaque instant, l'émergent d'un *nexus* d'assujettissements passés et présents⁸. Une personne est donc tout à la fois le produit – sans producteur – des institutions existantes, y compris ces institutions par lesquelles une société définit la capacité des personnes à changer le monde, entre « libre arbitre » et « destin » –, et une innovation en puissance – qui, toutefois, ne trouvera pas toujours à « s'exprimer » –, dans la mesure où le faisceau de ses assujettissements apparaît unique ou, du moins, *singulier*. Une personne est ainsi une singularité de l'espace institutionnel. Deux points sont à cet égard à souligner. Le premier point tient en ceci que, contrairement à ce que certaines des formulations employées jusqu'ici pourraient porter à croire, une personne dispose bien de *liberté*, mais d'une liberté potentielle, qu'il lui revient chaque fois de faire advenir : prenant appui sur certains de ses assujettissements, elle peut vouloir s'assujettir à de nouvelles institutions en espérant alléger, par cela même, des assujettissements devenus « tyranniques ». Ainsi en va-t-il de qui entreprend une psychanalyse, par exemple ; ou de qui change de partenaire de vie, ou de métier ; ou encore du chercheur qui bâtit une nouvelle théorie à laquelle il s'assujettira pour se délivrer de ses points de vue anciens, tenaces mais jugés par lui, désormais, contre-productifs ; ou, aussi bien, de qui emprunte pour se libérer d'une dette⁹ ; etc. Il y a ainsi, dans le « jeu » renouvelé des assujettissements d'une personne – qui correspond, on le voit, à un *je* renouvelé –, une dialectique entre *puissance* (ce sont nos assujettissements qui nous donnent notre puissance d'action) et *asservissement* (ce sont nos assujettissements qui limitent le champ de notre action).

Le second des deux points annoncés sera plus encore au cœur de notre travail. Toute personne cache, derrière sa singularité visible, une foule d'assujettissements, dont certains apparaissent définitivement incorporés à la personne, et dont d'autres semblent – peut-être à tort – n'être plus que des *traces* et des *vestiges* d'assujettissements anciens. De ce « moirage »

institutionnel de la personne, pourtant, il est en principe possible de faire ressortir tel ou tel assujettissement (ou tel « paquet » d'assujettissements) en créant momentanément, localement, des conditions auxquelles ces contraintes du comportement de la personne se révéleront *sensibles* – tandis que d'autres y seront insensibles ou moins sensibles –, et cela afin d'amener les assujettissements visés à s'exprimer dans le comportement de la personne. Nous verrons plus loin comment cette possibilité théorique sera exploitée dans la méthodologie de notre recherche. Mais il faut ajouter avant cela une autre notion théorique au corpus de la TAD : celle de *rapport*. On retrouve là, à nouveau, le phénomène d'approfondissement théorique déjà souligné.

Par delà les institutions et les personnes, la TAD suppose des *objets* (et des rapports à ces objets). Un objet peut être une institution ou une personne ; plus généralement, est objet tout ce qui *existe* pour au moins une institution ou une personne. Un objet existe pour une personne ou une institution si celle-ci a un certain rapport à cet objet – ou plutôt : a un rapport « non vide » à cet objet. On dit alors que la personne ou l'institution *connaît* l'objet – selon le rapport qui est le sien à cet objet. Un rapport (personnel ou institutionnel) subsume *a priori* tout ce qui entre dans le « commerce » que peut avoir la personne ou l'institution avec l'objet en question. À nouveau, ce qu'est le rapport à un objet peut varier depuis le cas limite du rapport vide – mais ce rapport n'est plus vide dès lors que quelqu'un a prononcé devant vous son nom, par exemple... – jusqu'à un rapport riche, complexe, intime, qui ne sera peut-être celui que de quelques personnes ou de très rares institutions. Le rapport d'une personne ou d'une institution à un objet peut changer : on dira alors que la connaissance qu'a cette personne ou cette institution change. Ainsi peut-on parler d'apprentissages personnels ou institutionnels.

Pour décrire le mécanisme de l'assujettissement d'une personne à une institution, il faut affiner un peu la description précédente. Une institution donnée accueille ses « sujets » dans des *positions* différenciées : on peut être sujet de l'école en tant qu'élève ou en tant que professeur, par exemple. Étant donné une position institutionnelle, l'institution définit et promeut un rapport institutionnel (à un objet donné) *pour les sujets de l'institution occupant cette position*. Le rapport d'un professeur enseignant en collège à l'objet « brevet blanc », par exemple, n'est ainsi pas le même que celui de l'élève au même objet. Mais ce dernier exemple peut être trompeur : le « rapport d'un professeur » à l'objet « brevet blanc » mentionné ici est en fait le « rapport institutionnel *pour un professeur* » à l'objet « brevet blanc ». Ce n'est jamais vraiment le rapport *personnel* de tel ou tel professeur à ce même objet. C'est le rapport en quelque sorte idéal proposé (en principe, bien que de façon parfois peu lisible) par l'institution comme « modèle » du rapport personnel qui devrait être celui de ses sujets venant

occuper la position : ce n'est en fait le rapport de personne, parce que toute personne est pluri-assujettie et que son rapport à un objet ne saurait être engendré totalement (même s'il l'est principalement) par son assujettissement à *une* institution, qui plus est dans *une* position au sein de cette institution. Pour que la personne occupant telle position apparaisse alors comme un « bon sujet » devant l'institution, il suffira que son rapport personnel soit regardé comme *conforme* au rapport institutionnel pour la position qu'il occupe – ce qui suppose une *évaluation* par l'institution, question dans laquelle nous n'entrerons pas plus ici. Bien entendu, telle personne pourra alors apparaître comme un « mauvais sujet » de l'institution, parce que ses assujettissements antagonistes à d'autres institutions se laisseront trop bien apercevoir depuis l'institution où, par exemple, elle tente de se faire agréer comme sujet « à part entière ». À cet égard, on notera que chacun oscille, dans la mise en scène de ses assujettissements (vécus souvent comme constitutifs de sa « personnalité ») entre histrionisme et retenue ; dans tous les cas, la « clinique » des personnes doit être attentive à l'idée d'explorer ce qui peut se tenir derrière ce que la personne exhibe ou cèle.

La théorie des rapports ainsi développée a de nombreuses implications, en particulier au plan de la méthodologie, nous le verrons. Notons ici seulement deux phénomènes qui concernent au premier chef les communautés scientifiques. Une institution ne définit de rapport (non vide), pour les positions existant en son sein, qu'à propos d'un certain ensemble d'objets, qui d'ailleurs peut varier avec la position (les personnes dans telle position n'ayant pas à connaître, *en tant que sujets de l'institution*, tel objet que d'autres sujets de l'institution, en une position différente en son sein, devront connaître selon tel rapport institutionnel). Lorsque, pour une position donnée, existe un rapport institutionnel non vide à un certain objet, on pourra dire que cet objet est visible depuis la position considérée au sein de l'institution. On dira que l'objet n'est pas visible depuis l'institution s'il n'est visible depuis aucune des positions existant en son sein. (Il revient au même de dire que l'objet n'existe pas pour l'institution ou que l'institution ne connaît pas cet objet.) Si, en outre, une personne n'a jamais été assujettie qu'à des positions institutionnelles d'où tel objet est (ou était) invisible, cet objet n'existera pas pour lui. Cela noté, lorsque l'institution considérée est une communauté scientifique, les recherches accomplies en son sein ont en principe pour effet ou d'y rendre visibles des objets qui ne l'étaient pas, et donc de proposer des rapports institutionnels là où il n'y avait rien – ou, du moins, *rien de problématique* –, ou de porter sur ces objets un autre regard, c'est-à-dire de proposer un remodelage du rapport institutionnel. Quand elle ne manque pas sa cible, une recherche porte en elle, ainsi, l'exigence d'apprentissages nouveaux de la part de la communauté où elle s'inscrit et, en conséquence, de la part de ses « sujets », chercheurs et responsables de la communauté. C'est ainsi que, comme on le verra, la

réception du travail dont nous exposons ici les préliminaires, qui porte déjà sur un objet – l’Internet – généralement péjoré, supposera en outre (quoique pas seulement) que l’on parvienne à suspendre un instant le rapport ordinaire, si invétéré qu’il semble parfois naturel, à la distribution sociale des connaissances, en tant que structurée selon diverses « spécialisations » qui reprennent à l’infini l’antique opposition clerc/laïc : hommes/femmes (ce savoir est pour les hommes, pas pour les femmes, *ou l’inverse*), adultes/enfants, jeunes/vieux, élites/masses, professionnels/amateurs, spécialistes/non-spécialistes, informaticiens/non-informaticiens, mathématiciens/non-mathématiciens, historiens/non-historiens, juristes/non-juristes, etc. Pour cela, une autre extension théorique sera, nous semble-t-il d’un grand secours : nous lui consacrerons la prochaine section.

La notion de praxéologie

La TAD n’aurait sans doute pas l’extension qu’elle a prise si la notion de *praxéologie* et les notions associées de *type de tâches*, de *technique*, de *technologie*, de *théorie*, qui en sont constitutives, n’avaient pas été introduites. Selon Y. Chevallard (communication personnelle, 24 mars 2006), le point de départ de cette élaboration se trouve dans la question suivante : de quoi naissent tout à la fois les objets et les rapports aux objets ? De cette naissance on trouvera une illustration – que nous ne reprendrons pas ici – dans la thèse de doctorat de Marianna Bosch (1994, pp. 15-21). À l’origine il y a l’activité humaine, dont l’unité élémentaire est l’accomplissement d’une *tâche*. Ce fait originaire suppose, bien entendu, des objets « donnés », par rapport auxquels la tâche sera définie : il n’est pas question ici de genèse *ex nihilo*. Notons que l’objet qui, pour la personne, doit avant tout exister – ou venir à l’existence, au cours d’un processus d’apprentissage – est le *type* de tâches, type dont chaque tâche particulière est un *spécimen* particulier. Dans la pratique majoritaire des sociétés, les *types* de tâches sont indiquées par des verbes d’action : « Mouche-toi ! » est une injonction indiquant à son destinataire d’accomplir ici et maintenant la tâche particulière du type « se moucher ». Tel est le point de départ de l’analyse de l’action que propose la TAD. Bien entendu, la référence anthropologique fait écho, ici comme ailleurs, à une exigence épistémologique précise : il n’est pas de types de tâches qui tomberaient hors du champ d’investigation de l’analyse praxéologique. Il n’est pas, ainsi, de type de tâches que l’on s’autoriserait, dans le cadre de la TAD, à regarder *a priori* comme futile, ou répugnant, ou sacré (et donc intouchable ou inaccessible), ou trop complexe, ou simplement inintéressant.

Dans les praxéologies dont la structure est la plus simple, le type de tâches, *T*, constitue le premier composant. Le deuxième composant est la *technique*, désignée par la

lettre grecque τ (*thêta*), initiale du mot grec *tekhnê*, savoir-faire. Pour accomplir une tâche d'un certain type T (se moucher, composer une fugue, nouer un hameçon au bout d'un fil à pêche, préparer un biberon, monter des œufs en neige, résoudre une équation du troisième degré, etc.), il faut mettre en œuvre une certaine manière de faire – une certaine technique. L'apport de l'anthropologie didactique est ici plus net encore qu'auparavant : il faut une technique pour se moucher (il en est plusieurs...), mais il en faut une aussi pour *marcher*, comme Mauss (1936/1950b) l'avait bien vu, ou pour rire, pour parler, pour respirer, etc. Les actes supposés « naturels » sont ainsi, selon la formule maussienne, des « idiosyncrasies sociales » : on ne marche pas selon la même technique – à l'époque de Mauss – si l'on est une femme française ou une femme américaine, une petite fille maori ou un soldat britannique qui défile. Une technique est clairement une construction institutionnelle, qu'il faut apprendre. Le couple formé d'un type de tâches et d'une technique, notée usuellement $[T/\tau]$ (où la barre oblique est un simple séparateur), constitue ce qu'on nomme le *bloc pratico-technique*, ou *praxis* : en langage familier, on a là essentiellement ce que le français nomme un *savoir-faire*. Bien entendu, rien n'indique que ce savoir-faire soit, si l'on peut dire, de « bonne qualité » : le type de tâches peut être mal défini ; la technique peut n'avoir qu'une portée – c'est-à-dire une zone d'efficacité – ridiculement réduite ; ou encore son usage peut provoquer régulièrement de multiples erreurs ; etc.

La volonté ou l'obligation d'accomplir des tâches situées hors de portée de la technique disponible sont le ressort qui, sous certaines conditions et contraintes, provoquera l'évolution adaptative de la technique utilisée (il y a en ce cas simple « travail » de la technique), ou son remplacement par une autre technique existante (il y a alors diffusion et transposition d'une technique allogène), ou encore la mise au point pure et simple d'une technique nouvelle (il y a ainsi création technique). On passe ici de la *statique* praxéologique à la *dynamique* praxéologique, qui est au cœur de la didactique et des problèmes de l'éducation. Notons à ce propos un phénomène quasi universel : si l'on est conduit à effectuer une tâche d'un type T , c'est en général parce que, ayant à effectuer une tâche d'un type T^* , on le fait à l'aide d'une technique τ^* qui oblige, par définition même de τ^* , à effectuer à un certain moment une tâche du type T . Ainsi, lorsqu'on effectue une multiplication (T^*) par la technique τ^* usuelle, on est conduit à devoir mémoriser (« retenir ») des retenues (T) à l'occasion des produits partiels : pour effectuer le produit 23×17 , on aura ainsi, à un certain moment, à « dire » quelque chose comme « 7 fois 3, 21 ; je pose 1 et je retiens 2 », en mémorisant la retenue 2 avant de passer au produit partiel suivant. Mais il est possible quelquefois de changer de technique pour accomplir les tâches du type T^* en adoptant une technique τ^{**} qui, elle, ne conduise pas à devoir accomplir des tâches du type T : s'agissant

par exemple de la multiplication, on pourrait ainsi recourir à la très ancienne technique *per gelosia*, qui possède la propriété recherchée (Chevallard, 1991, pp. 78-79). Un tel *évitement* d'un type de tâches *problématique* par le moyen d'un changement de technique « en amont » est un mécanisme fréquent dans les dynamiques praxéologiques. Mais le simple fait – qu'illustre l'exemple de la technique de multiplication *per gelosia* – qu'un tel mécanisme soit quelquefois obstinément refusé suggère que notre description de l'objet même des dynamiques praxéologiques est encore incomplète.

Une telle description doit en fait prendre en compte un troisième composant praxéologique. Toute technique τ appelle en effet un « discours » qui la *justifie* (dans l'ambition qui lui est prêtée, par une personne ou une institution, de permettre l'accomplissement de tâches du type *T*), qui l'explique, l'éclaire, lui donne une certaine *intelligibilité* – et qui, dans le meilleur des cas, permet même de *produire* cette technique. C'est un tel discours, qui peut prendre des formes diverses (incluant des formalismes, des calculs, des preuves expérimentales, cliniques, etc.), que l'on appelle en TAD une *technologie*, « discours raisonné » (*logos*) sur le savoir-faire (*technê*) qu'on notera en général par la lettre grecque θ (*tau*). Bien entendu, une technologie l'est *pour* une personne ou une institution : c'est d'abord une *fonction* – la fonction technologique (de justification, d'explication, etc.) – que tel discours concret pourra assurer auprès de telle personne ou institution, qui l'agréera, mais sera impuissant à assurer auprès de telle autre, qui le rejettera. On retrouve en cela la fragmentation du monde social déjà évoquée à propos des dichotomies du type clerc/laïc. Par contraste, nous rencontrerons plus loin le problème de la création de technologies (en matière de RII et, plus largement, de monde Informatique & Internet) qui soient « robustes », c'est-à-dire capables de créer *pour tous* une intelligibilité variable sans doute mais toujours satisfaisante.

En quelques cas, la technologie d'une technique se trouve en quelque sorte incorporée à la technique elle-même : décrire la technique, ou, comme l'on dit, « l'expliquer » est tenu souvent pour une justification satisfaisante, comme il en va dans les « recettes » (de cuisine, de statistique inférentielle, etc.), où il suffit parfois de « réciter » la technique pour la faire accréditer. Dans d'autres cas, la technologie peut apparaître comme un discours « opportuniste », engendré en fonction des besoins de l'instant. D'une façon générale, l'écologie praxéologique est soumise à des lois plus embrouillées qu'on ne pourrait le croire. Une technique peut apparaître trop simple pour s'installer durablement en telle institution : c'est le cas de la multiplication *per gelosia*, trop rudimentaire déjà au XV^e siècle pour être reçue dans les arithmétiques du commerce. Bien qu'objectivement simple, toutefois, une technique peut ne pas être « comprise » – et peut même heurter la « logique » – de ceux à qui

on la destinait. Pour obtenir le quotient entier d'un nombre entier (par exemple 124) par, disons, 15 ($= 3 \times 5$), on peut déterminer d'abord le quotient de ce nombre par 3 (ici, il est le même que le quotient de 123 par 3, soit 41), puis le quotient du quotient trouvé (ici, 41, donc) par 5 (ici, il est le même que le quotient de 40 par 5, soit 8) : on obtient alors le quotient cherché (ici, 8). On peut démontrer que cette technique (non enseignée au secondaire aujourd'hui en France) est valide dans le cas général (avec un diviseur et un dividende entiers positifs quelconques) ; mais qui le croira ? En un tel cas, une technologie acceptable semble trop complexe pour être reçue de ceux à qui elle serait destinée ; de là que nombre d'institutions – tel le collège d'avant-guerre pour la technique de division ci-dessus, qui y était effectivement enseignée – recourent alors à un acte d'autorité institutionnelle, qui permet seul de la faire accréditer par ses « sujets ».

La question de la technologie est ainsi d'une grande complexité. Mais, selon la TAD, un point doit être regardé comme assuré : du point de vue des besoins « anthropologiques » (et nullement d'un point de vue normatif), même si sa nature ou sa forme peuvent varier beaucoup, la présence éventuellement silencieuse d'une technologie *est inévitable*. Si la technique répond à la question « Comment ? » (accomplir des tâches du type considéré), la technologie répond à la question « Pourquoi ? » (cette technique marche-t-elle). Même si la « structure » d'une technologie peut être très diverse, la *fonction* technologique est, elle, nécessairement prise en charge : car la vie d'une technique comporte toujours des épisodes – qu'on peut regarder comme critiques, même s'ils sont banals – où quelqu'un « demande pourquoi » (la technique marcherait). Cette observation, au reste, s'applique aussi bien à la technologie elle-même : celle-ci ne saurait être entièrement auto-justifiée ; elle contient donc certaines assertions – ou certains silences autour de certaines assertions – qui, pour les personnes et les institutions destinataires du discours technologique, semblent mériter une justification supplémentaire. C'est là le rôle qu'assume en principe le quatrième composant d'une praxéologie : la *théorie*, qu'il est d'usage de désigner par la lettre grecque Θ (« grand » thêta). Bien entendu, la notion de théorie évoquée ici doit être entendue comme relative à l'institution dans laquelle un tel discours assume une *fonction* théorique. Les énoncés théoriques se distinguent généralement, dans les sciences comme dans l'ordinaire de la vie des personnes et des institutions, par le fait d'apparaître plus « généraux », plus « abstraits », susceptibles d'engendrer ou du moins de justifier et d'éclairer – éventuellement de façon opportuniste – un assez grand nombre d'énoncés technologiques. En même temps, ils apparaissent plus évanescents, évoqués de façon allusive plutôt qu'explicités, et donc moins facilement mis en débat.

Ce qu'on appelle praxéologie, dans le cas le plus simple, c'est, au total, une organisation, qu'on peut noter $[T/\tau/\theta/\Theta]$, unissant un type de tâches T , une technique τ , une technologie θ , enfin une théorie Θ . Le bloc $[T/\tau]$, on l'a dit, est celui de la *praxis*. Le bloc $[\theta/\Theta]$ est le bloc du *logos*, ou bloc *technologico-théorique*. Cette union des deux blocs se reflète dans le mot de praxéologie lui-même, dans lequel on peut continuer à lire, si on le souhaite, et selon l'usage de la formation des mots savants, la référence à une « science » (-logie) de la « pratique » (praxéo-) : mais il s'agit alors de la « science » (concrétisée par le bloc $[\theta/\Theta]$), *pour telle personne ou institution*, de telle de ses « pratiques » (celle désignée par le bloc $[T/\tau]$)¹⁰. Un tel fragment de « science » n'est en général pas isolé. Une praxéologie du type $[T/\tau/\theta/\Theta]$ est dite *ponctuelle*, parce que constituée autour de ce « point » qu'est le type de tâches T . Mais une technologie est souvent commune à plusieurs techniques. À moins qu'elle ne soit une technologie *ad hoc*, toute technologie a en général une *portée* qui ne se limite pas à une technique : elle peut engendrer, justifier, rendre intelligible (plus ou moins) *plusieurs* techniques, par des spécifications adéquates de son noyau assertif. On obtient alors des praxéologies dites *locales*, dont la « formule » pourrait être celle-ci : $[T_i/\tau_i/\theta/\Theta]_{i \in I}$, où I est un ensemble fini d'indices. La chose est plus vraie encore s'agissant des théories : on obtient ainsi des praxéologies dites *régionales*, où une même théorie Θ justifie *plusieurs* technologies θ_j ($j \in J$), chacune de ces technologies θ_j justifiant elle-même *plusieurs* techniques, τ_{ji} ($i \in I_j$), ce qu'exprime la formule suivante : $[T_{ji}/\tau_{ji}/\theta_j/\Theta]_{j \in J, i \in I_j}$.

Ajoutons qu'un bloc $[\theta/\Theta]$ (ou, plus largement, $[\theta_j/\Theta]_{j \in J}$) correspond à ce que la langue française d'aujourd'hui nomme un *savoir*. En pratique, il nous arrivera cependant, désignant le tout par le nom d'une de ses parties, de parler de *savoirs* pour désigner des praxéologies quelconques. Soulignons à cet égard que, parmi les raisons de créer la notion de praxéologie, on retrouve la volonté d'émanciper la TAD, en tant qu'outil de production de connaissances, des points de vue « indigènes » des personnes et institutions étudiées, et en particulier de se donner les moyens de prendre en compte les « petits » savoirs, ou tenus pour tels, et même ces savoirs – tels, encore fréquemment, les savoirs du monde Informatique & Internet – dont quelques-uns nieront absolument qu'ils soient en aucune façon des savoirs ! À cet égard, pour la TAD, il n'y a *a priori* que des praxéologies – et donc, dans un langage plus relâché, il n'y a que des savoirs, même quand on s'intéresse surtout aux savoir-faire qu'ils enveloppent. S'il y a des différences *pertinentes* (à certains égards) entre deux familles de praxéologies, il conviendra chaque fois de l'établir avec les outils de la TAD, sans se prévaloir des verdicts de la *vox populi*. Parce que c'est à propos des mathématiques que la TAD s'est d'abord construite, nous donnerons d'abord ici un (unique) exemple de praxéologie *mathématique*¹¹.

Le type de tâches T et la technique sont ceux vus plus haut : une tâche du type T consiste à trouver le quotient q d'un entier a par un entier b supposé non nul, strictement inférieur à a , et s'écrivant comme un produit $b = b_1 b_2$ de diviseurs « simples » : dans l'exemple déjà vu, on avait $a = 124$, $b = 15 = 3 \times 5$ et donc $b_1 = 3$ et $b_2 = 5$. On a vu la technique, τ : elle consiste à déterminer (« de tête ») le quotient q_1 de a par b_1 , puis à déterminer le quotient q_2 de q_1 par b_2 ; on a alors $q = q_2$. Comment justifier cette manière de faire ? Pour cela, il faut revenir à la définition du quotient. Quand on considère la suite des valeurs $b, 2b, 3b, \dots, kb$, etc., il arrive un moment où l'on dépasse a : il existe donc un entier k tel que $kb \leq a < (k+1)b$; par définition, le quotient q est égal au nombre k . Par exemple les multiples de 15 forment la suite suivante : $15, 2 \times 15 = 30, 3 \times 15 = 45, 4 \times 15 = 60, 5 \times 15 = 75, 6 \times 15 = 90, 7 \times 15 = 105, 8 \times 15 = 120, 9 \times 15 = 135 \dots$ Celle-ci montre que, dans ce cas, on a $q = 8$. Pourtant, dans ce que le lecteur vient de lire, il y a comme un silence : est-on sûr que, quels que soient les nombres a et b , il arrivera toujours un moment où la suite des multiples de b dépassera l'entier a ? Il s'agit là d'une hypothèse théorique qui va de soi pour tous, sauf pour les mathématiciens, qui n'ignorent pas que ce phénomène (le caractère dit « archimédien » de l'ensemble ordonné des entiers) est particulier. Mais admettons donc cette hypothèse. De l'encadrement $qb \leq a < (q+1)b$, qui en résulte, nous retiendrons d'abord que l'on a $qb \leq a$. Considérons alors l'autre « partie », l'inégalité $a < (q+1)b$: celle-ci équivaut à l'inégalité $a+1 \leq (q+1)b$, car, quand un entier n est strictement inférieur à un entier m ($n < m$), l'entier suivant, $n+1$, est soit inférieur strictement à m ($n+1 < m$), soit égal à m , et réciproquement. On a ainsi établi un énoncé technologique clé : q est le quotient de a par b si et seulement si on a d'une part $qb \leq a$ et, d'autre part, $a+1 \leq (q+1)b$. Considérons maintenant le quotient q_1 de a par b_1 : on a $q_1 b_1 \leq a$ et $a+1 \leq (q_1+1)b_1$. Soit ensuite le quotient q_2 de q_1 par b_2 ; le rôle de a dans ce qui précède étant joué par q_1 , on a de même : $q_2 b_2 \leq q_1$ et $q_1+1 \leq (q_2+1)b_2$. Cela étant, comme $q_2 b_2 \leq q_1$, on a $(q_2 b_2) b_1 \leq q_1 b_1$, soit $q_2 b \leq q_1 b_1$, et donc, puisque $q_1 b_1 \leq a$, il vient $q_2 b \leq a$. Par ailleurs, comme on a $a+1 \leq (q_1+1)b_1$ et $q_1+1 \leq (q_2+1)b_2$, il vient $a+1 \leq (q_2+1)b_2 b_1$ soit $a+1 \leq (q_2+1)b$. Ainsi le nombre q_2 vérifie-t-il les deux inégalités qui caractérisent le quotient q : $q_2 b \leq a$ et $a+1 \leq (q_2+1)b$. On en déduit donc, comme annoncé, que $q = q_2$.

Pourquoi la question des praxéologies importe-t-elle ? Lorsqu'une personne devient le sujet d'une institution, son assujettissement à cette institution peut maintenant se traduire ainsi : pour tous les types de tâches qu'elle aura à accomplir au nom de l'institution, elle est

conduite à intégrer à ce que nous nommerons son *équipement praxéologique* (personnel) les praxéologies prescrites par l'institution (dans la position qu'elle y occupera). Le *rapport institutionnel* à un objet donné pour une position donnée est alors l'amalgame de l'ensemble des façons d'activer cet objet à travers l'ensemble des praxéologies institutionnelles où cet objet apparaît (et qui sont mises en œuvre dans la position institutionnelle considérée). Une formulation analogue s'applique à propos du rapport *personnel* d'une personne à cet objet : il est le fruit de l'amalgamation de l'ensemble des façons d'activer cet objet à travers l'ensemble des praxéologies personnelles où cet objet apparaît. Si la personne est assujettie à l'institution en une certaine position, son rapport personnel, on l'a dit, devra se conformer au rapport institutionnel qui lui correspond. C'est là qu'un problème de *gestion praxéologique* de la société doit être soulevé. Ainsi qu'on l'a vu, les rapports personnels aux objets dérivent des systèmes praxéologiques en vigueur dans les institutions – de leurs équipements praxéologiques. Or ces rapports personnels vont constituer, dans les dynamiques praxéologiques ultérieures, autant de conditions et de contraintes portées par les personnes, de même que les rapports institutionnels constitueront des paquets de conditions et de contraintes portées par les institutions. Or ces conditions et contraintes vont entraver plus souvent qu'ils ne vont favoriser les innovations praxéologiques, pour cette raison simple qu'ils sont plus ou moins adéquats aux équipements praxéologiques établis. La racine du problème se trouve alors, à travers les rapports personnels et institutionnels, dans les équipements praxéologiques correspondants. Cette situation exige du chercheur qu'il fasse une *analyse praxéologique* attentive, exigence à laquelle répond, dans l'ordre de l'action, en toute institution, le devoir de chacun de manifester un *souci praxéologique* constant.

Nous prendrons encore un exemple, qui nous rapprochera de notre objet – la RII –, auquel les rapports ordinairement observables paraissent si souvent porter témoignage d'une réelle incurie praxéologique. Dans un ouvrage paru en février 2008 intitulé *Georges Canguilhem*, l'auteur, Dominique Lecourt, mentionne la dernière conférence que donne, le 10 mars 1990, Georges Canguilhem (1904-1995). Une note de bas de page précise ceci (Lecourt, 2008, p. 104).

Cette conférence publiée par la revue *Commentaire* (14.53) au printemps 1991 est disponible sur plusieurs sites internet en tapant son titre complet sur n'importe quel moteur de recherche.

La conférence de Canguilhem a pour titre *Qu'est-ce qu'un philosophe en France aujourd'hui ?* L'indication donnée par D. Lecourt tranche, par son style familier et sans apprêt, sur le reste de l'ouvrage, sévère et d'une haute tenue : on est là bien loin des

recommandations soignées du *Publication Manual* de l'American Psychological Association (APA, 2001). Le texte de Canguilhem, d'ailleurs, n'apparaît pas dans la liste bibliographique placée en fin d'ouvrage (p. 125), alors même que huit des dix-neuf titres qu'elle rassemble sont signés de Canguilhem lui-même. Il y aurait donc là comme une « exception bibliographique », peut-être liée à l'idée que ce qui est “retrievable from the Internet” ne doit pas être mélangé avec les ouvrages publiés qui n'existent que sous forme imprimée – de fait, la bibliographie ne comporte aucune référence à des documents en ligne. Ne prêtons pas à D. Lecourt des sentiments qu'il n'a pas besoin d'éprouver lui-même pour que son lecteur en soit éventuellement affecté. Ici tout conspire à suggérer que la technique à mettre en œuvre est des plus simples : choisir *n'importe quel* moteur de recherche, taper le titre de la conférence, mais, il est vrai, le titre « complet » ; et voilà. Le « n'importe quel moteur de recherche » est, bien entendu, une manière de parler – l'auteur n'a sans doute pas vérifié son affirmation sur plus de quelques moteurs, et il y a de toute façon neuf chances sur dix environ que, aujourd'hui, son lecteur choisisse Google ¹². La technologie de la technique indiquée, ici, est absente ou implicite : simplement la manœuvre *marche* – elle marche tellement bien, nous laisse-t-on entendre, qu'elle marche quel que soit le moteur de recherche ! Le lecteur est ramené alors à l'idée – dominante – d'un jeu *contre la nature*, dans lequel la « nature » – le moteur de recherche consulté – répond soit positivement (le joueur a alors gagné), soit négativement (le joueur a perdu, la partie est terminée, il n'y a rien d'autre à faire). Qu'en est-il au juste ? Consulté quelques mois après la rédaction de l'ouvrage ¹³, le moteur de recherche Yahoo! annonce environ 3 120 000 résultats ; Google, de son côté, en annonce 602 000. L'apparent foisonnement des réponses laisse voir, en première page, un ou plusieurs liens conduisant au texte recherché ¹⁴. La manœuvre réalisée est brutale, mais elle réussit (sur ces moteurs). Si elle a l'avantage de suggérer à l'internaute timoré qu'on peut « taper » dans la boîte de recherche bien autre chose qu'un ou deux mots « clés », détail technique qu'il pourra intégrer à son équipement praxéologique (mais sans en connaître l'explication), elle n'évoque pas moins une pratique de la RII où la réflexion n'a guère de part – *il suffit de faire*.

On aura noté que l'auteur n'invite pas l'utilisateur à mettre le titre *entre guillemets* (pour requérir des documents contenant l'expression exacte du titre). Si on le fait, le nombre de résultats annoncés n'est plus alors que de 42 pour Yahoo! et de 21 pour Google : l'apparent bric-à-brac fait de centaines de milliers de « résultats » est d'un coup renvoyé au néant. Notons que la recherche aurait pu cependant être moins heureuse : Live search annonce 136 000 résultats, mais l'article cherché n'apparaît pas dans les 10 premières pages de résultats ; si l'on met des guillemets, ce nombre de résultats tombe à 22, et l'on voit alors apparaître aisément ce qu'on cherchait. Il n'est pas difficile, cependant, de trouver des moteurs de

recherche qui semblent n'avoir pas trace, dans leur index, du document cherché : si Exalead, Ask.com, Excite ou Clusty ne sont pas dans ce cas, on ne peut en dire autant de Lycos, Hotbot ou Mirago par exemple. Comme l'illustre sinon la pratique de l'auteur, du moins la manière dont il s'adresse à ses lecteurs sur ce chapitre, le *souci praxéologique* en matière de RII ne semble pas être aujourd'hui une exigence largement partagée. Or, là comme en d'autres domaines, il est une condition essentielle des dynamiques praxéologiques qui peuvent à bon droit être recherchées en toute institution – de l'école à l'entreprise en passant par la vie quotidienne.

Didactique de l'enquête codisciplinaire

La TAD s'est développée au cours des dernières années par son extension à travers une autre théorie, celle des *parcours d'étude et de recherche* (PER). La théorie des PER a donné lieu à un ensemble déjà relativement substantiel de travaux, comme en témoigneront entre autres les actes (à paraître) du II^e congrès international sur la TAD tenu à Uzès (France) du 31 octobre au 3 novembre 2008. Nous en rappelons ici rapidement les principaux aspects utiles pour situer notre travail. Au point de départ se trouve une volonté d'agir face à la situation historique actuelle de l'enseignement scolaire, dont on peut dire, à partir des analyses permises par la TAD, qu'il est la victime d'un processus de dégénérescence didactique accéléré : pour reprendre une formule lapidaire, on tend à y enseigner des réponses à des questions qu'on a oubliées, ce qui signe le diagnostic de perte de sens que chacun peut constater. Chacun a dû, ainsi, en classe de mathématiques, étudier les triangles ; mais qui pourrait dire ce qui motive et justifie cette étude, du seul point de vue mathématique déjà ? La même question peut être soulevée à propos des angles, ou des expressions algébriques, ou des fractions, etc. À cet égard, on peut montrer qu'a existé un état historique de l'éducation scolaire où les raisons d'être des objets étudiés étaient apparentes ; où, dans une mesure variable sans doute, les connaissances enseignées apparaissaient comme constitutives de réponses à des questions dûment explicitées (Chevallard, 2006). À la question « Comment démontrer que deux droites sont parallèles ? », une réponse se construisait ainsi par étapes, sous la forme d'une praxéologie mathématique évolutive comportant d'abord la seule technique consistant à faire apparaître ces droites comme « perpendiculaires à une même troisième », à quoi s'adjoignait ensuite la technique consistant à les faire apparaître comme « parallèles à une même troisième », puis, un peu plus tard, la technique consistant à les faire apparaître comme portant une paire de côtés opposés d'un quadrilatère que l'on saurait par

ailleurs être un parallélogramme – ce qui est *l'une* des raisons pour lesquelles on étudiait au collège les propriétés du parallélogramme.

L'évolution historique de l'enseignement que nous venons de rappeler peut être modélisée, dans le cadre de la TAD, en partant d'un schéma qu'on peut d'abord écrire ainsi : $S(X; Y; Q) \rightsquigarrow R$. Ici, $S(X; Y; Q)$ est un *système didactique*. L'ensemble X est celui des *étudiants*, qui peuvent être des élèves du primaire, des étudiants d'université, des « savants », des militants associatifs ou politiques, etc. L'ensemble Y est celui des *aides à l'étude* et des *directeurs d'étude* : dans une classe scolaire traditionnelle, Y se réduit à *une* personne, le professeur de la matière dont relève la question Q , qui est la question étudiée, dans le système didactique $S(X; Y; Q)$, par le collectif X aidé et dirigé par l'équipe Y . Le but de l'étude – ou du moins son horizon – est d'élaborer une réponse R à la question. Comme le suggère l'exemple des droites parallèles, une réponse, en ce sens, est une praxéologie (ou un « fragment » de praxéologie). Dans un certain type d'enseignement traditionnel, où par exemple Y se réduit à une personne, le professeur y (on a donc, formellement, $Y = \{ y \}$), l'élaboration de R se fait ainsi : il revient à y d'enquêter sur Q , d'étudier Q et d'élaborer *sa* réponse R , que l'on note alors R^\heartsuit (ce qui se lit « r cœur »), peut-être parce que cette réponse lui tient à cœur. Le professeur y va alors présenter à la classe X *sa* réponse R^\heartsuit : « son cours » consistera en un compte rendu de l'enquête qu'il aura conduite – parfois depuis de nombreuses années – à propos de la question Q . Les élèves auront à « apprendre » la réponse R^\heartsuit puis à se montrer capables de la « restituer » (on verra qu'il n'en ira pas autrement dans le paradigme étudié par la TAD, même si les choses ne s'y disent pas de cette manière). Mais ces élèves n'auront pris aucune part à la *construction* de R^\heartsuit . Dans ces conditions, il devient tentant de supprimer l'étape inaugurale, où la classe se réunit autour de la question Q pour former le système didactique $S(X; Y; Q)$. La tentation sera grande alors d'enseigner *directement* la réponse R^\heartsuit , sans passer par la question Q , sans même formuler cette question, qui sera bientôt effacée.

Par contraste avec cette configuration scolaire aujourd'hui encore largement dominante, le dispositif du parcours d'étude et de recherche a pour objectif de revenir à une plus grande authenticité de l'éducation scolaire dans le rapport qu'elle institue à la connaissance, à sa genèse, à ses raisons d'être, à ses *fonctions*. Pour créer ensemble une réponse R^\heartsuit , la classe va créer un milieu didactique M , c'est-à-dire réunir un ensemble de ressources didactiques au service de l'élaboration de R^\heartsuit – lequel n'existe qu'à titre de projet, situation que l'on peut noter ainsi : $(S(X; Y; Q) \rightarrow M) \rightsquigarrow R^\heartsuit$. Le milieu M va rassembler deux types de ressources. D'une part, il réunira des réponses à la question Q existant « dans la culture » : on note une telle réponse R^\diamond (ce qui se lit « r poinçon »), pour signifier que cette

réponse est revêtue de l'estampille d'une institution ou, au moins, d'une personne ; et on notera $R_1^\diamond, R_2^\diamond, \dots, R_m^\diamond$ les différentes réponses « poinçonnées » ainsi collectées pour fournir des matériaux à la construction de R^\heartsuit . D'autre part, le milieu M va réunir des ressources d'une autre nature, qui pourront être des dispositifs et des techniques, des technologies et des théories utiles pour construire R^\heartsuit ; toutes ces œuvres de la culture sont notées de façon générique par la lettre O (pour *œuvre*) et on notera donc $O_{m+1}, O_{m+2}, \dots, O_n$ les œuvres ainsi apportées dans le milieu M . Les réponses R^\diamond sont le *matériau* de l'étude tandis que les œuvres O en sont le *matériel*, ensemble d'outils qui permettront de travailler adéquatement ce matériau.

Voici, à titre d'illustration, une question Q – que l'assertion précédente peut inspirer : « Quelle différence d'emplois existe-t-il entre “matériau” et “matériel” ? » Il s'agit là d'une question de technique d'expression linguistique – comment choisir le bon mot, s'il existe, en tel contexte d'emploi ? Voici trois « réponses » (et leur source).

R_1^\diamond . « **matériau** n. m. **1** – *matière* • **2** – *éléments* • *documents* • *données* • *informations* • *matière* • *support* [...] **matériel** n. m. **1** – *matériau* • *matière* • **2** – *équipement* • *instruments* • *outillage* • *matos*^{fam.} • **3** – *données* • *documents* • *matériau* (Le Fur, 2005, p. 722)

R_2^\diamond . « Le singulier **matériau** a été fabriqué d'après le pluriel **matériaux**, parfois seul considéré comme correct. **Matériau** est de plus en plus courant pour désigner une “matière première” : un matériau de choix. Il ne se confond nullement avec **matériel**, qui désigne “un ensemble d'outils ou d'instruments nécessaires à l'accomplissement d'une tâche”... » (Colin, 1994, p. 361)

R_3^\diamond . « Ce n'est qu'en 1822, dans un contexte militaire, que **le matériel** est attesté avec le sens collectif concret d'“ensemble des objets et de l'équipement nécessaire à une activité”. [...] Dans le langage des ethnologues et sociologues, de même qu'en psychanalyse, matériel est employé pour l'ensemble des éléments soumis à un traitement (les données). Le mot étant collectif, le singulier est de règle, sauf dans quelques contextes appropriés. [...] L'ancienne forme de matériel, material, s'est maintenue au pluriel materials ou materiaus, aujourd'hui **MATÉRIAUX** (1510) [...] Plus abstraitement, il s'applique aux éléments nécessaires à qqch. (1671), spécialement aux éléments de base pour écrire un ouvrage (1690). Quelquefois employé avec un article singulier (1867), emploi encore considéré comme anormal, il a produit un

singulier **MATÉRIAU** (1923) “type de matière entrant dans la construction d’un objet fabriqué”. » (Rey et al., 1993, p. 1204)

On pourrait multiplier les réponses R^\diamond . La tâche serait ici de construire la réponse R^\heartsuit (qui, par exemple, pourrait être la réponse à intégrer dans le « style » de telle revue de sciences de l’éducation).

Une fois le milieu M constitué, la situation se présente ainsi :

$$(S(X ; Y ; Q) \rightharpoonup \{ R_1^\diamond, R_2^\diamond, \dots, R_m^\diamond, O_{m+1}, O_{m+2}, \dots, O_n \}) \hookrightarrow R^\heartsuit.$$

Cette « formule » symbolise une configuration que Y. Chevallard a proposé d’appeler le *schéma herbartien*¹⁵. Deux points doivent être précisés d’emblée : tout d’abord, la constitution du milieu M peut être une affaire de longue haleine, et courir tout au long du processus d’étude et de recherche aboutissant à l’élaboration de R^\heartsuit , soit sur plusieurs mois s’agissant par exemple des TPE ; ensuite, le choix des réponses R^\diamond que l’on retiendra (et, corrélativement, de celles que l’on écartera) de même que le choix des œuvres O apportées dans le milieu sont des choix toujours risqués (ils peuvent se révéler faiblement pertinents, par exemple), même si X et Y tentent de maintenir un tel risque à un niveau raisonnable, en tout cas compatible avec les contraintes institutionnelles sous lesquelles le PER est amené à se dérouler. Ces deux aspects différencient déjà clairement le dispositif du PER du dispositif de la classe traditionnelle décrit plus haut. D’une façon plus large, le « remodelage » des conditions de l’étude peut être regardé comme hautement « mathémagène », selon l’adjectif introduit autrefois par Ernst Z. Rothkopf (1970), c’est-à-dire fortement générateur de connaissances¹⁶. La TAD distingue cinq « gestes » de base impliqués dans la fabrication de la réponse R^\heartsuit . Le premier est l’*observation* des réponses R^\diamond – nous allons y revenir. Le deuxième est l’*analyse* des réponses R^\diamond : là notamment interviennent les outils rassemblés dans le milieu M , ainsi que la « confrontation » entre réponses R^\diamond . Le troisième est l’*évaluation* des réponses R^\diamond ainsi observées puis analysées : il s’agit par là de déterminer la *valeur* de R^\diamond *pour le projet* de construire une réponse R^\heartsuit destinée à être mise en œuvre *sous certaines contraintes en concourant à créer certaines conditions*¹⁷. Le quatrième est le *développement* de R^\heartsuit (à partir des réponses R^\diamond auxquelles auront été reconnue une certaine valeur pour cela). Le cinquième, enfin, consiste en la *défense et illustration* de la réponse R^\heartsuit . Ici, « illustrer » signifie « rendre illustre » : dans le cadre scolaire, il s’agira pour les membres de X de présenter R^\heartsuit dans un examen oral ou écrit, devant un jury (ce qui est le cas de la thèse d’université, mais aussi des TPE), dans un « colloque » organisé par l’école (comme il en va avec les « doctoriales » par exemple à l’université), dans un article publié dans un journal de lycéens, etc. Et de même, le

cas échéant, pour les membres de Y . Ce « geste » de défense et illustration est à l'évidence fortement mathémagène (y compris pour les membres de Y !). Ce qui est *appris*, alors, c'est la réponse R^\heartsuit elle-même et tout ce qu'elle incorpore de connaissances apportées dans le milieu sous une forme intégrée aux réponses R^\diamond et aux œuvres O . Mais chacun des quatre autres gestes (observer, analyser, évaluer les réponses R^\diamond , développer la réponse R^\heartsuit), par le jeu souvent intense de questions/réponses qu'il implique¹⁸, a également des vertus mathémagènes importantes, qui font aussi la valeur formatrice des PER : tout au long du parcours, on apprend, même si certaines des connaissances momentanément acquises pourront être oubliées – parce qu'elles ne seront pas institutionnalisées avec R^\heartsuit .

Les *conditions* créées par un PER appellent la maîtrise et la mise en œuvre, de la part de X comme de Y , de praxéologies d'étude en grande partie scolairement inédites, à propos notamment (mais pas seulement) de la recherche de « ressources » pour constituer le milieu M . Une courte nomenclature a été d'emblée proposée (Chevallard, 2001a), qui comporte six « dialectiques », ce mot désignant en sciences humaines, selon le *Dictionnaire historique de la langue française* (Rey et al., 1993, p. 598), « toute méthode d'investigation qui repose sur le principe d'antinomie et de dépassement de l'antinomie ». La première de ces dialectiques est celle *du sujet et du hors sujet*. Dans ce cas, l'antinomie cruciale se nourrit de l'opposition entre l'injonction scolaire de ne pas faire, si peu que ce soit, du « hors sujet » et la nécessité tout au long d'un parcours d'étude et de recherche d'accepter le risque du hors sujet, par exemple quand on recherche des documents dont, avant un examen plus approfondi, on ne peut guère savoir s'ils sont « dans le sujet » ou en demeurent éloignés, ou encore quand on croit devoir examiner une œuvre – une théorie, une idée expérimentale, etc. – qui pourrait être précieuse dans le progrès de l'étude, tout en prenant le risque de déchanter, ou en soulevant telle question Q^* qui *pourrait* être l'une des clés ouvrant la route de la réponse à la question Q étudiée, etc. La deuxième dialectique, à la dénomination curieuse¹⁹, est celle *du parachutiste et du truffier*. En ce cas, l'antinomie s'établit avec le double habitus scolaire de la rareté documentaire et de la recherche de l'adéquation quasi immédiate du document cherché au projet qu'il doit servir – l'élaboration de R^\heartsuit . Par contraste, la dialectique du parachutiste et du truffier conduit régulièrement à assumer de « ratisser » (à l'instar du parachutiste militaire une fois au sol) de vastes zones, où l'on sait *a priori* qu'on ne trouvera sans doute pas grand-chose, mais où pourra advenir *de l'inattendu*, et où l'on apprendra (ici se situe la référence au truffier) à repérer de précieuses et rares trouvailles – souvent peu visibles dans un paysage indéfiniment recommencé –, qui feront progresser la recherche. La troisième dialectique est celle dite *des boîtes noires et des boîtes claires*. L'antinomie, en ce cas, naît du primat donné, dans l'éducation scolaire traditionnelle, à la connaissance *déjà* disponible et au travail de

clarification des boîtes... claires, sans considérer la pertinence des savoirs qu'elles recèlent pour le projet en cours supposé – qui souvent, au reste, n'existe pas ! Par contraste, dans un PER, il convient de rétablir le primat de la connaissance *pertinente*, quel que soit *a priori* son statut au regard des savoirs institutionnellement établis, tout en limitant le travail de clarification pour disposer d'un contrôle épistémologique solide, sans donc céder au fantasme scolaire de la clarté totale, car les boîtes réputées « claires » sont *toujours* des boîtes *grises*.

Les trois dialectiques suivantes ne sont pas indépendantes des précédentes. Mais elles marquent le passage du milieu « brut », tel que les trois premières dialectiques aident à le constituer, au *travail* sur ce milieu, d'où sortira éventuellement la réponse R^\heartsuit désirée. La quatrième dialectique, celle dite de *l'excription et de l'inscription* doit permettre d'échapper au *recopiage formel* de textes où ont été inscrites des réponses R^\diamond , réponses que cette mise en texte a souvent plus ou moins « dévitalisées » : le travail d'*excription* (il s'agit là d'un néologisme) a pour but de redonner vie aux réponses R^\diamond déposées dans les documents disponibles ; le travail d'*inscription* n'est pas alors un simple effort de réinscription textuelle de ces réponses R^\diamond mais un travail « rhétorique » inédit visant à constituer par étapes une réponse propre R^\heartsuit , laquelle prend forme peu à peu *par le croisement de plusieurs niveaux d'écrit* (carnet de bord du PER, notes de synthèse, glossaire, production finale, etc.). Dans le travail d'élaboration de R^\heartsuit , la cinquième dialectique est celle *de la conjecture et de la preuve*. L'antinomie met ici face à face l'usage scolaire de mises à l'épreuve plus ou moins réglées à l'avance d'assertions réputées sûres en vertu surtout de l'autorité de l'institution et la nécessité de soumettre les assertions surgies du travail sur le milieu M , qu'aucune institution « surplombante » ne garantit, à une critique spécifique, utilisant tous les instruments idoines, et conduisant, non à une vérité absolue donnée d'avance, mais à des assertions toujours conjecturales dont cependant le « *degré d'incertitude* » est raisonnablement petit. Cette cinquième dialectique a fait l'objet depuis quelques années d'un développement important, sous le nom de dialectique *des médias et des milieux*. Dans un document introductif au récent congrès d'Uzès déjà mentionné, que nous nous autoriserons à reprendre ici, on pouvait lire ceci.

Le mot de *média* désigne, dans ce cadre, tout système de mise en représentation d'une partie du monde naturel ou social à l'adresse d'un certain public : le « cours » du professeur de mathématiques, un traité de chimie, le journal d'un présentateur de télévision, un quotidien régional ou national, un site Internet, etc., relèvent en ce sens du système des médias. Un *milieu* est entendu ici dans un sens voisin de celui de *milieu adidactique* en TSD [théorie des situations didactiques] : on désigne en effet comme étant un milieu tout système qu'on peut regarder comme *dénué*

d'intention didactique dans la réponse qu'il peut apporter, de manière explicite ou implicite, à telle question déterminée. Le système considéré se comporte alors à cet égard comme un *fragment de « nature »*. Par contraste, à propos de nombre de questions qu'on entend leur poser, les médias sont en général mus par une certaine intention, didactique ou hypo-didactique, par exemple l'intention « d'informer ». Bien entendu, un média peut fort bien, à propos de telle question particulière, être regardé comme un milieu, et être utilisé comme tel.

L'enjeu des travaux sur ce thème était alors explicité comme suit.

L'existence d'une dialectique vigoureuse (et rigoureuse) entre médias et milieux est une condition cruciale pour qu'un processus d'étude et de recherche ne se réduise pas au recopiage acritique d'éléments de réponse épars dans les institutions de la société. Une telle exigence est en vérité consubstantielle à l'*esprit galiléen* caractéristique des sciences modernes de la nature et de la société, dans lequel la soumission à l'autorité cède la place à une culture partagée du questionnement, de la mise à l'épreuve par la construction de milieux idoines, déterministes ou statistiques, combinant dispositifs matériels et immatériels (enquête, expérimentation, raisonnement, déduction). En conséquence, l'un des grands problèmes éducatifs et citoyens de notre temps est celui de la généralisation de la capacité (de l'élève, du professeur, du formateur, du chercheur, du citoyen, etc.) à situer sa pensée et son action dans une dialectique des médias et des milieux adéquate à l'évaluation de ses assertions et de ses décisions.

On notera ici que, loin d'aller de soi (selon une croyance spontanée évoquée plus haut), le fait de pouvoir valablement « jouer contre la nature » est, en quelque contexte institutionnel que ce soit, le fruit d'un travail épistémologique et méthodologique toujours à reprendre.

La sixième et dernière des dialectiques actuellement considérées par la TAD est celle *de la diffusion et de la réception*. Alors que, dans l'usage scolaire, l'élève n'a pas à s'inquiéter de « faire connaître » la réponse qu'il ne fait guère que restituer au professeur qui la lui a apportée toute faite (par exemple), il en va autrement dans une situation de PER : dès lors que la réponse produite, R^\heartsuit , apparaît si peu que ce soit inédite dans l'institution, elle doit y être diffusée (par des canaux convenus) en même temps que sa réception par le « public » visé (les élèves de la classe, le jury de mémoire, etc.) doit y être gérée de manière éventuellement anticipée, par exemple en redéfinissant durant le PER le système des contraintes auxquelles on souhaite que la réponse R^\heartsuit à produire satisfasse.

Une ultime antinomie doit être encore soulignée entre culture des PER et culture scolaire traditionnelle. Dans cette dernière, les connaissances sont, dans une large mesure,

confinées à l'intérieur de *disciplines* – biologie, histoire, chimie, anglais, etc. – souvent très soucieuses de délimiter strictement leur territoire. De fait, il est possible de conduire certains PER dans un contexte fortement (sinon totalement) *monodisciplinaire* : les PER ont d'abord été conçus, à l'origine, comme des dispositifs didactiques en classe de mathématiques. Mais, ainsi qu'on le verra dans la suite de ce travail, pour nombre de questions Q qu'une personne ou une institution peut être amenée à étudier, les connaissances utiles – par exemple celles matérialisées dans les œuvres O qui entreront dans la composition du milieu M – relèvent de plusieurs domaines disciplinaires, qu'on se réfère à la carte des disciplines de l'enseignement secondaire ou à celle du monde savant. Ces disciplines doivent ainsi *concourir*, en synergie praxéologique dans le cadre du système didactique $S(X; Y; Q)$, à l'élaboration de R^\heartsuit : on parle alors de *codisciplinarité*. L'enquête sur la question Q , censée être monodisciplinaire lorsqu'elle se déroule dans la classe de biologie ou de mathématiques, devient ainsi enquête *codisciplinaire*. Pour être plus près de la vérité, on peut alors regarder toute enquête supposée monodisciplinaire comme une enquête codisciplinaire, dans laquelle le recours à des disciplines autres qu'une certaine discipline dominante peut parfois être réduit à un minimum compatible avec le projet d'élaborer une réponse satisfaisant certaines contraintes. Élaborée d'abord dans le contexte de l'enseignement des mathématiques au collège et au lycée, la théorie des PER apparaît dès lors comme située au cœur d'une *didactique de l'enquête codisciplinaire* en construction.

Dans le travail que suppose la mise en œuvre du schéma herbartien, l'utilisation des ressources du Web est aujourd'hui devenu essentiel, d'abord comme simple *moyen* au service d'une fin qu'on peut toujours concevoir comme indépendante de l'existence de l'Internet, même si de nombreuses enquêtes ne pourraient plus avoir lieu aujourd'hui sans un tel système de ressources. C'est ainsi que, s'agissant de la question Q relative à l'emploi des mots *matériau* et *matériel*, par exemple, s'ajouteront aujourd'hui aux réponses R_1^\diamond , R_2^\diamond et R_3^\diamond vues plus haut bien d'autres réponses, telle la suivante.

R_4^\diamond . « Au figuré, matériau et matériel sont encore plus proches. Matériau (ou son pluriel matériaux, si l'on veut insister sur la multiplicité ou la diversité des sources) désigne un ensemble de données, d'informations qui fournit la matière de base à un texte, un ouvrage, une recherche. Dans le domaine des sciences humaines et sociales (ethnologie, sociologie) ou en psychanalyse, on emploie le terme matériel pour désigner l'ensemble des éléments qui seront traités ou analysés. Et pour ajouter à la complexité sémantique, considérons de plus l'emploi possible de matière dans des contextes

similaires. En effet, matière renvoie au contenu, au sujet d'un ouvrage ou d'une étude ou encore au fond ayant servi à l'élaboration de ceux-ci.

Exemples :

- Les résultats du sondage ont servi de **matériau** (ou : de **matériel**) de base à l'étude en cours.*
- Elle a puisé le **matériau** (ou : la **matière**) de son roman dans ses propres souvenirs.*
- Les écrits de l'auteur ont fourni un **matériau** (ou : un **matériel**) riche et utile à la connaissance de cette période charnière dans l'histoire du pays.*
- On a confié à des étudiants le classement et l'analyse du **matériel** recueilli dans les enquêtes.*
- L'équipe est rentrée avec un **matériel** ethnographique d'une valeur inestimable.*
- La collecte du **matériel** linguistique qui a servi à l'élaboration de l'ouvrage s'est étalée sur plus de dix ans. »*

(Office québécois de la langue française, Banque de dépannage linguistique, Matériau et matériel : http://66.46.185.79/bdl/gabarit_bdl.asp?id=3861)

Cette exigence désormais irréfutable conduit à soulever, dans le cadre de la didactique de l'enquête codisciplinaire, le problème de la diffusion et de la maîtrise des praxéologies utiles au bon usage du Web. Selon un scénario classique, dans une première étape (qui n'est sans doute pas achevée entièrement aujourd'hui), le recours au Web a d'abord été regardé comme se faisant à coût praxéologique quasi nul – pour faire de la RII, il n'y aurait « rien à savoir ». Dans un second temps (dont ce travail pourrait avec d'autres contribuer à hâter l'avènement), on redécouvre les exigences de la dialectique outil/objet (Douady, 1986) en retrouvant bientôt que le bon usage de l'*outil* suppose une *étude* de cet outil en tant qu'*objet* : comme on le verra, le développement des pratiques d'enquête codisciplinaire, c'est-à-dire, désormais, d'enquête codisciplinaire *avec Internet*, rencontre aujourd'hui l'obstacle de la faiblesse persistante de la culture courante en matière de RII. C'est sur ce point capital que notre travail vise à apporter une contribution dont nous allons maintenant nous efforcer de préciser la nature.

Problématiques de recherche connexes

La didactique, on l'a dit, étudie la diffusion des praxéologies, ses conditions sous des contraintes données, ses mécanismes. En matière de RII, une première problématique doit être mentionnée qui n'est pas sans affinité avec le postulat plusieurs fois évoqué selon lequel « il

n’y aurait rien à savoir » ou presque, puisqu’un équipement praxéologique minimal, dont l’acquisition serait à la portée de tous, est tout ce qu’on pourrait désirer. Dans cette problématique, qui consiste à neutraliser quasiment la variable placée au cœur de notre travail – l’équipement praxéologique de l’usager –, que ce soit par idéologie (comme on l’a vu) ou par méthode (seul point de vue scientifiquement acceptable), il ne reste plus alors, du jeu à trois joueurs mentionné plus haut, que deux joueurs, les *pages et sites Web*, d’un côté, et les *outils de recherche*, de l’autre. Du point de vue de la didactique de la RII, qui est le nôtre ici, il y a bien là deux personnages clés d’une culture de la RII, sur lesquels les travaux existants apparaissent on ne peut plus utiles, ainsi qu’on le verra dans la suite de ce travail. Ces travaux, qui ne s’adressent pas *en principe* à l’usager ordinaire du Web, s’étagent au long d’une montée dont le « camp de base » est constitué de publications plus pratiques que théoriques (tels Pillou, 2005 ; Warbesson, 2007 ; Lemainque, 2008) portant à la connaissance des créateurs amateurs ou semi-professionnels de pages Web le fruit d’observations diverses et de conclusions pratiques, quelquefois lacunaires ou hâtives, sur le fonctionnement des objets qui peuplent le Web et l’art de se comporter avec eux : c’est là une littérature sur le contenu de laquelle nous aurons amplement l’occasion de revenir. Au-delà, on arrive à une zone à l’atmosphère déjà plus raréfiée, celle des spécialistes professionnels (ou semi-professionnels) du Web. Un large tableau de ce niveau d’activité autour des objets du Web a été récemment brossé par Olivier Andrieu dans son livre *Réussir son référencement web* (Andrieu, 2008) : c’est à ce niveau²⁰ qu’évoluent les grands sites Web dédiés aux problèmes de la *Web search* – la RII –, tels que les sites français *Abondance* (le site de d’O. Andrieu lui-même) ou *WebRankInfo* (qui est le site d’un autre spécialiste, Olivier Duffez), ainsi que les sites internationalement connus que sont *SearchEngineWatch* (créé par Danny Sullivan) ou *Search Engine Showdown*, le site de Greg R. Notess, auteur d’un ouvrage important sur la RII (Notess, 2006), et bien d’autres encore dont on trouvera une riche webographie dans le livre d’Andrieu (pp. 293-297).

Au sommet de ce parcours, dans le prolongement du passionnant témoignage de Tim Berners-Lee sur la naissance historique du Web (Berners-Lee, 2000), on doit placer le projet récemment formulé d’une *Web science*, une « science du Web », présenté dans le numéro d’août 2006 de la revue *Science* dans un article intitulé “Creating a Science of the Web”, où l’on peut lire notamment ceci (Berners-Lee, Hall, Hendler, Shadbolt & Weitzner, 2006).

Web science is about more than modeling the current Web. It is about engineering new infrastructure protocols and understanding the society that uses them, and it is about the creation of beneficial new systems. It has its own ethos: decentralization to avoid social and technical bottlenecks, openness to

the reuse of information in unexpected ways, and fairness. It uses powerful scientific and mathematical techniques from many disciplines to consider at once microscopic Web properties, macroscopic Web phenomena, and the relationships between them. Web science is about making powerful new tools for humanity, and doing it with our eyes open.

Ce projet, intitulé *Web Science Research Initiative*, s'est évidemment donné un site (<http://webscience.org/>), sur lequel on trouve notamment un texte important, fournissant un cadre de développement à la « science du Web » (Berners-Lee, Hall, Hendler, O'Hara, Shadbolt & Weitzner, 2006). Nous sommes là sur un toit du monde Internet, où l'on croise également des travaux universitaires à l'accès exigeant (tels Liu, 2007 ou, en français, Boughanem & Savoy, 2008) ainsi que de rares travaux venant de l'intérieur de l'industrie du Web, qu'on a bien voulu porter à la connaissance des *outsiders* : ainsi en va-t-il par exemple de tel article (Burges et al., 2005) qui lève le voile sur certains aspects du moteur de recherche de Microsoft. Un texte de vulgarisation de haut niveau – il s'agit de *Web dragons* (Witten, Gori, & Numerico, 2007), déjà cité –, assure l'interface avec les usagers non spécialisés mais curieux ou soucieux de comprendre ce qui se passe derrière la *query box*, en leur révélant “the myths of search engine technology”. À cela, il faut ajouter des études d'histoire immédiate ciblant par exemple la saga de Google (Battelle, 2006 ; Vise & Malseed, 2006), qui apportent souvent une foule d'informations, notamment sur les acteurs principaux (comme Sergey Brin et Larry Page, les fondateurs de Google) mais aussi « secondaires » (le livre de Vise et Malseed comporte ainsi un chapitre sur le rôle joué par Danny Sullivan, dont le nom a été cité plus haut), sur les événements les plus marquants, tout en mettant l'accent sur les événements financiers et les logiques économiques surtout.

Le rêve caressé par les fondateurs de Google, celui de la *perfect search*, la « recherche parfaite », soit “the ability to ask any question and get not just an accurate answer, but *your* perfect answer, an answer that with eerie precision is informed by who you are and why you're asking” (Battelle, 2006, p. 252), et que concrétise en attendant mieux le bouton « J'ai de la chance » (*I'm Feeling Lucky*), qui concurrence le bouton « Recherche Google » (*Google Search*) sur l'interface du moteur de recherche Google²¹, répond subrepticement, dans un registre plus ambitieux, certes, à ce qui semble bien être, pour beaucoup d'usagers, l'essentiel de leur équipement praxéologique en matière de RII : le Web ou plutôt le moteur de recherche étant regardé comme un fragment de nature, on choisit (des mots clés), on clique, on attend (très peu) ; si la manœuvre « échoue », on peut seulement la recommencer, quasi à l'identique. Telle est fréquemment, de fait, notre attitude face à beaucoup de systèmes regardés à tort ou à raison comme « naturels » : si l'on met de l'eau à bouillir sur une plaque

chauffante, le système ainsi « lancé » sera regardé comme *autonome*, doté de ses propres lois, sur lesquelles on ne saurait intervenir ; on n'a donc plus qu'à attendre que l'eau bouille – ou pas. Dans un tel paradigme praxéologique, l'utilisateur minore son propre rôle face à la « machine », censée faire ce qu'elle veut – ou plutôt ce que veulent les lois qui la gouvernent.

Ces remarques nous conduisent à nous arrêter maintenant sur des travaux relevant d'une deuxième problématique de recherche en matière de RII : celle de *l'exploration de l'équipement praxéologique actuel des usagers*. Très tôt des enquêtes par questionnaire ont été réalisées et continuent de l'être sur divers aspects de cette question. Une des plus récentes a été conduite en avril-mai 2008 aux États-Unis par le *The Pew Research Center's Internet & American Life Project* (Fallows, 2008), organisation qui se présente comme “a nonprofit ‘fact tank’ that provides information on the issues, attitudes and trends shaping America and the world” et dont les travaux, en principe, “are nonpartisan and take no position on policy issues”. Son titre attire l'attention sur un fait qui s'est imposé depuis la précédente enquête : “Almost half of all internet users now use search engines on a typical day.” L'usage de l'Internet pour le courrier électronique reste certes en tête, 60 % des utilisateurs de l'Internet y ayant recours « un jour typique », alors que seulement 49 % de ces mêmes utilisateurs sollicitent des moteurs de recherche ; mais en six ans environ, le premier pourcentage a gagné 8 points tandis que le second augmentait de quelque 15 points, ce qui est en effet digne d'être souligné : comme l'écrit l'auteure, “these new figures propel search further out of the pack, well ahead of other popular internet activities, such as checking the news, which 39% of internet users do on a typical day, or checking the weather, which 30% do on a typical day”. Pour le reste, l'étude établit que, dans l'état présent, l'emploi de moteurs de recherche croît avec le niveau d'éducation et les revenus, décroît avec l'âge, est plus important dès lors que l'on dispose d'une connexion à haut débit chez soi et que l'on est un homme (le pourcentage des hommes utilisant des moteurs de recherche « un jour typique » est passé de 33 % en 2002 à 53 % en 2008, tandis que le pourcentage des femmes faisant de même passait de 25 % à 45 %). Dans un communiqué de presse du 10 octobre 2007, ComScore, qui se présente comme “a Global Internet Information Provider”, rapportant les résultats d'une enquête conduite en août 2007, indique ²² : “The study found that more than 750 million people age 15 and older – or 95 percent of the worldwide Internet audience – conducted 61 billion searches worldwide in August, an average of more than 80 searches per searcher.” Ce dernier chiffre s'établit à 77,4 pour les États-Unis et 85,1 pour l'Europe. Pour la France en particulier on dispose des enquêtes annuelles du CRÉDOC (Bigot, 2005 ; Bigot, 2006 ; Bigot & Crouette, 2007) ; seule l'enquête réalisée en juin 2005 comportait, dans le questionnaire utilisé, la modalité de réponse « Rechercher des informations » – en général – à la question « Parmi les

utilisations suivantes, désignez celles que vous pratiquez à domicile sur votre micro-ordinateur ? ». Parmi les personnes équipées d'une connexion à Internet, lit-on dans le compte rendu de l'enquête (p. 12), « 93 % s'en servent pour rechercher des informations, 82 % pour envoyer des mails, 39 % pour télécharger des films, de la musique ou des logiciels, 32 % fréquentent des forums de discussion et 8 % utilisent leur ordinateur pour téléphoner ». Ici, contrairement au cas des États-Unis, la RII rend des points au courrier électronique ; pour le reste, on retrouve des corrélations classiques (p. 42) : la recherche d'informations « est un peu plus répandue parmi les diplômés, les cadres et les personnes vivant dans un ménage aux revenus les plus élevés ; elle reste néanmoins assez fréquente dans les autres catégories de la population ». Il semble en revanche que les différences liées au sexe n'interviennent guère (Bigot, 2005, p. 134).

Ce qui précède suggère que l'activité de RII joue un rôle croissant dans les sociétés contemporaines. Mais cela ne dit rien cependant quant aux praxéologies de la RII diffusées dans la société. Une autre enquête conduite pour le *Pew Internet & American Life Project* par l'auteure déjà citée a été réalisée mai-juin 2004 (Fallows, 2005). Le titre en résume l'essentiel : "Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve". Le premier résultat énoncé par l'auteure est que "Internet users are very positive about their online search experiences": il s'agit là d'un point essentiel, qui apparaît d'emblée comme une source de contraintes sur la diffusion des praxéologies de la RII – *en favorisant leur non-diffusion*. Voici à cet égard la conclusion de l'auteure (p. i).

Users paint a very rosy picture of their online search experiences. They feel in control as searchers; nearly all express confidence in their searching skills. They are happy with the results they find; again, nearly all report that they are usually successful in finding what they're looking for. And searchers are very trusting of search engines, the vast majority declaring that search engines are a fair and unbiased source of information.

En fait, ils sont 92 % à se dire confiants dans leurs "searching skills", 48 % – presque la moitié – se déclarant même "very confident". S'ils ne sont que 17 % à déclarer qu'ils trouvent *toujours* ce qu'ils cherchent, ils sont 70 % à dire qu'ils le trouvent "most of the time". Le deuxième résultat est que "most searchers use search engines conservatively", ce qu'il faut entendre ainsi : "Nearly all settle into a habit of using one or just a couple search engines, with only a very few searchers branching out to try more than three." C'est déjà là une indication praxéologique : l'emploi d'une variété d'outils de recherche, souvent évoqué dans la littérature prescriptive en matière de RII, reste une frontière à conquérir – même si 48 %

des usagers interrogés disent utiliser « deux ou trois » moteurs de recherche différents. Le troisième résultat est plus informatif : l’auteure le résume en écrivant que “most internet users are naïve about search engines”. Cette naïveté est un fait culturel, et non pas individuel : alors que nous sommes tous capables de distinguer, dans un journal ou à la télévision, ce qui est une publicité – non subliminale – et ce qui ne l’est pas, l’enquête fait apparaître que “only about 1 in 6 searchers say they can consistently distinguish between paid and unpaid results”. Cette incapacité technique va de pair avec l’idée que “nearly half of all users say they would stop using search engines if they thought engines were not being clear about how they present their paid results”. Le quatrième résultat est que les usagers de l’Internet “turn to search engines for both their important and their trivial questions”. Sur cet aspect des choses, les études sont anciennes (voir ainsi Spink, Wolfram, Jansen & Saracevic, 2001). Parmi les questions proposées figurent celle-ci : “How much of the information you search for online is information that you wouldn’t bother looking up if you didn’t have access to an Internet search engine?” Parmi les usagers interrogés, 8 % répondent *All*, 25 % *Most* et encore 41 % *Some*. Ils sont 24 % seulement à répondre *Very little*. À ce propos, un enquêté commente ainsi sa réponse.

I just completed a course on the Operas of Giacomo Puccini. With the internet I was able to easily obtain his biography and a complete listing of all his works, various reviews of the pieces and as much detail as I wanted. All of this without leaving my desk and to obtain this when it was convenient to me.... (without a search engine) I probably would have not bothered to obtain that information.

Cette stratégie, que l’on peut estimer *a priori* « raisonnable », renvoie à un phénomène classiquement repéré en TAD sous le nom de déni de problématicité²³ : on tend à s’en tenir aux tâches qui n’apparaissent pas *a priori* problématiques, en y renonçant dès lors que la problématicité de la tâche envisagée devient évidente, en « oubliant » ensuite de tels épisodes. Ce phénomène – on ne cherche que ce que l’on sait pouvoir trouver et l’on ne se souvient avoir cherché que ce que l’on sait avoir trouvé – explique sans doute en grande partie la satisfaction déclarée par nombre d’utilisateurs des moteurs de recherche vis-à-vis de leur équipement praxéologique en matière de RII.

L’attitude faiblement critique à l’endroit des informations recueillies à l’aide des moteurs de recherche – 68 % des usagers interrogés pensent que ceux-ci sont “a fair and unbiased source of information” – se conjugue sans doute avec le phénomène précédent pour donner des usagers de l’Internet “confident, satisfied and trusting”²⁴. N’accomplir que des

tâches de recherche non problématiques, se satisfaire des résultats trouvés aboutit régulièrement à croire et à dire que la RII ne suppose aucun savoir. Cela bien sûr ne peut faire que les usagers ne soient dotés, de fait, d'un équipement praxéologique en matière de RII, même quand celui-ci s'est construit de bric et de broc ! L'étude de cet équipement ne s'est pas limitée aux quelques aspects déjà examinés. Dans cette perspective, l'une des explorations les plus suggestives est sans doute celle dont les auteurs, Leah Graham et Panagiotis Takis Metaxas (2003), ont rendu compte sous ce titre accrocheur mais révélateur : *"Of course it's true; I saw it on the Internet!" Critical Thinking in the Internet Era*. Un premier résultat de leur étude – conduite auprès de 180 étudiantes du collège de Wellesley²⁵ qui suivaient un enseignement intitulé "Computers and the Internet" (Informatique & Internet) – est que, "though the survey was not in any way limited to Internet resources, less than 2% of students' responses to all questions included non-Internet sources". La conclusion des auteurs est nette et va dans le sens opposé au « Il n'y a rien à savoir » déjà commenté : ils insistent au contraire sur "the importance of teaching good Internet research skills, as students rely so heavily on the Internet." L'étude menée à bien redécouvre à la fois le fait massif du recours majoritaire à un moteur de recherche et la confiance spontanée dans les résultats que celui-ci apportera. Mais deux autres résultats méritent surtout d'être soulignés. Tout d'abord, le passage du temps n'a pas d'influence significative sur la qualité de l'équipement praxéologique des étudiantes concernées ; à cet égard, les auteurs notent (p. 74) : "Using class year, we hoped to see if students became better Internet researchers over the course of their years at Wellesley. The results indicate there was no significant difference in performance based on class year." Ensuite, le fait de se déclarer très confiant (ou assez confiant) dans ses capacités en matière de RII ne se traduit pas de façon évidente par une meilleure réussite pratique – "The confidence a student has in his or her ability to effectively search the Internet does not significantly affect the student's performance."

Les étudiantes ayant participé à l'étude ont eu à répondre, au moyen de recherches sur Internet, à différentes questions. L'une d'elles était la suivante : "List three major innovations developed by Microsoft over the past 10 years." Une majorité parmi elles répondirent à cette demande en consultant uniquement le site de... Microsoft. Seules 12 % d'entre elles récupèrent plusieurs réponses R_1^\diamond , R_2^\diamond , R_3^\diamond , ..., pour les confronter les unes aux autres. Une autre question avait été choisie pour engendrer à coup presque sûr une réponse clairement erronée lorsqu'on ne consultait que la réponse R^\diamond la plus probablement rencontrée : "Did the 1999 Rambouillet Accords allow NATO to operate in all of Yugoslavia or only in Kosovo?" Les auteurs écrivent : "The correct answer—

all of Yugoslavia—can be found in the actual document, though it is difficult to wade through its 82 pages.” Bien que le texte en question ait été consultable en ligne, il était en effet tentant de ne s’arrêter que sur des documents le résumant ou en donnant un bref compte rendu. Or voici le piège dans lequel il était facile de tomber : “A frequently found summary is the U.S. Department of State Bureau of European Affairs fact sheet released on March 1, 1999, which implies that NATO presence is limited to Kosovo.” Cette fois, 10 % seulement des étudiantes établirent la réponse correcte en la vérifiant à l’aide de deux réponses R_1^\diamond et R_2^\diamond . Deux autres questions, également anodines, mettaient en outre les étudiantes dans des situations en vérité très différentes : la première demandait de trouver l’identité du créateur de Linux, ce qui ne pose guère problème : 78 % des étudiantes trouvèrent la bonne réponse, mais sans la vérifier ; la seconde question consistait à trouver l’aire de la ville de Lisbonne. Selon les auteurs, “while this sounds elementary, it can take a tremendous amount of time to locate any answer on the Internet, and even longer to find a second source”. Pourtant, dans ce cas aussi, 75 % des étudiantes donnèrent la première réponse trouvée, sans vérification ! Le fait que les étudiantes observées aient eu à chercher de l’information pour répondre à de « vraies » questions permet de poser plus rigoureusement le problème de la RII, trop souvent rabattu sur l’exemple de recherches « immédiates » (ce qui en gomme la problématique). Notons que, dans la même veine, Y. Chevallard avait proposé aux auditeurs de son séminaire 2006-2007 de formation de formateurs une question que l’on peut formuler ainsi (Chevallard, 2007a, p. 214). Dans un texte intitulé Qu’est-ce qu’un dispositif ? (Agamben, 2007), le philosophe italien Giorgio Agamben écrit ceci : « En philosophie, les questions terminologiques sont importantes. Comme l’a dit un philosophe pour lequel j’ai le plus grand respect, la terminologie est le moment poétique de la pensée. » Qui est donc ce philosophe qui a dit que « la terminologie est le moment poétique de la pensée » ? Question véritable (elle peut venir à l’esprit de quiconque ouvre le petit ouvrage d’Agamben) et en même temps question bien plus redoutable que celle portant sur le créateur de Linux !

L’examen à travers la littérature ou de façon directe – notamment par l’observation naturaliste de classes, que nous avons pratiquée durant plusieurs années – des praxéologies de la RII conduit au constat de l’absence actuelle d’une culture praxéologique en la matière qui soit à la fois bien développée et bien diffusée, dans laquelle baigneraient largement personnes et institutions. En quelque domaine que ce soit, une telle culture rendrait en effet à peu près impossibles certains « gestes » (au plan de la *praxis*) et certaines « explications » (au plan du

logos), dont nous ne donnerons ici que quelques exemples. Dans son livre déjà mentionné, Notess rapporte le cas d'une personne qui, après que l'auteur lui eut montré où saisir l'URL de la page qu'elle souhaitait voir s'afficher – à savoir, dans la barre d'adresses du navigateur installé sur son ordinateur, soit Netscape (qui a cessé aujourd'hui d'exister comme tel) –, s'acquitta tout d'abord de cette tâche de façon parfaitement correcte. Notess poursuit ainsi.

However, since the Web was new to her, she then looked for somewhere to click on the screen rather than just pressing the Enter key as I had done (something I hadn't thought to emphasize). Seeing the Search button above the address box in the top tool bar (...), she clicked on it. (p. 20)

Le "Search button" en question est en fait le bouton de recherche du moteur de recherche configuré dans la barre d'outils figurant en l'espèce au-dessus de la barre d'adresses : lorsque la boîte de requête de ce moteur est vide, comme ce devait être le cas dans l'épisode qui nous est rapporté, si l'on clique sur ce "Search button", on voit s'afficher l'interface de recherche du moteur – sa "Search page" –, en attente d'une requête. Voici donc la suite de l'aventure.

The Search button did bring up a new Web page—not for the URL she had painstakingly entered into the address box, however, but for the Netscape Search page. The Netscape Search page contained another box. Seeing the new box, she entered the URL there and clicked on the Search button above it. The Netscape Search page reloaded once more, with the unfortunate consequence of erasing the URL she had just entered for a second time.

Soulignons un fait crucial : loin de conforter le point de vue courant sur le comportement de la personne comme exprimant une « propriété » *intrinsèque* de cette personne, la TAD y voit l'effet sur la personne d'assujettissements institutionnels anciens et actuels, dont ce comportement est simplement révélateur. Ce qu'on peut dire de prime abord, ici, c'est que la personne mentionnée n'est pas assujettie, lorsque cet épisode a lieu, à cette culture du monde Informatique & Internet (et en particulier de la RII) dans laquelle, par contraste, baigne l'auteur cité – qui, lui, en est un spécialiste. En un sens, donc, on peut simplement avancer que cette culture n'est pas parvenue jusqu'à elle. Mais de semblables faits sont trop fréquents, en dépit du pudique silence qui les entoure – et l'antienne « Il n'y a rien à savoir » n'est-elle pas à cet égard comme un voile sonore jeté sur la nudité des faits ? –, pour qu'on n'y aperçoive pas un fait essentiel : une culture de la RII adéquate aux besoins praxéologiques des personnes et des institutions, si elle existe, ne baigne aujourd'hui que de rares et étroits milieux institutionnels.

Au plan pratico-technique, celui de la *praxis*, les faits du type précédent sont sans doute innombrables : il est frappant de voir comment d'une part de nombreux types de tâches restent ignorés (et donc non pratiqués), faute sans doute d'une technique adéquate disponible, et comment d'autre part fleurissent les idiosyncrasies techniques – la manière d'accomplir les tâches de tel type variant de façon inattendue d'une personne à l'autre au sein d'un milieu humain *a priori* homogène, l'un faisant ainsi et ignorant la manière de faire de l'autre, pourtant son semblable à bien des égards²⁶. Faute d'une culture développée et partagée qui homogénéise, cette hétérogénéité résulte d'un *bricolage praxéologique personnel*, plus ou moins articulé à un bricolage institutionnel qui ne concerne guère, au mieux, que des micro-institutions – l'équipement praxéologique se fragmentant au-delà d'une certaine taille de l'institution. Il n'en va pas autrement au niveau technologico-théorique, celui du *logos*. Dans une étude un peu ancienne (Raya et al., 1999) réalisée par une équipe de chercheurs de la *Graduate School of Library and Information Science* de l'Université de Washington (Seattle, État de Washington), la *high school* constituant le terrain de l'enquête venait d'être dotée par Microsoft de matériels performants, dont dix ordinateurs munis de logiciels Microsoft. Les élèves participant à l'étude, au nombre de huit, suivaient un cours d'horticulture et étaient très inégalement familiers de l'Internet. Les auteurs écrivent alors ceci.

Given that Microsoft donated the computers the students used, and that they were searching with Microsoft's Internet Explorer®, it was not completely surprising to discover that half of the participating students believed that Microsoft was responsible for the information on the Internet. Four students assumed that Microsoft collected or generated the information, and stored it on its computer, but their opinions varied about the level of control exercised by the company. While two saw Microsoft involvement on the technical level only, the other two suggested that the office of Bill Gates selected the information on the Internet and directly monitored its use.

Nous sommes là bien loin d'une culture « bien développée » de la RII. Mais on doit constater ceci : lorsqu'une technologie manque, chacun s'efforce, parfois de façon quasi inconsciente, d'inventer une technologie *ad hoc*, qui trouve alors ses matériaux dans son univers proche : c'est là, croyons-nous, un fait incontournable au plan anthropologique. Des situations plus délicates surviennent lorsque « l'inventeur » en question est, non pas un élève (qui a le droit d'être ignorant *au départ*), mais un enseignant, qui, lui, est censé savoir. À cet égard, il nous a été donné d'observer l'épisode suivant, sans doute typique, quoique ordinairement masqué. Dans une classe de 3^e, le professeur de technologie teste par écrit la culture informatique des élèves. À la consigne demandant de citer un navigateur Web, un élève, qui paraît avoir par

ailleurs une implication personnelle forte dans le monde Informatique & Internet, croit bon de répondre en mentionnant *Firefox*, alors que le navigateur utilisé en classe est autre. Le professeur rejettera cette réponse, en « inventant » que Firefox serait, non pas un navigateur, mais une... *secte*. Bien entendu, il est *en principe* facile de s'informer sur Firefox ; mais il est aussi facile de vérifier que l'idée que Firefox serait une secte peut venir à un esprit peu au fait du sujet, utilisateur pressé de l'Internet, comme le montrent les résultats obtenus en présentant au moteur de recherche Google la requête **secte Firefox** ²⁷.



Là encore, il convient de voir d'abord les effets de l'extériorité de cet enseignant par rapport à une culture peut-être introuvable, ou, du moins, qu'il n'a pas véritablement rencontrée – alors même qu'il a reçu mission d'y initier les générations montantes.

Loin d'être isolé, le bricolage praxéologique incontrôlé est ainsi partout. Lors d'un entretien radiodiffusé le 5 août 2008 (Jarraud avec Davidenkoff, 2008) sur France Info, François Jarraud, rédacteur en chef du site *Le Café pédagogique*, qui s'exprime alors en tant que professeur d'histoire et de géographie, déclare ceci en réponse à une question de l'intervieweur – le journaliste Emmanuel Davidenkoff – à propos des difficultés de ses élèves avec l'Internet.

Alors mes élèves, ils ont de très grosses difficultés avec les moteurs de recherche comme ils ont des difficultés en général avec les outils éducatifs. Quand je les observe, je vois bien des jeunes qui

évaluent la réponse de Google en fonction de sa rapidité. Quand il met plus de temps, pour eux il hésite, effectivement, et la réponse est moins sûre.

Si l'interviewé dit vrai, on a là à nouveau, de la part d'élèves, un *bricolage technologique* typique – qui pèse en principe sur le bloc pratico-technique. On notera toutefois que le professeur lui-même, par la formulation qu'il emploie, énonce implicitement une technologie douteuse : est-ce bien Google qui « met plus de temps » à répondre, ou est-ce l'affichage des résultats qui tarde à se faire, parce que le débit de la connexion Internet serait moindre, par exemple ? Ici, le professeur occupe manifestement la place du maître (lequel, contractuellement, est censé savoir, même si son savoir porte la marque d'une certaine incertitude) ; à ce point qu'il regarde les moteurs de recherche comme des « outils éducatifs » particuliers, ce qui, symboliquement, étend *ipso facto* sa maîtrise à ces outils de recherche. Tout cela participe d'une posture que, à l'opposé, le didacticien doit absolument s'interdire vis-à-vis des praxéologies dont il étudie les conditions et les contraintes de diffusion.

Problématique de la recherche

L'étude de l'équipement praxéologique en matière de RII des personnes et des institutions relève de la *statique* didactique. Son apport à la didactique est important, mais partiel : la didactique, nous l'avons souligné, a pour objet la *dynamique* praxéologique. La question « Que sait-elle (*logos*) et que sait-elle faire (*praxis*) ? » en tel domaine d'activité, c'est-à-dire « Quel est son équipement praxéologique ? » en ce domaine, à propos d'une personne ou d'une institution (une classe de collège, par exemple), est une question *instrumentale* en didactique, et non pas une question *définitoire* de la didactique. La question génératrice de la didactique, en effet, posée à propos d'une personne ou d'une institution toujours, est : « Que peut-elle apprendre et que peut-elle apprendre à faire ? », c'est-à-dire « Quel équipement praxéologique peut-elle acquérir ? »

Bien entendu, cette dernière question est incomplète si on ne lui ajoute pas ceci : « Quel équipement praxéologique peut-elle acquérir, *sous des contraintes fixées et dans des conditions à déterminer* ? » La question ainsi complétée appelle alors deux séries de commentaires, qui nous permettront de dégager progressivement notre propre problématique de recherche. Un premier paradigme de recherche conduit à considérer que les contraintes essentielles sont celles dont est porteuse, *hic et nunc*, la personne ou l'institution examinée. Ainsi distinguera-t-on des personnes de différents « types », la typologie correspondante dépendant en partie de l'institution dont se réclament les chercheurs : on parlera par exemple,

en anglais, de *eighth graders* (Yi-Wen, Meilun & Chin-Chung, 2008), de *postgraduates* (Chu & Law, 2007), de *first-year students* (Goodfellow & Wade, 2006-2007), de *teens* (Watkins, 2008), etc. Notons que, chaque fois, ici, le « type » évoqué n'est en rien un concept produit par la recherche, mais un emprunt aux catégories de pensée de l'institution à laquelle les chercheurs se réfèrent, que ce soit l'école (*eighth graders*, *first-year students*, *postgraduates*) ou la société (*teens*) par exemple. Ce qui reste largement absent, alors, ce sont ces contraintes dont les effets ne sont pas supposés être « portés » par les acteurs directs ou très proches de l'institution – de la classe, par exemple. Cet oubli procède, nous semble-t-il, d'un postulat non débattu selon lequel le comportement d'un « système » s'expliquerait par des propriétés *propres au système*, sans qu'on ait à se préoccuper des contraintes *environnementales*. Dans une classe, il y aurait ainsi les propriétés du professeur, les propriétés de tel ou tel élève, celles des praxéologies étudiées, voire celle de « la classe » saisie comme un tout. C'est là bien entendu un point de vue restrictif, qui oblige à prêter au professeur, à l'élève, au savoir enseigné *beaucoup plus* qu'il n'est nécessaire – et juste. Rappelant récemment les circonstances de la naissance de la didactique (des mathématiques) et de la TAD, Y. Chevallard écrivait ceci (Chevallard, 2007b).

Let me first say a few words about the *origins* of the field. The first argument that urged us to create a new field of study was a dual argument. On the one hand, we dismissed psychology, as Piaget had made it popular at the time, as a very partially relevant approach to the problems of teaching and learning. The reason was essentially the following – of course this is *my* version of the story. What a pupil would do, or fail to do, or even resist or elude is not determined only by some inner properties of his (...). In his lecture series *The Sciences of the Artificial* (1969), Herbert A. Simon made an illuminating statement: “An ant, he wrote, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself.” Simon then substituted “man” for “ant”, thus suggesting that man's behaviour was also largely a reflection of his environment.

À l'image de la fourmi qui chemine sur le sol de façon étrangement complexe pour l'observateur humain (parce que le sol est, à son échelle de fourmi, très tourmenté), on peut ajouter cette autre image : une personne est lâchée dans un labyrinthe ; son comportement ne peut guère s'expliquer si on méconnaît la structure du labyrinthe – en particulier, ce comportement ne peut s'expliquer seulement par les lois autonomes supposées de la personne regardée comme un “behaving system”. Au lieu de parler seulement, de façon générique, de *personnes* ou – plus précisément – d'*usagers* de l'Internet, nous avons mentionné plus haut,

comme il est d'usage, la « qualité », définie par les institutions où on les observe, des personnes dont nous commentons l'équipement praxéologique en matière de RII : des étudiantes d'une université d'élite, une habituée d'une bibliothèque universitaire, des élèves de *high school* suivant un cours d'horticulture, un professeur de technologie enseignant en 3^e, un professeur d'histoire et de géographie responsable d'un site pédagogique connu. Mais ces descriptions « classiques » ne doivent pas tromper : d'un côté, elles ne nous disent pas tout, loin de là, sur le paquet des contraintes dont la personne considérée est porteuse ; d'un autre côté, elles risquent fort de nous conduire à oublier qu'il est bien d'autres contraintes à prendre en compte pour analyser la conduite de cette personne. On peut observer un usager s'affairant avec des mots clés et un moteur de recherche (Ladage, 2007a) ; mais si, comme il en va souvent, cet usager ignore l'usage (et jusqu'à l'existence même) des *opérateurs booléens*, c'est d'abord cela que, en nombre de cas, la praxéologie qu'il se sera bricolée (ou qu'il sera en train de se bricoler) manifestera (et nous révélera), en même temps que certains assujettissements allogènes de l'usager (qui pourra par exemple poser au moteur de recherche une question complète, ainsi qu'il le ferait avec un interlocuteur dans la vie courante, ou recopier une partie de la consigne écrite proposée par la maîtresse s'il s'agit d'un jeune élève, etc.), et non telle ou telle des propriétés intrinsèques supposées de cet usager.

Nous serons amenée, dans ce qui suit, à nous éloigner, théoriquement mais aussi pratiquement, du postulat selon lequel les « systèmes » étudiés seraient « auto-explicatifs ». Mais nous devons nous démarquer maintenant d'une grande problématique de recherche en didactique, celle de nombre de recherches qui répondent au schéma de question suivant : « Étant donné les contraintes pesant actuellement sur elle, sous quelles conditions – existantes ou à créer – une institution ou une personne donnée peut-elle intégrer à son équipement praxéologique telle organisation praxéologique *désignée* ? » Il s'agit là du questionnement le plus classique, tout au moins quand on l'applique à cette institution qu'est une classe scolaire et à ces personnes que sont les élèves d'une classe. En nous référant à l'échelle des conditions et des contraintes présentée plus haut, nous pouvons en effet avancer que, selon un schéma d'intervention classique, lorsque, en une civilisation et une société données, une école a été établie, une pédagogie réglée *et un programme d'enseignement fixé*, et seulement alors, le didacticien « classique » est censé entrer en scène pour démêler, sous des contraintes auxquelles il n'est pas censé toucher, les conditions « proches » – en termes de situations didactiques –, qui permettront ou faciliteront une certaine diffusion dans la classe des praxéologies que le programme désigne. À titre instrumental ou auxiliaire, certains des travaux conduits dans une telle problématique se donnent souvent pour objectif, partiel ou non, de mieux connaître quelques-uns des objets qui y interviennent : ainsi étudiera-t-on les

praxéologies désignées comme « à diffuser », leur composition, leurs cohérence, leur complétude, leur degré d'intégration, etc. De même pourra-t-on examiner leur transposition (dans les manuels, les classes, etc.) sous les contraintes imposées, étude qui devrait se prolonger par l'examen des effets transpositifs des conditions envisagées, que celles-ci soient établies dans l'institution observée ou que le didacticien imagine simplement de les y voir établies. On s'attachera encore, classiquement, et comme nous l'avons vu plus haut, à explorer plus ou moins finement l'équipement praxéologique préalable des acteurs de l'institution où la diffusion des praxéologies désignées doit se faire, c'est-à-dire à mettre au jour certaines des contraintes sous lesquelles ladite diffusion devrait s'opérer. Dans le domaine qui nous occupe ici, on pourrait citer à cet égard un nombre croissant de travaux, parmi lesquels nous mentionnerons ici, car par beaucoup d'aspects exemplaires de la problématique classique, la thèse pionnière de P. Brandt-Pomares (2003) et, d'ampleur très différente sans doute, le travail plus récent de N. Gauducheau, F. Cuisinier et C. Garitte (2008).

On ne peut toutefois ignorer que, peut-être parce que – implicitement ou non – leurs auteurs « prêtent trop » à l'élève, une part non négligeable des travaux de type didactique consacrés à la RII s'en tiennent à des descriptions larges, qui n'évoquent que de manière bien imprécise les praxéologies à enseigner ou qui apparaîtraient comme des candidates sérieuses à l'être au terme de l'étude réalisée. À titre d'illustration, voici la présentation résumée, par l'auteure elle-même, des résultats d'un article choisi parmi beaucoup d'autres (Enochsson, 2005) : “In my research I have identified six different skills, which the students point to as fundamental. The youngest students focus only on (1) language and (2) knowledge of technology, while the oldest students regard these skills as obvious and do not always mention them. These can be seen as basic skills, and are prerequisites for seeking information on the Internet. But there is always a need to develop them further. According to the students it is also important to have (3) knowledge of different ways of information seeking, and (4) how search engines work. These skills are very closely connected to the Internet. (5) Setting goals and (6) being critical, are skills that are also useful outside the Internet, and can also be seen as skills needed for understanding when not to use the Internet.” Par contraste, on peut imaginer, sur le modèle de ce qui existe à propos des disciplines scolaires établies (mathématiques, biologie, etc.), que des didacticiens s'efforcent de concevoir et de mettre au point des situations didactiques visant à provoquer l'apprentissage d'une praxéologie constituée autour du type de tâches suivant :

proposer à un moteur de recherche une requête demandant l’affichage des pages Web contenant le mot M_1 mais ne contenant ni le mot M_2 , ni le mot M_3 .

Pour introduire la seconde grande problématique de recherche en didactique, que nous ferons nôtre dans la suite de ce travail, il nous faut utiliser une notion clé : celle de *projet* – d’une institution ou d’une personne²⁸. Il pourra s’agir par exemple, pour une institution de formation professionnelle, du projet de créer en son sein une nouvelle filière de formation, qui conduise à des métiers anciens ou nouveaux ; pour une personne, du projet de lancer une société de création de sites Web dont elle sera d’abord l’unique salarié ; etc. Pour concevoir et mettre en œuvre un projet, un équipement praxéologique Π sera *utile*, et parfois indispensable.

Une question cruciale se pose alors : de quel équipement praxéologique Π a-t-on *besoin* ? De quelles praxéologies sera-t-il fait ? Cette question est au cœur de la didactique de l’enquête codisciplinaire. La question soulevée initialement – quelles sont les praxéologies *utiles* à la conception et à la réalisation du projet ? – y trouve un principe de réponse : ce sont les praxéologies qui se révéleront *utiles* à la fabrication de réponses R^\heartsuit aux questions Q que soulèvera la poursuite du projet. Tout cela se situe d’abord au niveau de l’économie didactique : décider d’étudier telle question Q , en venir à décider, un peu plus tard, de « se rendre disponible » – par l’étude – telle œuvre O sont autant de décisions qui relèvent de cette économie didactique²⁹. Pour comparer cette situation avec celle supposée dans la problématique classique, il faut imaginer que quelque institution s’efforce de proposer aux institutions et personnes « intéressées » une *formation à la conception et à la réalisation de projets du type considéré*. Surgit alors l’un des deux grands problèmes qui nous occuperons dans toute la suite : le problème de la détermination des contenus praxéologiques d’une formation donnée, précisée d’abord par son objet déclaré. Ainsi arrive-t-on à la problématique qui sera la nôtre ici, que Y. Chevallard qualifie de *postclassique* et que l’on peut formuler ainsi : « Étant donné les contraintes pesant sur elle et un ensemble de conditions existantes ou à créer, à quel équipement praxéologique est-il possible qu’une institution ou une personne donnée accède pour opérer au mieux dans un domaine d’activité donné ? »

Ce schéma de question appelle tout de suite un commentaire crucial : en règle générale, nombre d’acteurs des institutions de formation *ne se posent pas ce genre de question*, parce qu’ils se trouvent réellement ou se placent en imagination dans une position institutionnelle pour laquelle la réponse serait *réglée d’avance* – au double sens du terme – par des institutions surplombantes jugées compétentes – mécanisme qui, au reste, peut se répéter à des niveaux fort différents des hiérarchies institutionnelles. Pourtant, quand, comme

nous le faisons dans ce travail, on examine un tant soit peu ce que, pour tel domaine d'activité, pourrait être la réponse, on s'aperçoit que la situation est on ne peut plus complexe et difficile à clarifier.

Le schéma de question précédent doit bien sûr, pour ce qui nous concerne, être spécifié de la façon suivante : « Étant donné les contraintes pesant sur elle et un ensemble de conditions existantes ou à créer, à quel équipement praxéologique est-il possible qu'une institution ou une personne donnée accède pour opérer au mieux *dans le domaine de la RII* ? » Mais cette question elle-même se laisse scinder en deux sous-questions que nous examinerons chacune à son tour, quoique solidairement. La première d'entre elles est celle-ci : que sont les praxéologies qui pourraient être regardées comme *utiles* à la RII ? La seconde, indissociablement liée à la première, peut-être formulée ainsi : en regard de la nébuleuse des praxéologies éventuellement utiles, qu'est-ce qui, dans l'écologie sociale des praxéologies est susceptible d'expliquer la faiblesse actuelle de la diffusion de ces praxéologies et, corrélativement, de la culture praxéologique courante en matière de RII ? *C'est pour l'essentiel à examiner ces questions que sera consacré le travail que nous présentons ici.*

Ainsi qu'on l'a vu, élaborer une réponse R^\heartsuit à une question Q prend un sens complet en rapport avec un certain *projet* où la question Q est censée se poser, même si une large partie de la construction de R^\heartsuit est indépendante de ce projet (ce qui assurera sa *valeur* ultérieure de réponse R^\diamond par rapport à d'autres projets possibles). Même lorsque la construction de R^\heartsuit a pour habitat une institution scientifique, ainsi qu'il en va pour ce qui nous concerne ici, il n'y a aucune raison pour que le projet considéré soit lui-même de nature « scientifique ». En l'espèce, ce travail se réfère à un projet nécessairement collectif, qui n'est bien entendu qu'un horizon pour notre travail, et qui relève à nos yeux d'une politique d'*instruction publique* (au sein de l'Éducation nationale aussi bien qu'hors d'elle) : celui de travailler de façon ouverte et transparente, scientifiquement informée, à construire une culture praxéologique en matière de RII (et, plus largement, en matière de monde Informatique & Internet) susceptible d'éclairer et de nourrir les *pactes d'instruction* (Chevallard, 1998a, 1998b) qui pourront se nouer, en matière de RII, au sein des institutions les plus diverses.

Notons ici, pour prévenir toute méprise, que nous ne visons pas à définir les contenus d'une formation à la RII en telle ou telle institution, ainsi qu'il en irait peut-être en une étude curriculaire. Nous ambitionnons simplement, *en amont de telles études*, de contribuer à élucider les conditions et contraintes sous lesquelles semblable projet pourrait être mené à bien. Notre travail peut donc être lu comme une étude préparatoire à l'élaboration de curriculums de formation à la RII. C'est donc, si l'on veut, une étude *précurriculaire*, dont

l'ambition est de dégager un outil permettant d'interroger adéquatement tout projet comme toute réalité curriculaire en matière de RII.

Aspects méthodologiques

L'objet central de notre étude n'est ni une population de personnes, ni une population d'institutions, mais bien une population (aux contours nécessairement flous) *de praxéologies*, que nous interrogerons et sur sa composition, et sur les conditions et contraintes de ses interactions avec des personnes et des institutions, c'est-à-dire sur son écologie sociale. Cet ensemble Π de praxéologies est, idéalement, celui qu'on obtient en réunissant l'ensemble des équipements praxéologiques de toutes les personnes (y compris nous-même...) et de toutes les institutions, dans une certaine période de temps. Il est clair qu'un tel ensemble, que l'on peut *supposer*, n'est « observable » dans sa totalité qu'en un sens tout théorique – un peu à la manière dont est observable l'« univers observable » des cosmologistes. Mais des observations partielles en sont évidemment possibles et, comme nous l'avons indiqué plus haut, nous nous y sommes livrée à l'aide de différentes techniques. De plusieurs de ces observations, faites directement auprès de personnes ou institutions, seules des traces apparaissent ici : ainsi avons-nous été amenée à évoquer, dans la conclusion de la deuxième partie de ce travail, l'un des entretiens approfondis que nous avons eus avec une série d'interlocuteurs. Mais c'est une forme d'observation « oblique » que nous devons surtout mentionner, parce qu'elle donne son matériau à la *première partie* de ce travail : elle consiste à saisir les praxéologies d'une personne en matière de RII, et plus précisément à observer la place accordée à la RII dans le cours d'une enquête codisciplinaire déterminée, à travers le *récit* rédigé par l'enquêtrice elle-même – dans l'ignorance, peut-on penser, de l'usage « clinique » que nous en faisons ici.

L'observation et l'analyse de ce long texte à vrai dire exceptionnel constituent méthodologiquement une transition vers ce qui est la stratégie d'observation fondamentalement mise en œuvre dans ce travail : nous avons en effet, *pour l'essentiel*, travaillé sur ces praxéologies que l'on peut saisir dans des *textes de tous types*, parce qu'elles y sont décrites, évoquées, mises en œuvre ou parce qu'elles y apparaissent au contraire comme notoirement manquantes. À la population « théorique » Π des praxéologies nous avons ainsi substitué la sous-population de praxéologies présentes – d'une manière ou d'une autre – dans l'ensemble Σ des textes touchant à la RII ou, comme nous dirons, des *exposés* relatifs à la RII – question sur laquelle nous reviendrons dans la *deuxième partie* de notre travail. Bien entendu, ces textes ont des *auteurs*, connus ou anonymes, et qui sont assujettis,

dans leur rapport à la RII, à certaines institutions : à travers leurs productions, ce sont ainsi ces personnes et ces institutions dont nous observons, indirectement, l'équipement praxéologique. En substituant ainsi Σ à Π , nous avons tenu compte de deux faits indépendants l'un de l'autre (ils appartiennent à des registres bien distincts), mais se renforçant l'un l'autre. Tout d'abord, nous étudions ici non pas les praxéologies simplement *utilisées* mais bien les praxéologies *déclarées utiles* (de façon explicite ou implicite). Ensuite, du fait de l'existence du traitement de texte et du Web, une partie substantielle des praxéologies personnelles ou institutionnelles même les plus singulières est désormais mise par écrit et rendue publique.

La « clinique praxéologique » (Chevallard, 2008, séance 5), ici relative à Π , devient ainsi *clinique des exposés sur la RII*, relative à Σ . Bien entendu, l'ensemble Σ , quoique fini, est vaste et, dans sa partie « en ligne » du moins, changeant, en sorte que, du fait de l'état d'abandon, d'obsolescence ou d'incurie de certains sites Web, nous avons par exemple dû modifier plusieurs fois, au long de notre recherche, la webographie d'exposés en ligne sur la RII qui précisait notre corpus d'exposés. Cette clinique des textes sur la RII nous a amenée à prendre connaissance d'une foule d'exposés ; mais seuls un petit nombre d'entre eux, que nous espérons adéquatement choisis³⁰, feront l'objet ici d'une observation et d'une analyse détaillées. Étant donné l'objet de notre recherche, en effet, il ne pouvait être question de réduire les exposés consultés à de simples fiches signalétiques ; en sorte que, sauf à augmenter démesurément le volume de ce travail, nous ne pouvions envisager de présenter que quelques « cas cliniques » dont la représentativité d'ensemble vis-à-vis de Σ nous paraisse assurée.

Nous laisserons le lecteur découvrir les techniques d'analyse d'exposés que nous avons mises en œuvre, ainsi que les commentaires que nous en avons faits : banales mais indispensables en certains cas (dans la deuxième partie par exemple), elles se conjuguent avec des techniques beaucoup plus originales, croyons-nous, dans la quatrième partie. Mais nous voudrions revenir ici à la question des praxéologies *présument utiles*. Nous prendrons pour cela deux « exposés » brefs, centrés sur les services de Google, mais qui diffèrent par la langue ainsi que, si l'on peut dire, par la « niche » qu'ils occupent dans le système des médias. Le premier est extrait du livre de David A. Vise et Mark Malseed déjà mentionné : celui-ci comporte un Appendice I, intitulé “23 Google Search Tips” qui court sur cinq pages (pp. 293-297) : comme son titre l'indique, il se compose de 23 « trucs » (*tips*) dont nous reproduisons ci-après les intitulés.

1. Google can be your phone book / 2. Google can be your calculator / 3. Longer is better, but shorter is okay / 4. Use quotation marks when precision matters / 5. Google can be your dictionary /

6. Capitalization doesn't matter / 7. Forget pluralism / 8. Get the picture / 9. Maps, driving directions, and satellite views are one click away / 10. Where do you want to go today? / 11. Browse the world's bookshelves online / 12. Dial GOOGL when you're on the go / 13. I'm Feeling Lucky / 14. Google can be your newscaster / 15. Google can be your weatherman / 16. Become a researcher / 17. Become a scholar / 18. Take a magic ~ride / 19. Pack more results onto each page / 20. Translate into other languages / 21. Get an instant stock quote / 22. Get PG-rated results / 23. Peer inside Google

Considérons ainsi le *tip* n° 3, “Longer is better, but shorter is okay”; les auteurs le développent ainsi.

Google is designed to return high-quality results even for one- or two-word queries, so you can keep your searches short. But adding a few more words often yields better results. *Example:* While gathering information on applying to colleges, include the word *admissions* after the name of a university you are searching to get more relevant results.

Il est donc utile, selon ces auteurs, d’abord de savoir qu’une requête peut fort bien ne pas se limiter à un ou deux mots, ensuite de savoir exploiter cette possibilité pour amener aux premiers rangs des pages plus susceptibles de répondre à ces besoins. Le *tip* n° 4, “Use quotation marks when precision matters”, fait l’objet de ce commentaire.

Typing “*the search is over*” into Google will return Web pages about the rock song by Survivor—but leaving off the quotes will produce an assortment of unrelated pages. The reason: adding quote marks around a query tells Google to look for occurrences of the exact phrase as it was typed. That makes quote marks especially helpful when searching for song lyrics, people’s names, or expressions such as “*to be or not to be*” that include very common words.

À nouveau, on a là un savoir proposé comme utile. Voici maintenant le *tip* n° 18, “Take a magic ~ride”.

The tilde character “~” in the corner of your keyboard is a handy tool in Google searches. Put it before a word, with no space between, to have Google look for pages with both that term *and* its synonyms.

La même remarque s'impose. Le *tip* n° 16, "Become a researcher", préconise, on le verra, une culture de la RII déjà plus approfondie.

Google tends to list popular and fresh pages at the top of its results, but dig beyond the first page or two of search results and you will often find older, forgotten pages that have just what you need for a research project. Also check out the "Cached" versions of Web pages that Google collects as it crawls and downloads the Web, which are available through a blue link at the end of every search result. The cached version is an old version of the page; and often has the content you are seeking even if the current version of the page has changed—say, a news site that removed the original story. *Bonus tip:* The cached version also highlights your search terms in color wherever they appear on the page, an especially helpful feature when combing through long documents.

Ici, les auteurs évoquent l'utilité de la praxéologie indiquée *for a research project*. D'une façon générale, ils se réfèrent à l'utilité d'une praxéologie, si simplette soit-elle, *pour* un type de situations.

Voici maintenant le second exposé annoncé : il s'agit d'un billet signé de Diane Mottez et Alexia Guggémos pour *Femme Actuelle.fr*, mis en ligne le 11 août 2008 et intitulé *Google, mode d'emploi*. Après un court préambule, les deux auteures proposent à leurs lectrices supposées « tous [leurs] trucs et astuces pour devenir une vraie "Sherlock Holmes" de la Toile ! ». Ce sont ces trucs et astuces que nous reproduisons ici.

1. Utilisez les guillemets

Si votre recherche comporte plusieurs mots pouvant être associés les uns aux autres, utilisez les guillemets.

Par exemple, la saisie de Carla Sarkozy sans guillemet [*sic*] recherchera les résultats parmi les mots Carla ou Sarkozy. En revanche, la saisie de "Carla Sarkozy" (avec guillemets) provoquera une recherche avec les deux mots demandés, obligatoirement l'un à côté de l'autre. Il s'agit alors de la syntaxe la plus efficace.

2. Attention à l'ordre des mots !

Vous comprendrez aisément que "JO 2008" ne donne pas le même résultat que "2008 JO".

3. Choisissez les mots-clés

La qualité des réponses dépend avant tout de la qualité des mots choisis. Par exemple, "forme" ne donne pas les mêmes résultats que "bien-être" : le premier peut vous orienter vers des formes géométriques plutôt que des exercices de yoga.

4. Utilisez les "+" et les "-"

Ces signes ajoutent des contraintes à votre recherche. Par exemple, vous pouvez rechercher des documents sur Vincent Perez qui ne parlent que de Fanfan la tulipe (Vincent Perez + fanfan la tulipe) ou au contraire qui ne parlent pas du film (Vincent Perez - fanfan la tulipe).

5. Installez la barre de recherche

La barre d'outils Google se télécharge et s'installe dans son [sic] navigateur. Elle intègre un filtre anti pop-up, un vérificateur d'orthographe, un traducteur mot à mot et permet de limiter la recherche aux pages d'un pays. Pratique pour gagner du temps.

6. Faites une recherche en images

Avec le logiciel Girafa, visualisez en un coup d'œil l'ensemble des vignettes correspondant aux requêtes. Vous chercher [sic] "MP3" sur Google et vous avez [sic] un jour repéré un excellent site sur le sujet, vous vous souviendrez alors plus de l'apparence de la page mais pas de son url. Idéal quand on surfe beaucoup.

7. Exploitez les forums de discussion

Baptisé Google Groups, ce nouveau service de Google abrite plus de 700 millions de messages postés sur les forums de discussion. Il est possible qu'un internaute ait posé la même question que vous quelques jours avant. Par exemple, sur un mode de garde ou un problème technique sur votre ordinateur. Chaque réponse a été archivée. Ces forums sont une vraie mine d'or !

<http://groups.google.fr>

Ces deux « exposés » présentent, de par leur nature même, ce que ceux qui les ont signés ont jugé à un moment donné, dans des circonstances données, à l'endroit d'un certain public, *utiles* en matière de RII. On peut penser qu'il s'agit, dans les deux cas, et notamment dans le second, d'un choix parmi une liste plus vaste : en ce sens, il s'agit là des praxéologies jugées *les plus utiles* par qui les proposent. Ce qui importe, cependant, tient pour nous dans le fait suivant, sur lequel nous allons revenir : ces exposés ne sont pas des déclarations que nous aurions recueillies « frontalement » auprès de leurs signataires, par exemple dans le cadre d'une enquête par questionnaire ou par entretiens. En ce sens, il s'agit d'observations *obliques* (et non pas *frontales*), non influencées par l'enquête que nous menons – même si elles le sont par d'autres conditions et contraintes. On aura observé ainsi que le texte de D. Mottez et A. Guggémos comporte de nombreuses imperfections formelles qu'un travail d'édition plus soigné (tel celui qui a cours sans doute dans la version imprimée du magazine) aurait pu éliminer. De façon générale, les aspects secondaires de leur production, ceux qui ne sont pas clairement des enjeux didactiques – ou plutôt, ici, hypodidactiques – semblent moins soumis à

l'infléchissement que les circonstances peuvent apporter à l'exposé. Dans un court préambule, les auteures écrivent ainsi : « Google référencent [*sic*] aujourd'hui 1000 milliards de pages Web. » Sans doute se prévalent-elles ici d'un *post* récent – il est daté du 25 juillet 2008 – de deux ingénieurs de Google (Alpert & Hajaj, 2008) dans lequel ceux-ci écrivent : “Recently, [...] our systems that process links on the web to find new content hit a milestone: 1 trillion (as in 1,000,000,000,000) unique URLs on the web at once!”. Derrière le sens de l'actualité de ces journalistes, qui se montrent ainsi en phase avec les nouveautés du monde Internet, on voit cependant transparaître un rapport fort incertain aux notions d'index et de référencement, alors même que les auteurs du *post* prennent soin d'indiquer : “We don't index every one of those trillion pages [...] But we're proud to have the most comprehensive index of any search engine, and our goal always has been to index all the world's data.” Ils prennent soin aussi, il est juste de le souligner, de ne rien dire exactement de la taille de l'index de Google ! Mais cet index ne référence certainement pas un milliard de pages. Alors que l'on pourrait voir une intention derrière le fait de mentionner l'existence d'au moins un milliard d'URL, il n'y en a vraisemblablement aucune dans la confusion entre ce qui est « crawlé » et ce qui est « indexé ». Quoique les auteures aient une familiarité évidente avec le monde Internet, on aperçoit ici une lacune de leur culture de la RII.

Pour dégager la matière d'un débat et d'un travail à venir – les praxéologies que l'on peut estimer utiles à la RII –, donc, nous nous en sommes tenue à interroger des personnes par le truchement de leurs productions écrites, personnes particulières qui, soit de leur propre chef (comme dans le cas de David Vise et Mark Malseed, sans doute), soit en tant que sujets d'une institution qui prescrit ou, du moins, qui accorde son investiture, comme dans le cas de Diane Mottez et Alexia Guggémos), se posent comme *auteurs*. Pourquoi n'avoir pas interrogé aussi, de façon directe, frontale, des personnes « quelconques », que seul le questionnement à elles adressé change en auteurs d'un moment ? Pour deux raisons au moins. La première touche spécifiquement à ce qui aurait été alors le thème de l'enquête, les savoirs (utiles) en matière de RII : les entretiens que nous avons conduits, les interactions plus informelles que nous n'avons pas cessé d'avoir au long de cette recherche nous ont montré la faiblesse actuelle de la culture de la RII un tant soit peu diffusée, en sorte que nous avons rapidement abandonné une idée dont les fruits nous semblaient ne pas devoir tenir la promesse des fleurs. La seconde raison a déjà été suggérée : elle a trait au contrôle des effets engendrés par la situation d'interaction elle-même. La chose est connue et l'on peut à son propos se référer à l'article ancien mais fondamental de Martin T. Orne (1962) intitulé “On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications”. Soumis aux contraintes et aux conditions de la situation « expérimentale », les

personnes ayant accepté de « participer à l'expérience » vont momentanément endosser un des quelques rôles typiques que la situation engendre, dont, principalement, celui du “good experimental subject”, à propos duquel Orne observe notamment ceci.

A volunteer who participates in a psychological experiment may do so for a wide variety of reasons ranging from the need to fulfill a course requirement, to the need for money, to the unvoiced hope of altering his personal adjustment for the better, etc. Over and above these motives, however, college students tend to share (with the experimenter) the hope and expectation that the study in which they are participating will in some material way contribute to science and perhaps ultimately to human welfare in general. (p. 778)

Lorsqu'on s'adresse à un *groupe* de personnes – de « sujets » – constituant un échantillon « disponible » – c'est-à-dire, dans les institutions de formation, constitué le plus souvent d'étudiants, voire de formateurs³¹ –, ce que Orne nomme les *demand characteristics*, les « caractéristiques de la demande », vont souvent porter les participants à adopter l'un ou l'autre d'un petit nombre de rôles « normalisés » : il y a ainsi, à côté du *good participant role*, son opposé, le *bad participant role*, mais aussi le *faithful participant role*, qui conduit à suivre les instructions à la lettre, et encore l'*apprehensive participant role*, dont le titulaire est obsédé par l'idée positive ou négative que, croit-il, l'expérimentateur ne manquera pas de se faire de lui... Ces phénomènes, qui peuvent être atténués par une « demande » soigneusement conçue et réalisée, ne sauraient être éliminés entièrement. Une manière somme toute classique de le faire consiste, lorsque la chose est possible, à augmenter la taille du groupe pour que les *régularités statistiques* recouvrent les idiosyncrasies comportementales individuelles et permettent que, au-delà du « bruit » engendré par les caractéristiques de la demande, le *signal* émerge, *quand signal il y a* – ce qu'on ne peut savoir, en règle générale, qu'après coup, en analysant les données recueillies. De cela, nous donnons un exemple typique dans l'annexe 1 à cette *Introduction générale*, à propos de la « péjoration de l'Internet » : on verra comment l'emploi de la technique du *différenciateur sémantique* (*semantic differential*) révèle la présence, dans l'intimité des sentiments, de cette contrainte imposée par la culture dominante actuelle.

Dans un cas important, auquel nous avons consacré toute notre *troisième partie*, nous avons *nous-même* tenté de dégager de textes multiples touchant à des domaines divers, sur la base d'une analyse chaque fois spécifique, des praxéologies qui pourraient être prises en compte dans le travail collectif de construction d'une culture de la RII telle que nous l'avons évoquée plus haut. Il s'agit là de *notre contribution propre la plus importante* à un projet dont

il nous a paru ainsi indispensable d'ébaucher certains contours que les auteurs dont nous avons par ailleurs sollicité les « avis » nous semblaient avoir moins nettement dessinés. Mais c'est dans la *quatrième partie* de notre travail que, après l'annexe à cette *Introduction générale*, l'on retrouvera le recours direct, frontal le plus significatif à des personnes (autres que nous-même !), selon un dispositif « expérimental » qui sollicite les « sujets » selon des modalités que nous présenterons et commenterons alors.

Il resterait à dire un mot sur un sujet qui agite beaucoup certains esprits lucides, chagrins ou inquiets : le recours à l'encyclopédie libre *Wikipedia* (en anglais) ou *Wikipédia* (en français), que nous avons nous-même beaucoup sollicitée, sans que cela se voie toujours nettement. Personne n'ignore que des critiques ont été élevées à son encontre : nous y reviendrons. Mais deux points peuvent être d'ores et déjà précisés. Tout d'abord, il n'est pas anormal, dans un tel cas, d'entendre la défense, et en l'espèce de se reporter par exemple aux articles "Criticism of Wikipedia" et "Reliability of Wikipedia" de cette même encyclopédie *Wikipedia* ainsi que quelques autres textes qu'on y trouvera mentionnés. Ensuite et surtout, aucun auteur ne saurait s'abriter derrière l'autorité supposée d'aucune source lorsqu'il avance une assertion quelle qu'elle soit : en aucun cas sa responsabilité d'auteur ne saurait être diminuée du fait qu'il cite une source que d'aucuns regarderaient comme sûre ou plus sûre que telle autre. Une source ne saurait être une preuve : la dialectique des médias et des milieux retrouve ici sa place pleine et entière.

Note sur le « style »

Le mot de *style* est, en français, souvent réservé au style *d'écriture*, le *writing style* des auteurs de langue anglaise. À cet égard, on ne s'étonnera pas de ce que, ayant suivi depuis plusieurs années le séminaire d'Yves Chevallard, qui a en outre bien voulu diriger ce travail, nous nous soyons imprégnée non seulement des manières de penser propres à la TAD (ce qui était le but recherché), mais aussi de beaucoup de ses manières de dire, du style écrit ou oral dans lequel la conceptualisation anthropologique en didactique reste très largement formulée. Mais il est une autre notion de style, qui reste encore largement confinée dans la langue anglaise, comme lorsqu'on parle du « style » de l'APA : cette notion de style éditorial, *editorial style*, inclut alors non seulement l'orthographe, la ponctuation, la typographie, mais encore une foule d'aspects de l'édition d'un texte auxquelles nous avons tenté – sans y réussir complètement ! – d'être attentive. Nous ne ferons ici, à cet égard, que quelques remarques sommaires.

Notre texte comporte beaucoup de citations et de documents divers qui constituent le matériau essentiel de notre travail. Nous n'avons donc pas choisi de les regrouper dans un volume d'annexes, même si nous avons désigné comme annexes, à la fin de certaines des parties en lesquelles se scinde ce volume, des développements qui nous ont paru utiles, mais par trop susceptibles de rompre la dynamique du texte principal s'ils y avaient été intégrés. Beaucoup de ces citations, en outre, sont en anglais. Nous n'en avons pas donné de traduction, considérant qu'il s'agit là d'une langue de travail académique qui est le bien commun de tous les chercheurs (et pas seulement des anglophones « natifs »). Du point de vue typographique, nous avons respecté, sauf erreur, les règles propres à chaque langue (pour le deux-points, le point-virgule, le point d'exclamation ou d'interrogation notamment), en faisant en outre le choix d'utiliser les guillemets anglais pour les citations en langue anglaise, et les chevrons pour les citations en français (avec les règles classiques gouvernant dans chaque langue les citations dans une citation). Pour les citations de références (dans le texte ou dans les notes) et pour les références elles-mêmes (en fin de volume), nous avons utilisé des normes en cours d'élaboration, celle d'un « style TAD » à venir, qui visent modestement à démarquer celles de l'APA. Nous avons, ici, adopté l'orthographe *Web*, avec une majuscule à l'initiale, tout en respectant l'orthographe adoptée par les auteurs que nous citons, et en parlant de *webographie* (sans majuscule) plutôt que de « sitographie » (suivant en cela les recommandations de l'Office québécois de la langue française). La même remarque s'applique à propos de *l'Internet*, que nous avons fait précéder de l'article défini et écrit avec une majuscule à l'initiale. S'agissant des documents en ligne, nous avons rendu *retrieved*, *retrievable*, par « consulté », « consultable », et *available* par « disponible ». Comme on l'aura remarqué déjà, nous avons parlé quelquefois de *savoirs* (plus rarement de *savoir-faire*) là où la TAD demanderait *praxéologies* : ce léger abus de langage, sans conséquence dès lors que l'on place son propos sous le contrôle de la TAD, nous a paru utile en certains cas pour aller plus directement à notre objet.

ANNEXE 1

La péjoration de l'Internet à l'épreuve du différenciateur sémantique

Ce que nous avons appelé la « péjoration » de l'Internet se traduit-elle, dans le rapport des personnes à cet objet largement marqué par la modernité, de façon suffisamment nette pour émerger de la variété des histoires et des postures à l'endroit de l'Internet ? Pour apporter réponse (partiellement) à cette question, nous utiliserons ici la technique du *différenciateur sémantique* (*semantic differential*), qui permet d'accéder à ce qui paraît être l'intimité singulière des rapports personnels. Cette technique, due au psychologue Charles E. Osgood (1916-1991), a en effet pour objet d'évaluer "the connotative meaning of individual words" (Reber, 1985, p. 681) ; pour cela, "a subject rates the meaning of each word along a number of polar dimensions such as hot-cold, weak-strong, tense-relaxed, rough-smooth, etc.". Au cours de l'année universitaire 2007-2008, nous avons ainsi demandé à divers groupes d'étudiants en sciences de l'éducation de l'université de Provence de bien vouloir participer, en marge de leur formation, à une telle exploration. La procédure était la suivante : une liste d'échelles de 0 à 6, telles les échelles laid/beau ou froid/chaud, leur était présentée, avec un mot servant d'exemple pour indiquer ce qu'il leur serait demandé de faire ; puis une suite de mots leur était proposée au fur et à mesure, chaque participant disposant pour chaque mot de cette suite d'une fiche qu'il devait remplir (en assignant un score au mot sur chacune des échelles qui y était portée), avant qu'on n'annonce le mot suivant. La liste des échelles et la liste des mots ont un peu fluctué au fil des groupes sollicités. En particulier, les quinze échelles proposées au premier groupe ont été ramenées à onze puis à dix, pour limiter la durée des passations. Les résultats des divers groupes sont présentés ci-après dans l'ordre chronologique des passations.

Nous avons d'abord demandé à un *premier groupe* de 22 étudiants de la licence professionnelle de sciences de l'éducation – dont le lieu d'étude est le site de Lambesc de l'université de Provence – d'attribuer à un certain nombre de mots, dont « Internet », un score de 0 à 6 sur quinze échelles, parmi lesquelles nous ne retenons ici que les onze suivantes : superficiel/profond (S/P), froid/chaud (F/C), peu utile/utile (PU/U), irritant/apaisant (I/A), peu fiable/fiable (PF/F), abstrait/concret (A/C), inorganisé/organisé (I/O), laid/beau (L/B),

passif/actif (P/A), artificiel/naturel (A/N), facile/difficile (F/D). Voici les scores, totalisés sur l'ensemble des 22 « juges », obtenus par « Internet » sur ces onze échelles : ces scores sont exprimés en pourcentage du score maximal, égal en cas à $22 \times 6 = 132$.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	A/N	F/D
Score	38	41	82	43	70	56	52	71	61	25	34

On voit que le score du mot « Internet » sur l'échelle laid/beau (L/B), soit 52/100, se situe presque exactement à mi-chemin entre le minimum théorique et le maximum théorique : à juger selon ce critère numérique, « Internet » ne serait ni laid ni beau pour notre groupe de juges – même si l'un d'eux lui attribue le score maximal, 6, le trouvant donc « très beau ». À l'évidence, « Internet » est jugé *utile* : sur l'échelle peu utile/utile (PU/U), le plus *bas* score qu'il obtient est 3 (la « moyenne » sur une échelle de 0 à 6), qui lui est attribué deux fois. Le score sur l'échelle A/N (artificiel/naturel), soit 25/100, est le plus bas obtenu sur *l'ensemble des mots* proposés à ce groupe : « Internet » est ainsi *fortement artificiel*.

On peut comparer les scores obtenus par « Internet » à ceux obtenus par « Web », puisque ce mot était également proposé à l'appréciation des 22 étudiants. Sur le tableau suivant, la deuxième ligne rappelle les scores totaux en pourcentage de « Internet » ; la troisième présente ceux relatifs à « Web ».

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	A/N	F/D
Internet	38	41	82	43	70	56	52	71	61	25	34
Web	42	39	80	46	65	59	54	63	67	26	42

Les deux mots ont des profils voisins, ce qui n'est guère étonnant. Sont-ils les seuls à avoir ce type de profil ? On trouvera ci-après le tableau précédent complété par les lignes relatives aux autres mots proposés aux 22 juges, à savoir, après « Internet » et « Web », « statistique », « didactique », « université », « éducation », « Lambesc ». Dans plusieurs des colonnes de ce tableau s'observe un changement lorsqu'on passe du bloc des trois premières lignes (Internet, Web, statistique) à celui des quatre dernières (didactique, université, éducation, Lambesc). Ainsi, sur l'échelle superficiel/profond (S/P), on découvre que « Internet » et « Web » sont plutôt *superficiels*, que c'est moins vrai pour « statistique », mais que « didactique », « université », « éducation » et « Lambesc » sont regardés bien davantage comme *profonds*.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	A/N	F/D
Internet	38	41	82	43	70	56	52	71	61	25	34
Web	42	39	80	46	65	59	54	63	67	26	42
Statistique	55	29	72	36	61	58	42	64	81	35	61
Didactique	76	61	80	55	77	67	59	67	69	54	63
Université	82	57	85	51	81	60	67	80	73	55	74
Éducation	87	72	93	52	80	69	77	86	64	61	71
Lambesc	83	64	89	58	79	64	77	82	67	60	64

Un schéma analogue s'observe à propos de l'échelle froid/chaud (F/C) : « statistique » surtout, suivi par « Internet » et « Web », sont *froids*, tandis que « didactique », « université », « éducation » et « Lambesc » sont plutôt chauds. Sur l'échelle peu utile/utile (PU/U), tous les mots sont rangés franchement du côté de l'*utile*, y compris « statistique », le mot « éducation » recevant le score le plus élevé. La moindre appréciation de « statistique » s'exprime encore sur l'échelle irritant/apaisant (I/A), où il est le moins bien noté. Mais ici, en vérité, aucun mot n'a de score impressionnant puisque le maximum est atteint par « Lambesc », dont le score ne représente pourtant que 58 % du score maximal. On saisit ainsi concrètement ce que nous avons appelé plus haut la « péjoration » de l'Internet – en découvrant, si on ne le savait pas déjà, que la statistique est encore plus mal lotie !

La vision précédente est-elle propre au groupe des 22 étudiants de Lambesc ? Ou au contraire est-elle un invariant d'une certaine « culture étudiante », pour ne pas parler de la culture, tout court ? Un formateur qui doit travailler avec un groupe « pris au hasard » parmi les groupes d'étudiants en sciences de l'éducation que l'administration de l'université crée doit-il s'attendre à retrouver la tonalité que fait entendre le groupe examiné jusqu'ici ? Pour tenter de répondre, nous examinerons d'abord les scores attribués par un *deuxième groupe* composé de 10 étudiants de première année du master de recherche en sciences de l'éducation, dont le lieu d'étude est le site d'Aix-en-Provence de l'université de Provence. Un groupe si réduit permet-il d'entendre le même « signal » par delà le « bruit » ?

Dans ce cas, les mots proposés étaient (dans cet ordre) « université », « statistique », « éducation », « Internet », « mémoire (le) » ; quant aux échelles, elles n'étaient plus que 10, l'échelle artificiel/naturel (A/N) ayant été écartée. Comme précédemment, les scores obtenus ont été exprimés en pourcentage du score maximal.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	F/D
Université	75	57	85	52	75	55	55	78	53	67
Statistique	57	45	78	47	62	63	60	52	83	50
Éducation	92	80	92	65	65	70	78	90	65	82
Internet	55	58	82	58	60	70	60	62	60	38
Mémoire	68	72	80	45	65	62	62	83	77	73

Pour « lire » ces résultats, on peut classer les mots sur chacune des échelles ; on obtient ceci.

➔ **Superficiel/profond**

Internet (55) < statistique (57) < mémoire (68) < université (75) < éducation (92)

➔ **Froid/chaud**

statistique (45) < université (57) < **Internet** (58) < mémoire (72) < éducation (80)

➔ **Peu utile/utile**

statistique (78) < mémoire (80) < **Internet** (82) < université (85) < éducation (92)

Irritant/apaisant

➔ mémoire (45) < statistique (47) < université (52) < **Internet** (58) < éducation (65)

Peu fiable/fiable

➔ **Internet** (60) < statistique (62) < éducation (65) ~ mémoire (65) < université (75)

Abstrait/concret

➔ université (55) < mémoire (62) < statistique (63) < éducation (70) ~ **Internet** (70)

➔ **Laid/beau**

université (55) < statistique (60) ~ **Internet** (60) < mémoire (77) < éducation (78)

➔ **Passif/actif**

statistique (52) < **Internet** (62) < université (78) < mémoire (83) < éducation (90)

➔ **Inorganisé/organisé**

université (53) < **Internet** (60) < éducation (65) < mémoire (77) < statistique (83)

➔ **Facile/difficile**

Internet (38) < statistique (50) < université (67) < mémoire (73) < éducation (82)

Le mot qui a sans doute la meilleure image est « éducation » : il obtient 92 % du maximum sur deux échelles : superficiel/profond et peu utile/utile ; mais il obtient encore 90 % du maximum sur l'échelle passif/actif. Son portrait : *très profond*, *très utile*, *très actif*, mais *difficile* aussi (82 %), *chaud* (80 %), de loin le plus *beau* de tous les mots proposés (78 %),

loin devant ses suivants (dont le score reste inférieur ou égal à 62 % du maximum), et aussi *le plus concret* (avec « Internet »), et le *moins irritant*. Seule ombre au tableau : il n'est que moyennement *fiable* et moyennement *organisé*... Pour « université », le tableau est plus contrasté. Quant à *l'utilité*, « université » obtient un score de 85 % et n'est « battu » que par « éducation ». Mais il est vrai que les cinq mots proposés sont tous trouvés des plus *utiles*, le plus faible score, celui de « statistique », étant tout de même de 78 % : on retrouve ici *la forte adhésion d'ensemble aux études poursuivies*. Avec ce même score exactement, « université » est jugé *actif*, même si « éducation » le surclasse à cet égard. « Université » se rapproche encore de « éducation » sur l'échelle superficiel/profond : il est jugé à 75 % *profond* ; il obtient le même score sur l'échelle peu fiable/fiable, se distinguant en cela de « éducation », jugé moyennement fiable. Bien que *fiable*, donc, « université » est en revanche nettement *inorganisé* ; mais il est vrai que « éducation » ne brille guère plus à cet égard. Il est en outre une échelle sur laquelle « université » et « éducation » sont aux antipodes l'un de l'autre : alors que « éducation » est le plus *beau* de tous les mots, on l'a vu, « université » est *le moins beau* – il est remarquable qu'il soit même *plus laid* que « statistique » ! Sur deux autres échelles encore, « université » a des performances décevantes : il n'est pas très *chaud* (même s'il l'est un peu plus que « statistique »), et surtout il est le plus *abstrait* de tous les mots. Il occupe une position médiane sur l'échelle irritant/apaisant (où les scores sont resserrés entre 45 % et 65 % du maximum) ainsi que sur l'échelle facile/difficile, tout en étant sensiblement plus difficile que « Internet » et surtout que « statistique ».

Le mot « mémoire » a été proposé dans une période sensible, durant laquelle les « enquêtés » avaient le souci de mener à bien le mémoire de première année de master qui leur était demandé. On ne s'étonnera guère, en conséquence, que « mémoire » soit tenu pour le plus *irritant* des mots proposés, au même niveau que « statistique ». De même, s'il est bien entendu *utile* (avec un score de 80 %), il ne l'est guère plus que « statistique ». Il est trouvé relativement *abstrait*, même s'il l'est moins que « université ». Il est assez *profond*, mais moins que « éducation » ou même que « université ». Plus positivement, il occupe le *second rang* derrière « éducation » du point de vue de *l'activité*, de la *difficulté*, de la *chaleur*, de la *beauté*, derrière « université » pour la *fiabilité*, derrière « statistique » pour ce qui est, enfin, du niveau d'*organisation*.

Les mots « statistique » et « Internet » n'appartiennent pas, à l'évidence, à la même famille que « éducation », « université », « mémoire ». Ce sont, grosso modo, d'assez « mauvais objets » : ainsi sont-ils respectivement *dernier et avant-dernier* sur les échelles superficiel/profond, peu fiable/fiable, passif/actif, facile/difficile : *superficiels, peu fiables, passifs*, et, coup de pied de l'âne, *faciles* ! Le portrait n'est pas flatteur. Bien entendu, l'un et

l'autre sont *utiles*. Ils ne se séparent guère non plus sur l'échelle laid/beau, où ils se positionnent entre les deux extrêmes – ils sont toutefois nettement moins beaux que « éducation ». Cela dit, « Internet » l'emporte sur « statistique » sur l'échelle froid/chaud (« statistique » est sensiblement *plus froid*), sur l'échelle irritant/apaisant (c'est « statistique » qui irrite le plus, avec « mémoire »), sur l'échelle abstrait/concret (« Internet » est le plus concret de tous les mots proposés). Une seule inversion, qui n'est véritablement à l'avantage d'aucun des deux mots : « statistique » est très *organisé* (son score est de 83 %), « Internet » l'est beaucoup moins (60 %). Si le meilleur des mots est « éducation », le pire est donc « statistique ».

Un *troisième groupe* comprenant 11 étudiants de deuxième année du master professionnel suivant l'option « Ingénierie de formation par alternance » a été ensuite sollicité. Ces étudiants ont attribué des scores à sept mots : examens, statistique, éducation, Internet, Lambesc, didactique, université. Le score maximal sur chaque échelle était ici égal à $6 \times 11 = 66$, la « moyenne » étant donc de 33. Nous ne comparerons ici que les scores obtenus d'une part par « éducation », d'autre part par « Internet ». Les voici.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	F/D
Internet	44	38	80	42	56	56	53	62	67	44
Éducation	86	76	92	52	70	61	77	82	64	80

Comme on le voit, « Internet » n'est pas bien haut dans l'échelle des valeurs ! Il est peu profond, en tout cas nettement moins que « éducation ». Ce dernier mot obtient en outre un score *double* du sien sur l'échelle froid/chaud : « Internet » est froid, « éducation » est chaud. (Parmi les mots proposés, seuls « examens », avec un score de 30 %, et surtout « statistique », avec un score de 20 %, sont *plus froids* que « Internet ».)

Sur beaucoup d'autres échelles, les deux mots se distinguent moins. Toutefois, « éducation » est beau (77 %), quand « Internet » ne l'est guère (53 %), même s'il l'est plus que « statistique » (36 %). Sur l'échelle facile/difficile, également, les choses sont claires : « éducation » est difficile, et « Internet » est facile. À nouveau, on retrouve ici la péjoration de « Internet », même si « Internet » est trouvé utile (avec un score de 80 %), sans pouvoir rivaliser pourtant sur ce point avec « Éducation » (qui obtient un score de 92 %).

Un *quatrième groupe* composé également de 11 étudiants, mais de licence de sciences de l'éducation cette fois, et travaillant à Forcalquier, a eu à situer, sur les dix échelles déjà

utilisées, les cinq mots suivant (dans cet ordre) : Forcalquier, éducation, Internet, université, statistique. Voici les scores obtenus par ces différents mots.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	F/D
Forcalquier	52	48	62	56	55	77	88	64	56	58
Éducation	95	77	97	55	61	62	68	91	79	83
Internet	56	36	83	47	50	61	59	74	58	48
Université	67	41	79	52	71	56	55	58	61	79
Statistique	67	27	73	38	67	61	41	67	83	59

L'examen de ces scores fait retrouver des conclusions déjà rencontrées, avec quelques nuances. Ainsi, de manière un peu inhabituelle s'agissant du lieu de formation, « Forcalquier » reçoit-il des scores mitigés : s'il obtient certes le maximum sur l'échelle laid/beau (88 %), il n'est jugé ni vraiment profond (52 %), ni vraiment chaud (48 %), et pas beaucoup plus apaisant ou organisé (56 %). Ce tableau semble désigner un problème dans le rapport des étudiants à l'environnement de la formation suivie. On retrouve un écho de ce sentiment à propos de « université » : si le mot est jugé plutôt profond (67 %), utile et difficile (79 %), il est jugé froid (41 %), peu apaisant (52 %), et assez dépourvu de beauté (58 %).

Pour le reste (éducation, Internet, statistique), on retrouve des traits bien installés, quoique toujours avec quelques nuances. « Éducation » est facilement premier pour la profondeur (95 %), pour la chaleur (77 %), pour l'utilité plus encore (97 %), et aussi pour le caractère actif (91 %) et la difficulté (83 %) ; il est jugé assez beau (68 %), assez bien organisé (79 %), d'une fiabilité un peu juste (61 %), mais peu apaisant (55 %), ce qui surprend. « Internet » reçoit son lot habituel de qualificatifs : il est – après « Forcalquier » – le *moins profond* des mots proposés (56 %) ; surtout, il est *froid* (36 %), même s'il l'est moins que « statistique » ! S'il suppose une dose certaine d'activité (74 %), si, bien sûr, il est reconnu très utile (83 %), il est aussi estimé facile (48 %) et peine à être jugé fiable (50 %) et plus encore apaisant (47 %). Par ailleurs, il est tenu pour plutôt organisé, plutôt concret, et, de façon surprenante, plutôt beau aussi (ces trois scores totaux tournent autour de 59 %). On ne s'étonnera pas non plus de l'image de « statistique » que le groupe renvoie. Ce mot est le *plus froid de tous* (27 %) ; c'est aussi le *plus irritant* (38 %), le *moins beau*, et de loin (41 %). Bien entendu, il est trouvé plutôt difficile (59 %) quoique relativement concret (61 %). Il est aussi assez profond, fiable, actif (à 67 %). Et, nous le savions, il est jugé utile (73 %). Mais surtout, il est *organisé* (83 %).

Un *cinquième* groupe a fait l'objet du même « geste exploratoire » de notre part : il rassemblait 17 étudiants du centre de Lambesc, tous en deuxième année de master professionnel, dont 12 avaient choisi l'option « Responsable de formation » et 5 l'option « Ingénierie de la formation multimédia ». Voici le tableau des scores attribués sur les dix échelles déjà considérées.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	I/O	L/B	P/A	F/D
Université	75	62	86	51	70	58	55	73	66	64
Lambesc	68	78	74	43	62	60	50	62	70	61
Internet	49	31	75	35	45	59	62	51	58	38
Statistique	62	44	75	26	62	60	86	42	70	61
Éducation	80	77	86	49	65	60	63	73	76	72

« Internet » est le moins profond des cinq mots proposés. Il est aussi, et de loin, le plus froid. Utile à l'égal de « statistique », il l'est sensiblement moins que « université » et « éducation ». Il est irritant – seul « statistique » irrite plus que lui. Il est en outre, ici, placé bon dernier pour la fiabilité. Il obtient en revanche des scores semblables à ceux des autres mots pour le caractère concret. Pour l'organisation, cependant, s'il bat « université » et « Lambesc » et fait jeu égal avec « éducation », il est largement dépassé par « statistique ». On retrouve ensuite qu'il n'est ni beau ni laid, et que, sans qu'on puisse le dire passif, il est, du point de vue de l'activité, sensiblement dépassé par tous les autres mots, qui le distancent plus nettement encore du point de vue de la difficulté – « Internet », c'est facile !

On peut agréger les réponses des cinq groupes sollicités. À titre d'exemple, on a calculé ci-après le tableau des scores en pourcentage sur les dix échelles utilisées et pour les seuls mots « Internet » et « éducation ». On obtient ce qui suit.

Échelle	S/P	F/C	PU/U	I/A	PF/F	A/C	L/B	P/A	I/O	F/D
Internet	47	40	80	44	58	59	54	66	61	39
Éducation	87	76	92	54	69	65	75	85	66	76

À nouveau, « Internet » apparaît comme *plutôt superficiel* (« éducation » étant au contraire *très profond*), *froid* et *facile*. Au plan scolaire et universitaire, la « facilité » est un trait péjoratif, qui porte à « regarder de haut » ce qui est jugé ainsi. Nous avons observé que « Internet » était, en même temps, jugé *utile* : on le retrouve ici. Sans doute cela repose-t-il,

parmi les enquêtés, sur une expérience partagée de l'usage de l'Internet. Mais il est peu contestable qu'il existe aussi toute une idéologie très prégnante qui ne place pas bien haut ce qui est (simplement) *utile* : la problématique de l'utilité des savoirs, centrale en TAD, ne va nullement de soi (Chevallard, 2007c).

Le différenciateur sémantique fait ainsi apparaître, au niveau de groupes même de taille réduite, sinon toujours au niveau de chacune des personnes, des attitudes bien partagées. Du point de vue de la clinique didactique, l'exploration *individuelle* révèle ainsi, à un premier niveau, une polarisation de *groupe*. À un second niveau, quand on enchaîne les prises d'information sur des groupes différents, elle fait apparaître des *permanences* qui transcendent chaque groupe. En l'espèce, les attitudes dégagées ici à l'aide du différenciateur sémantique désignent bien l'Internet comme une réalité dénigrée.

¹ Le sigle DESS désigne un Diplôme d'Études Supérieures Spécialisées ; le signe DEA, plus loin, désigne, lui, un Diplôme d'Études Approfondies. Ces dénominations ont laissé place aujourd'hui, respectivement, à celles de *master professionnel* et de *master de recherche*.

² Nous avons introduit ce sigle dans notre article *Apprendre la recherche d'informations sur Internet à l'école* (Ladage, 2007a). Nous le reprenons ici pour sa concision. On notera sa proximité avec le sigle classique RI, pour « recherche d'information », qui lui-même fait écho à l'anglais IR, pour *Information Retrieval*.

³ Voir Spink, Wolfram, Jansen & Saracevic, 2001.

⁴ Voir par exemple la thèse de doctorat de Landy Rajoson (1988).

⁵ Sur les notions de dystopie et d'anti-utopie, on pourra voir l'article "Dystopia" de l'encyclopédie *Wikipedia*.

⁶ Dans un entretien d'octobre 2006 à CNT News, Timothy Hornyak (2006b) mentionne le témoignage d'un roboticien japonais qui, ayant rencontré dans un train une dame sexagénaire, eut avec elle une conversation révélatrice : "... she asked him, 'So what do you do?' and he said 'I'm a roboticist.' And she said, 'Oh, I'm really looking forward to the time when robots are going to take care of me.' Le vieillissement de la population japonaise et la diminution corrélative de la force de travail conduit à envisager l'usage social massif de robots, notamment dans l'aide aux personnes âgées.

⁷ Nos activités professionnelles nous ont conduite toutefois, pendant quelques années, à travailler dans le monde de la bibliophilie et de l'édition de livres rares.

⁸ Voir par exemple Chevallard, 1990.

⁹ On peut noter en passant que *nexus*, « lien, nœud, étreinte », désigne aussi le débiteur insolvable, qui devient par là esclave de son créancier – jusqu'à ce qu'il s'acquitte de sa dette.

¹⁰ On ne confondra pas l'emploi fait en TAD du mot praxéologie et l'usage plus ancien du mot pour désigner une hypothétique « science de l'action humaine ». Il n'y a rien d'autre, en TAD, que des praxéologies : parler de *la* praxéologie, absolument, n'y a aucun sens.

¹¹ Nous devons cet exemple à Y. Chevallard : il pourra éclairer le lecteur « mathématicien ». La référence dans ce qui suit au caractère « archimédien » des entiers doit être entendue ainsi : il existe des structures mathématiques « non archimédiennes », où donc, dans certains cas, la suite des multiples d'un élément *b* ne parvient pas à dépasser un élément donné *a*. Pour un exemple, on pourra se reporter à l'article « Archimédien » de l'encyclopédie Wikipédia.

¹² Voir le « Baromètre Xiti » : <http://barometre.secrets2moteurs.com/index.php/Barometre-xiti>.

¹³ Les recherches dont nous faisons état ici ont été réalisées le 5 août 2008, vers 12 h 30. Le dépôt légal est de l'ouvrage est de février 2008.

¹⁴ Notons que la fiche de la base de données Cat.Inist (CNRS) permet de comprendre que la mention sibylline figurant dans la note de Lecourt reproduite plus haut, à savoir (14.53), est supposée indiquer que l'article a paru dans le volume 14, n° 53 de la revue *Commentaire*. Pour le texte de Georges Canguilhem lui-même, on pourra *par exemple* aller voir à l'adresse indiquée ci-après, que fournit Yahoo! mais que Google ne semble pas connaître : http://217.126.81.33:501/psico/sesion/ficheros_publico/ficheros.php?opcion=textos.

¹⁵ Ce qualificatif renvoie au philosophe et pédagogue Johann Friedrich Herbart (1776-1841). Le lien – apparemment ténu – ainsi établi entre le « schéma herbartien » de la TAD et l'œuvre de Herbart trouve son origine dans ce passage de celui qu'on a appelé parfois « le père de la pédagogie scientifique » (cité dans Cauvin, 1970, p. 38) : « Le professeur [...] n'est plus un enseignant (*Lehrender*), l'étudiant n'est plus un enseigné (*Lernender*) ; mais ce dernier poursuit des recherches personnelles, le professeur ayant pour tâche de le guider et de le conseiller dans ces recherches. » À propos de la référence à Herbart, Y. Chevallard (communication personnelle, 24 mars 2006) mentionne le *théorème de Boyer* – Carl Benjamin Boyer (1906-1976) était un historien des mathématiques –, aux termes duquel “mathematical formulas and theorems are usually not named after their original discoverers”. Cette « loi » s'applique, semble-t-il, au théorème de Boyer lui-même (voir Hatzipolakis, 2000) ; et sans doute aussi au schéma de Herbart... Toutefois, on peut se demander s'il n'existe pas une affinité spontanée entre la TAD et « l'ontologie faible » de Herbart (Maigné, 2007).

¹⁶ À l'entrée MATHÉMATIQUE, parlant du mot latin *mathematicus*, le *Dictionnaire historique de la langue française* (Rey et al., 1993) précise : « Le latin l'a repris au grec *mathematikos* “qui désire apprendre, scientifique” et spécialement “qui concerne les mathématiques”, substantivé dans *ê mathematikê* (*tekhne*) comme nom de science. Ce mot est dérivé de *mathêma* “ce qui est enseigné”, employé au pluriel pour “connaissances”, par opposition à *mathêsis*, qui met l'accent sur le fait d'apprendre. L'un et l'autre sont dérivés de *manthanein*, verbe passé de sa signification première, “apprendre, par l'expérience, apprendre à connaître, à faire”, au sens

plus abstrait de “comprendre”. » On notera le déplacement du sémantisme de *manthanein* depuis le plan de la *praxis* vers le plan du *logos* (ces mots étant pris au sens de la TAD).

¹⁷ La TAD contient une théorie de l'évaluation ; celle-ci exclut toute idée de valeur *en soi*, absolue, d'une praxéologie. Il n'y a de valeur qu'au service d'un projet.

¹⁸ On peut rapprocher cette affirmation de certaines des conclusions de Rothkopf : voir Theilheimer, 2006.

¹⁹ Nous reviendrons sur ce point dans la conclusion de la troisième partie de notre travail.

²⁰ Le titre de l'ouvrage pourrait porter à croire qu'il appartient à la catégorie précédente (celle du « camp de base »). Le titre original, ou du moins celui de la version de l'ouvrage mise en ligne avant sa parution sous forme imprimée, *Référencement 2.0*, paraissait en revanche plus « avancé », quoique, il est vrai, plus énigmatique. Cette impression était cependant corrigée par un sous-titre explicite : *Comment optimiser son site pour améliorer sa visibilité sur les moteurs de recherche ?*

²¹ Ce bouton ouvre directement la page Web à laquelle renvoie le *premier* résultat affiché sur la première page de résultats fournie par Google – supposé être le résultat (presque) “parfait”.

²² Rappelons que, en anglais américain, le mot *billion* désigne un milliard, soit 10^9 (ou mille millions). Le même mot désignait autrefois, en anglais britannique, mille milliards ou 10^{12} , ce que l'anglais nomme aujourd'hui un *trillion*.

²³ Pour un exemple, voir Chevallard, 2007d.

²⁴ Le rapport examiné (Fallows, 2008) comporte un cinquième résultat, déjà évoqué, qui concerne les “demographic differences among searchers” : “men and younger users are more plugged into the world of searching than women and older users.”

²⁵ Université privée féminine située à Wellesley, près de Boston, l'une des meilleures des Etats-Unis pour les sciences humaines.

²⁶ Notre expérience de la formation, comme « formée » et comme formatrice, nous a montré combien cette absence d'uniformité est préjudiciable à la diffusion praxéologique.

²⁷ La copie d'écran ci-après présente les premiers résultats affichés par Google lors d'une recherche effectuée le 11 août 2008. Selon les informations que nous avons pu recueillir, c'est pour avoir vu ce genre d'affichage que le professeur en question a pensé que Firefox était une secte.

²⁸ Sur ce point, nous suivons certains développements que l'on trouvera dans Chevallard, 2008, séance 5.

²⁹ La justesse d'une décision économique n'est jamais certaine : il n'apparaîtra que plus tard, peut-être, que telle œuvre était décidément incontournable, tandis que telle autre aurait pu, sans que cela nuise en rien à la conception et à la réalisation du projet, être ignorée.

³⁰ Nous reviendrons sur la question des critères de choix dans la dernière partie de ce travail.

³¹ Sur la notion d'échantillon disponible, voir par exemple McGee, 1975, p. 30.

PREMIÈRE PARTIE

**ENTRER DANS LE MONDE INTERNET ?
ANALYSE D'UNE ENQUÊTE CODISCIPLINAIRE**

Introduction

On aime tant la vérité que, si on aime autre chose qu'elle,

on veut que ce soit la vérité.

Augustin d'Hippone, *circa* 400

Les praxéologies de l'enquête codisciplinaire, les conditions et les contraintes gouvernant leur dynamique historique et leur diffusion sociale constituent l'horizon de notre travail. Il était donc normal que nous nous arrêtions d'abord sur cette réalité empirique qu'est l'enquête codisciplinaire en donnant à voir et en analysant une ou plusieurs enquêtes dont le résultat, sinon la chronique, nous soit connu de façon raisonnablement documentée. Le choix, aujourd'hui encore, ne saurait porter que sur très peu de spécimens d'enquête répondant aux critères qui étaient les nôtres. Le livre récent publié par une journaliste justement estimée à propos de « l'affaire Monsanto » (Robin, 2008a) illustre sans doute excellemment, dans une enquête où de multiples disciplines entrent en synergie, le recours systématique (et apparemment bien intégré et bien vécu) aux ressources de l'Internet – du moins si l'on en croit le documentaire compagnon du livre, disponible en cédérom, où la chose est illustrée de façon récurrente (Robin, 2008b). Mais la recherche d'information sur Internet n'est en ce cas qu'un discret contrepoint, dont le rôle et les manques propres sont mal évaluables tant l'enquête mobilise de ressources variées, spécifiques de la question étudiée. Étant donné notre problématique, il nous a paru alors préférable de choisir une enquête dont l'objet aurait à voir avec les outils même de l'Internet ; où l'objet d'étude, donc, serait en quelque sorte homogène à certains au moins des outils de l'étude à conduire. Il existe certes des études du monde Internet ou de certaines des créatures qui le peuplent, tel Google (Ippolita, 2008). Mais ces études sont souvent le fait de spécialistes dont la profonde science, fruit de longs travaux et de durs labeurs, ne témoigne d'abord que d'un fait connu de chacun : le temps passé, l'effort, la persévérance sont – ici comme ailleurs, et sans doute ni plus ni moins qu'ailleurs – les gages d'une réussite du petit nombre, réussite dont la *scalability*, pour parler comme ces spécialistes, c'est-à-dire la capacité à s'élargir au (plus) grand nombre, n'est toutefois pas le fort.

L'idée fut donc de se tourner vers une enquête, s'il s'en trouvait, dont l'objet se découperait, certes, dans le monde Internet, mais qui ne serait pourtant pas le fait de quelque expert patenté en la matière. Au contraire, notre souhait était d'observer – et de donner à voir – un « profane », si possible lettré, aux prises avec un monde qu'il n'aurait abordé jusque-là que d'une façon fort limitée, prudemment ancrée dans des usages techniques presque sciemment dépourvus d'épaisseur technologique « savante ». Une exigence supplémentaire était que le travail en question ait été publié, de façon à être, en droit comme en fait, offert *ipso facto* tant au débat public – ce qui n'était pas notre affaire – qu'à l'analyse didactique – que nous entendions conduire avec bienveillance, mais sans complaisance. Sous ces conditions, l'établissement d'une *short list* devenait chose facile. Trois ouvrages nous ont paru un temps susceptible de constituer le matériel empirique sur lequel faire porter la recherche : le livre mince et dense de Jean-Noël Jeanneney intitulé *Quand Google défie l'Europe*, paru en 2005 ; le gros ouvrage, d'une érudition sûre et généreuse, de Clarisse Herrenschmidt, *Les trois écritures*, sous-titré *Langue, nombre, code*, paru en 2007 ; et enfin le livre touffu, impétueux et imprécateur de Barbara Cassin, *Google-moi. La deuxième mission de l'Amérique*, paru la même année.

Nous retrouverons à plusieurs reprises dans notre travail l'étude ciblée, allègre, maîtrisée, efficace de Jean-Noël Jeanneney ; mais ce n'est pas elle que, finalement, nous avons retenue. L'éditeur de l'ouvrage de Clarisse Herrenschmidt présentait son travail, en le situant « au carrefour de plusieurs disciplines, la philologie, l'histoire, l'anthropologie et la linguistique » (et en oubliant donc tout de même les mathématiques, l'informatique et, plus généralement, les sciences du numérique, auxquelles il n'échappe pas), comme le fruit d'une immense enquête sur « l'aventure sémiologique » de l'homme, depuis Sumer jusqu'à l'Internet. Très vite, pourtant, le cadre général de l'enquête nous est apparu démesurément étendu par rapport au projet qui était le nôtre : analyser, sinon exhaustivement, du moins assez précisément, les conditions et les contraintes d'une enquête codisciplinaire à propos du monde Internet – pourvu qu'elle fût conduite, rappelons-le, par une personne docte sans doute, mais relativement novice en la matière, en sorte qu'apparaisse ou se devine *l'effort didactique* exigé d'elle par l'enquête à mener. Nous avons alors arrêté notre choix sur l'ouvrage de Barbara Cassin, auquel cette première partie de notre travail est toute entière consacrée.

Notre étude porte ainsi sur le compte rendu public d'une enquête codisciplinaire ambitieuse, conduite par une intellectuelle de renom, à propos d'une question *a priori* étrangère à son champ de spécialité. Disons ici notre estime, voire notre admiration pour l'effort de l'auteure, pour l'énergie, la détermination, l'ardeur même sans lesquelles, croyons-nous, le livre que nous allons examiner n'aurait pu voir le jour. Étant nous-même passée, en

quelques années, d'une culture essentiellement linguistique, philosophique et politique à la culture propre au monde de l'informatique et de l'Internet, nous croyons savoir ce qu'il faut d'opiniâtreté pour pénétrer si peu que ce soit dans un univers sans doute fascinant, mais rempli de chausse-trapes, et où, bien souvent, ce qui brille est aussi ce qui rebute.

On peut se demander, il est vrai, s'il n'en est pas ainsi en tout apprentissage un tant soit peu substantiel, à quelque âge qu'on l'accomplisse. Ce que nous avons appelé compte rendu d'enquête devrait alors être regardé *aussi* comme une forme masquée de *Bildungsroman* – de roman d'apprentissage. Nous en resterons pourtant au point de vue déjà annoncé. Une enquête est lancée. Il y a une enquêtrice, et qui doit apprendre, et parfois apprendre beaucoup, *ab initio*. Pour cela, elle dispose d'aides – des proches, des amis, des collègues, des relations, qui dessinent autour d'elle une petite « tribu » dont elle peut obtenir un certain secours. (De quelle nature et à quel degré, cela, nous ne le savons pas.) Elle détient aussi un outillage intellectuel, praxéologique, qu'elle va abondamment solliciter. Le fait-elle adéquatement ? Surtout, cet outillage suffira-t-il ? Pour répondre, pour identifier certaines limitations éventuelles de l'étude que nous observions en ses résultats, nous avons lancé, ponctuellement, mais en des points cruciaux, des contre-enquêtes dont, par méthode, nous avons limité l'outillage aux seules ressources de l'Internet, et même généralement à très peu d'entre elles. Tout cela sera fait ici dans un double but solidaire : d'une part, contribuer à la clinique didactique – encore peu développée – de l'enquête codisciplinaire, menée en l'espèce à la frontière des mondes savant et profane ; d'autre part, saisir, dans un cas singulier sans doute mais à bien des égards exemplaire, la structure et la dynamique d'un certain rapport à l'Internet et à ses objets – notamment à « Google », cet objet aux mille visages –, en nous donnant une chance de débusquer ce qu'un tel rapport porte en lui au plan praxéologique et ce que, du même mouvement, il exclut ou, du moins, éloigne.

CHAPITRE 1

L'enquêtrice, la question, les semblables, les savoirs

Entrer dans le monde Internet semble aujourd'hui à beaucoup de gens chose facile. Mais sans doute ce sentiment procède-t-il d'une vision très partielle des choses. Dans son livre déjà cité, John Battelle (2006) le note dans les termes suivants.

The short of it is this: we are incredibly lazy. We type in a few words at most, then expect the engine to bring back the perfect results. More than 95 percent of us never use the advanced search features most engines include, and most search experts agree that the chances of ever getting that number lower are slim to none. (pp. 24-25)

L'abord du problème, en ce cas, se centre étroitement sur le seul système de *recherche d'information sur Internet* (RII), ses potentialités, son emploi réel. Il ignore, presque par principe, l'épaisseur anthropologique de la situation qui conduit une personne à taper quelques mots dans la fenêtre de recherche d'un moteur de recherche. D'une façon très générale, en effet, cette situation est déterminée – d'une manière plus ou moins rigide, certes – par des conditions et des contraintes qui ne sont pas seulement celles créées par les outils de la RII, mais qui tiennent également à *la personne elle-même*, à ses *assujettissements*, lesquels donnent à la personne sa puissance de pensée et d'action en même temps qu'ils la limitent. Or, de ces assujettissements, on ne saurait rendre compte en examinant simplement le "*clickstream*" de l'utilisateur d'Internet, le flux de ses clics de souris au cours d'une session de recherche sur Internet.

Nous tenterons ici d'explorer quelques-unes de ces conditions et contraintes écologiquement liées à la personne et (donc) à ses assujettissements en examinant un cas singulier, pour lequel une certaine documentation est disponible : celui de l'auteur du livre *Google-moi. La deuxième mission de l'Amérique* (Cassin, 2007), que nous désignerons seulement par ses initiales, BC, pour marquer que notre propos ne se réfère pas tant, si attrayantes soient-elles, à la vie et à l'œuvre de cette philosophe de renom, qu'à certaines conditions et contraintes d'une rencontre effective (et donc possible) avec le « monde

Internet ». En particulier, nous ne prendrons nullement parti sur les aspects que d'aucuns pourront juger polémiques de l'ouvrage mentionné, aspects que nous regarderons seulement comme des éléments déterminés (et déterminants), parmi d'autres, du rapport évolutif de BC au monde Internet.

La question inaugurale : comment vivre (au) mieux avec Internet et Google ?

Au point de départ du parcours d'action qui conduit une personne à mobiliser et à solliciter un ou plusieurs outils de recherche sur Internet, il y a en règle générale une *question* que cette personne se pose ; une question qui, très généralement, est déterminée par une situation du monde essentiellement indépendante de l'existence d'Internet ¹. Ce peut être, prosaïquement, une question sur des horaires d'avion ou de train, sur les concessionnaires d'une certaine marque automobile, sur la couleur politique de l'équipe municipale d'une ville où l'on a l'intention de se rendre. Ce peut être aussi une question plus ambitieuse, de plus vaste ampleur, suggérée par exemple par le désir de s'informer en matière d'architecture navale ou d'orthotypographie. S'agissant de BC, la question à la fois *inaugurale* et *génératrice* de sa quête peut sans doute être trouvée dans les lignes qui suivent.

... comment négocier entre l'anathème heideggérien sur l'essence de la technique – réactionnaire ; lucide – et la béatification du présentisme mondialisé, qui parie sur l'amélioration par la technique du sort de tous les hommes dans le monde ? La question est pourtant imparable : nous (qui nous ?) ne vivrons plus sans le Net et quelque chose comme Google, alors comment vivre vraiment mieux, ou au mieux, avec ? (pp. 12-13)

Retenons donc cette question : « Nous (qui nous ?) ne vivrons plus sans le Net et quelque chose comme Google, alors comment vivre vraiment mieux, ou au mieux, avec ? » Et notons que la quête de BC se réfère au monde Internet comme à une partie intégrante de notre monde commun. Car l'Internet fait partie de la situation du monde qui est à ses yeux problématique, et qui motive sa quête d'une solution. Répondre à la question qu'elle se pose implique donc d'établir et de nourrir un certain rapport au monde Internet en général et à « quelque chose comme Google » en particulier. C'est alors dans cette dynamique de construction d'un certain rapport que s'impose à BC, subrepticement ou ouvertement, parcimonieusement ou à profusion (nous ne le savons pas), le recours à des recherches sur Internet et l'emploi corrélatif d'outils de recherche sur Internet.

Le point de départ : une “vanity search” expéditive

Dans le cas que nous suivons ici, la description et l'analyse des conditions et contraintes sous lesquelles la personne agit et pense son action doit porter certainement sur l'environnement social, culturel, idéologique, philosophique, saisi dans ses rapports avec la question soulevée. La formulation de cette question (reproduite plus haut) montre d'emblée un certain positionnement personnel (mais qui n'est pas *que* personnel) vis-à-vis de l'Internet, ou plus justement de Google. Une première part de notre description se doit de mentionner la présence, chez l'enquêtrice, d'une certaine animosité à l'endroit de Google, une hostilité qu'atteste cette citation figurant sur la quatrième de couverture de son livre ² : « Google est un champion de la démocratie culturelle, mais sans culture et sans démocratie. Car il n'est un maître ni en culture (l'information n'est pas la *paideia*) ni en politique (la démocratie des clics n'est pas une démocratie). »

L'auteure est une spécialiste des textes de l'antiquité grecque, de ceux des sophistes en particulier. Derrière l'âpreté de son propos, on saisit le souffle d'une culture et d'une tradition qui n'entendent pas renoncer au droit du premier occupant. Or voici que, selon le récit qui nous est fait, Google tout à coup fait irruption dans cet univers de haute et ancienne culture, jusque-là maître chez lui, et qui brusquement se trouve aux prises avec un nouveau venu inattendu et que l'on trouve brutal. Un propos fortuit tenu à BC – « Je sais qui vous êtes, je vous ai tapée sur Google » – lui révèle un jour, au détour d'une conversation presque triviale, que son monde est envahi par une matérialisation de la figure de *Big Brother*. Elle-même se livre alors, avoue-t-elle, à ce que John Battelle (2006, p. 29) nomme une *vanity search* : elle tape son nom sur Google, s'aperçoit qu'elle est « googlée » – notamment pour un ouvrage qu'elle a dirigé sur le *Vocabulaire européen des philosophies* (Cassin, 2004) –, et s'y retrouve dans la compagnie improbable – même si, convient-elle après examen, elle n'a pas à en rougir – de deux autres Barbara Cassin, l'une américaine, l'autre néo-zélandaise !

L'opération de recherche sur Internet qui permet d'obtenir ce résultat n'est en vérité pas tout à fait immédiate. Il faut ainsi aller interroger, non le Google donné par défaut lorsqu'on est en France, mais le Google international, en anglais. Au demeurant, conclure que Google ne connaîtrait en tout et pour tout que *trois* Barbara Cassin ne va pas de soi. En vérité, BC a confondu deux d'entre d'elles : la Néo-zélandaise, officier de la marine de son état, et une certaine Barbara Cassin-Gionfriddo, militante de la *Connecticut association of the deaf*, l'association des sourds du Connecticut ³. À cela une recherche un peu plus persévérante viendrait ajouter une cinquième Barbara Cassin, vivant, elle, en Californie, et généreuse donatrice d'un club de voile pour enfants défavorisés ⁴ ; une Barbara Cassin, donc, dont

nonobstant la moindre noblesse intellectuelle de ses intérêts, BC ne devrait pas à avoir rougir non plus.

Ce que BC a dû faire pour parvenir aux informations – partiellement erronées – qu’elle restitue aux pages 10 et 11 de son livre, elle n’en dit rien à son lecteur : le problème à résoudre – trouver les Barbara Cassin mentionnées sur l’Internet – est sans doute vu comme une affaire secondaire, en elle-même peu digne de faire l’objet d’une relation détaillée. Un point n’est pourtant pas anodin : imagine-t-on BC commettant, sur les textes grecs (ou autres) qu’elle affectionne, une bétise semblable à celle qui conduit à la confusion que nous avons relevée – sans parler de l’oubli que nous avons constaté ? Dans un entretien (avec Colette Briffard) pour la revue *Texte !* à propos du *Vocabulaire européen des philosophies*, BC déclare ceci (Cassin avec Briffard, 2006).

Quand j’ai enseigné la philosophie dans les lycées ou même quand je l’ai enseignée dans des endroits plus bizarres que les lycées, par exemple les PTT (j’ai en effet eu un parcours très atypique), je me suis aperçue que ce qui m’intéressait et ce qui intéressait les élèves, à quelque niveau qu’ils soient, c’était le travail qu’on faisait sur les textes. C’était là où il se passait vraiment des choses, et pas dans les idées générales. Ce travail se faisait sur les textes en langues, forcément. C’est au plus loin possible des QCM, des contrôles de connaissance que l’on fait aujourd’hui et du prêt-à-penser. Je pense que le Dictionnaire des intraduisibles est un outil pour montrer que c’est comme cela que le travail se fait, que le travail quotidien peut se faire.

Sans doute BC n’a-t-elle pas, à l’endroit du « corpus Internet », la même exigence tenace qu’on l’imagine manifester à l’égard des corpus de textes qui lui sont chers. Ajoutons que, si l’opération n’est pas décrite, le nécessaire travail d’apprentissage qui permet la recherche sur Internet évoquée ici l’est moins encore ; mais on sait que, d’une façon générale, le refoulement du didactique est un fait de civilisation massif.

Le trauma et le collège invisible

Si rien ne laisse entendre de façon explicite que cet épisode de recherche sur Internet ait été pour BC le tout premier du genre – peut-on avoir dirigé le *Vocabulaire européen des philosophies* et être innocente de tout commerce avec Google et autres outils de recherche ? –, on doit conclure, à suivre l’auteure, que ce fut là une forme de *trauma*, à l’origine de son enquête. Un second épisode « traumatique » aura lieu au printemps 2005 (le premier épisode, lui, n’est pas daté), lorsque, dans le cadre d’un colloque tenu à Thessalonique, BC tombe des

nues en entendant le représentant de Google Europe déclarer sans ambages : “*Our mission is to organize all the information in the world.*” Toute l’information du monde ! Pour BC, l’arrogance de la chose ne passe pas. Dans la même période paraît le pamphlet de Jean-Noël Jeanneney, *Quand Google défie l’Europe* (Jeanneney, 2006), dont le sous-titre est explicite : « Plaidoyer pour un sursaut ». Sa publication est une réaction au projet Google Print⁵. Tout cela participe d’une attitude globalement anti-Google. Le paradoxe est alors que, dans un milieu de haute culture et de forte exigence intellectuelle, cette hostilité commune va conduire BC à se lancer dans une étude sans doute dépourvue d’aménité, mais tenace, sérieuse sinon toujours exacte, de l’empire Google.

Cet empire n’est pas *tout* l’Internet. Comme BC le souligne, Google n’est pas le Net : c’est un « révélateur du Net », selon le titre du chapitre I de son ouvrage. Mais vouloir s’interroger sur Google oblige à entrer, d’une manière ou d’une autre, dans le monde Internet. Parmi les conditions de possibilité de l’étude d’un objet donné, nous retiendrons d’abord, ici, cet impératif que la TAD met en avant : sauf exception, on n’étudie pas *seul*, mais ensemble, à *plusieurs*, au sein d’une communauté d’étude, d’une « tribu » qui autorise l’étude, en soutient la légitimité, et lui insuffle une dynamique – à moins que, au contraire, elle ne l’entrave ou même ne l’interdise explicitement. Dans le cadre scolaire habituel, la tribu qui permet à chacun de progresser dans l’étude est, officiellement, la *classe*, en laquelle de petits « clans » d’élèves peuvent se former qui fonctionnent comme accélérateurs de l’étude, ou au contraire comme instance inhibitrice. Dans d’autres cadres sociaux, la tribu dont l’étude envisagée reçoit le *nihil obstat* apparaît comme un ensemble plus flou, quoique non moins effectif pour cela, jusqu’à prendre les allures de ce que Robert Boyle et ses pairs appelaient au XVII^e siècle un “*invisible college*”, un collège invisible⁶.

Qu’en est-il dans le cas de BC ? Quelle « tribu » a soutenu son étude de Google – même s’il s’agit bien d’un travail d’auteur tout personnel, marqué de singularités fortement assumées ? Quel collège, quelle *société de semblables*, avec qui le dialogue est possible dans une certaine connivence et, en tout cas, non sans une certaine confiance, peut-on imaginer autour de la quête de BC ? Une réponse sans doute partielle est apportée à cette question par la liste de noms figurant à la page 7 de l’ouvrage, sous le titre « Remerciements ». La voici.

Je remercie Éric Alliez, Françoise Balibar, Jean-Marie Borzeix, Christine Buci-Glucksmann, Jacques Cassin, Hélène et Olivier Cayla, Laurent Catach, Souleymane Bachir Diagne, Jean Dixsaut, Xavier Fels, Daniel Ferrer, Jean-Pierre Guillemant, Anne-Sophie Legendre, Jacqueline Lichtenstein, Daphné Mamat, Maurice Matieu, Roland Moreno, Xavier North, Irène Rosier-Catach, Daniel Saadoun, Valérie Tesnières, Caroline Wiegandt, Heinz Wismann, Peter Wittenburg et, bien sûr,

Aalam Wassef, qui tous, à un moment ou un autre, m'ont fait bénéficier de leur savoir, de leur lecture, de leurs conseils. Je remercie particulièrement Xavier Perrot, qui m'a permis de rectifier beaucoup d'erreurs dans un domaine qui n'est pas le mien, mais le sien.

Le lecteur aura peut-être reconnu quelques noms que ses lectures ou ses fréquentations lui auront rendus familiers. Le collège des semblables, ou du moins la part de ce collège invisible qui apparaît ici, est plus homogène qu'il ne paraîtra peut-être à qui n'y repère que quelques noms de loin en loin. À de rares exceptions près, en effet, la plupart des membres de ce collège ont pour « genre prochain » le monde des arts, des lettres, de la philosophie, à travers l'enseignement, l'édition, l'érudition, la recherche ; ce qui renvoie, certes, à un vaste continent ! Mais c'est une « différence spécifique » largement partagée qui apparaît alors comme emblème de la tribu : plus d'un, en effet, a dû, souvent pour des raisons professionnelles, engager et développer un commerce technique, intellectuel, culturel avec le monde Internet et donc aussi avec la double infinité de ses ressources et de ses dangers.

Nous ne nous arrêterons ici, brièvement, que sur l'un d'eux, Jean-Marie Borzeix, qui, à l'instar de BC, a participé au colloque de Thessalonique. Journaliste (il fut notamment rédacteur en chef des *Nouvelles littéraires*) puis directeur littéraire au Seuil, directeur de France Culture, membre du Haut Conseil de la Francophonie, avant d'être nommé en 2002 conseiller du président de la BNF pour la francophonie et les relations extérieures (à une époque, donc, où le président de la BNF n'était autre que Jean-Noël Jeanneney), il a publié récemment, à Montréal, des *Carnets d'un Francophone* (2006) dont nous extrayons le passage suivant, à la tonalité subtilement ambivalente à l'endroit du monde Internet ⁷.

... une révolution est en cours qui va changer la donne. L'espace imaginaire, indissociable de l'utopie francophone, devient la réalité. Grâce à Internet. Un monde multipolaire, où chacun peut aspirer à devenir le centre d'émission d'un message, où chacun peut entrer en contact dans une relation particulière avec quiconque parle sa langue, n'est-ce pas à la fois l'utopie de la francophonie et la réalité de la Toile ? Une fois encore, les visions des poètes ont précédé les découvertes des scientifiques. Le monde de la communication d'hier était vertical, celui de la communication de demain sera horizontal. À l'égard de la Toile, beaucoup de francophones, à l'exception notable des Canadiens, ont dans un premier temps manifesté plus de méfiance que d'intérêt. Ils ont redouté que la machine et le système, élaboré par des militaires américains, ne soient un rouleau compresseur appelé à écraser les différences linguistiques et culturelles, à uniformiser la planète. Disons-le tout net (sans jeu de mot), ils ont eu une grande frayeur pour leur langue. Ils découvrent depuis peu qu'il s'agit peut-être pour elle d'une chance inespérée. Le français étant l'une des rares langues présentes

sur tous les continents et la seule avec l'anglais dont les pratiquants sont à ce point éloignés les uns des autres sur la Terre, la Toile peut tisser entre eux des liens jusqu'ici inimaginables.

Spécialiste des questions de droit d'auteur, Jean-Marie Borzeix a pu conseiller BC, en relation avec les projets de Google, sur ces problèmes complexes. La « tribu dialogique » qui transparaît dans les remerciements que l'auteure adresse à certains de ses membres est riche de beaucoup de compétences et représente à n'en pas douter une somme rare d'expériences diversifiées du monde Internet. Sans que, sauf exception rarissime, nous ne puissions avancer rien de certain à ce propos, il est hautement vraisemblable que, en dépit de la forte personnalité de BC, ce collègue invisible aura joué un rôle non négligeable dans l'enquête dont son livre nous donne un compte rendu ⁸.

À la petite tribu que l'on devine ainsi discrètement présente autour du livre de BC, il faut sans doute ajouter, pour constituer le « collège des semblables », d'autres partenaires, pairs et compagnons. Ainsi en va-t-il certainement d'Alain Badiou, qui dirigeait à l'époque avec BC la collection *L'Ordre philosophique* aux éditions du Seuil. Sans doute ne faut-il pas sur-interpréter le fait que Badiou ne soit pas compté au nombre de ceux qui ont soutenu, symboliquement ou très concrètement, l'effort de BC aux prises avec Google ; car bien d'autres noms attendus manquent aussi, par exemple ceux d'Isabelle Stengers, de Jacques Rancière ou de Yann Moulier-Boutang. La tribu, voire le clan au sein de quoi se pense, se vit, s'anime un projet d'étude et de recherche est en effet, en règle générale, découpé dans un collectif plus large, celui que nous avons nommé ici « société des semblables », dont il mobilise quelques-uns des membres en créant ainsi une totalité plurielle de circonstance, dont l'apparent bariolage ne doit pas faire oublier le partage de certaines références essentielles en ce qui touche de plus ou moins près à l'objet de l'enquête. C'est avec ce soutien étagé en degrés différenciés – depuis ceux qui seront fréquemment sollicités jusqu'à ceux qui se contenteront d'attendre ce que l'auteur donnera finalement sur le sujet étudié – que BC va se colleter avec les questions en quelque sorte *vitales* que soulèvent à ses yeux l'existence et le projet de développement de Google.

Cet abord ne se fait qu'avec le soutien de la tribu, dans la référence au collège des semblables et, souvent, comme ici, dans une attitude défiante à l'endroit de ce qu'on peut appeler l'*anti-collège*, soit l'ensemble flou de ceux à qui BC reprochera par exemple de vouloir remplacer le bon *savoir* par la mauvaise *information*, les langues de culture – anciennes ou modernes – par le *globisch*, la *qualité d'une œuvre* par la quantité de clics : autant de thèmes martelés par BC au long de son ouvrage.

Le globish est le fruit d'une tentative pour promouvoir un code de communication international à base d'anglais (voir «Globish»). Sa fortune récente dans les pays francophones semble liée à la publication du livre de Jean-Paul Nerrière, Philippe Dufresne et Jacques Bourgon, Découvrez le globish (2005). BC, qui mentionne Nerrière à la page 228 de son livre, y précise qu'elle « ne partage pas sa conviction que la promotion du globish permette de sauver le français comme langue de culture ».

On retrouve en tout cela une structure quasi universelle et donc au fond banale au plan de l'anthropologie didactique : une coalition est invoquée au service d'un projet qui a sa figure emblématique – ici BC – et qui s'oppose du même mouvement à d'autres points de vue sur l'objet d'étude – ou, pour parler comme BC, au point de vue de l'Autre, de l'*allos* –, et cela même si le collège des semblables n'a pas, vis-à-vis de la question étudiée, avant comme après l'enquête, un point de vue complètement homogène – « semblable » ne signifie pas « identique ». Pour autant, c'est bien une *personne* singulière qui se met en route ; ce sont ses *assujettissements personnels* qui détermineront son action, ses choix d'outils, ses audaces, ou au contraire ses renoncements explicites ou implicites.

Savoirs indigènes, savoirs allogènes

Une personne aborde l'étude d'une question avec tout un équipement de connaissances, d'habitus, de savoirs ; bref, avec tout un *équipement praxéologique*. Toute étude suppose en outre l'identification et la mobilisation de savoirs neufs pour l'« étudiant », donc de savoirs *allogènes*, et leur articulation avec des savoirs (ou plus généralement des praxéologies) *indigènes* – non pas avec la *totalité* de ces praxéologies, sans doute, mais avec un florilège d'outils praxéologiques tenus par la personne pour pertinents, de façon au moins conjecturale.

Dans cette combinaison *a priori* indéterminée de savoirs indigènes familiers et de savoirs allogènes souvent inconnus à l'orée de l'étude (ou connus seulement à travers quelque signifiant), plusieurs cas de figure s'observent. On peut ainsi être tenté de privilégier les savoirs indigènes, en lesquels on a confiance, au détriment de savoirs allogènes parfois introuvables et, quand on les « trouve », reçus souvent avec défiance ou incrédulité. On peut, à l'inverse, être tenté d'oublier l'acquis, en privilégiant ce qui, par rapport à soi, se laisse regarder comme neuf, inédit, naissant, et auquel on prête alors parfois, à son insu même, des vertus quasiment magiques. Dans le cas qui nous occupe ici, la combinaison observée échappe à l'un et l'autre extrême : il s'agit d'une configuration « médiane », que nous examinerons maintenant ici à gros traits.

La problématique de BC semble d'abord mettre en avant l'indigène plutôt que l'allogène : les problèmes qu'elle affronte, BC ne veut les aborder « ni en informaticien, ni en politique, mais en philosophe ». Ajoutons : en *philologue* aussi, amante du *logos*. C'est donc d'abord dans ce monde des savoirs philosophiques et philologiques que BC va tenter de trouver des outils appropriés à son projet. Une prospection rapide, purement quantitative, permet d'établir le nombre d'occurrences des noms de philosophes anciens ou modernes dans son ouvrage. Parménide y est ainsi mentionné 3 fois, mais Platon l'est 13 fois et Aristote, 9. Spinoza l'est une fois, et Leibniz 7 fois, tandis que le nom de Kant apparaît deux fois, comme celui de Hegel. Le nom de Wilhelm von Humboldt se rencontre 4 fois, celui de Marx une fois, et celui de Nietzsche, deux. Si Heidegger n'est mentionné que deux fois, le nom de Hannah Arendt est cité 12 fois. Ajoutons à cela 5 mentions des sophistes, chers au cœur de BC⁹, deux occurrences du nom de Jacques Lacan, une de Gilles Deleuze, une de John Rawls. Quant à l'écrivain et philologue Victor Klemperer (1881-1960), il est mentionné 6 fois, pour des raisons sur lesquelles nous allons revenir.

Pour saisir les savoirs indigènes dont s'outille le travail de BC, nous suivrons linéairement le texte de son ouvrage. Ces savoirs indigènes lui servent d'abord à penser *ce qui est* – qui se révèle parfois brutalement à elle. Ainsi en va-t-il de l'existence googléenne des multiples « Barbara Cassin ». Comme s'il lui fallait assumer cette pluralité dépersonnalisante en la conceptualisant, BC cite ici Jacques Lacan : « une langue, entre autres, n'est rien de plus que l'intégrale des équivoques que son histoire y a laissé persister » (p. 10). Il y a là une référence et un contenu de pensée à quoi BC paraît tenir : dans l'entretien déjà cité à propos du *Vocabulaire*, elle la mentionne aussi, comme elle l'avait fait déjà dans un entretien « officiel » de présentation de ce même ouvrage¹⁰.

À Colette Briffard, elle déclare en effet ceci : « Quelle que soit la langue, et même pour la nôtre, le français, on vérifie à travers le dictionnaire qu'“une langue, entre autres, n'est rien de plus que l'intégrale des équivoques que son histoire y a laissé persister” – une phrase de Lacan que je trouve remarquablement exacte, dans L'Étourdit. Chaque langue est au fond l'intégrale de ces équivoques, de ces homonymies, qui n'apparaissent telles qu'à partir d'une autre langue. » Le texte de Lacan a paru d'abord dans le numéro 4, pp. 5-52, de la revue Scilicet (1973) ; il a depuis été repris dans Autres écrits (Lacan, 2001).

Les savoirs indigènes l'aident aussi à concevoir *ce qui serait possible* – et, de son point de vue, souhaitable. Évoquant ainsi des « stratégies de réponse » à l'offensive de Google, elle en

distingue de deux ordres, qui conduiraient l'une et l'autre à un « autre que Google », la première stratégie se situant dans une altérité que le mot grec *allos* exprime, la seconde étant par elle associée au mot grec *heteros* – et cela, à vrai dire, sans trop se soucier du lecteur peu au fait des subtilités du grec ancien ¹¹.

Allos s'applique à « un autre de la même sorte » (un autre homme, par exemple) ; heteros s'applique à « un autre d'une sorte différente » (une femme par rapport à un homme, par exemple). D'une manière générale, BC n'évite pas une certaine morgue lorsqu'elle exhibe des connaissances prises aux savoirs qu'elle pratique en tant que philosophe. Cela n'a pas échappé aux commentateurs, comme le montre la critique suivante du livre de BC rencontrée sur un blog (Bortzmeyer, 2007) : « Le ton est très pédant (peut-être pour montrer que, contrairement à ces sauvages d'États-Uniens, en France, on est cultivés), avec citations de Platon, Kant, Spinoza et Hannah Arendt à toutes les pages... »

Les savoirs indigènes sont manifestement, pour l'auteure, comme il en va sans doute pour quiconque, une base de réassurance, où se pensent, s'élaborent, se préparent, les offensives analytiques et critiques concernant l'objet examiné – ici Google. Le passage suivant est à cet égard un excellent représentant de cette manière de faire ¹².

Le Net est donc quelque chose de relativement récent, plus jeune que moi, je l'ai vu naître.

C'est encore une merveille, *stricto sensu*. La merveille de la communication réalisée. Concrètement, un monde plutôt deleuzien, une esthétique support-surface sur fond de cerveau-machine : réseau, rhizome, centre partout circonférence nulle part, multiplicités, subsidiarités, puissances directes de la multitude, simplicité de la connexion, auto-organisation... Un monde fait par tous, en immanence, et non par un. Où Deleuze rejoint Lautréamont : « la poésie sera faite par tous et non par un », et Rimbaud lu par Char : « sa date incendiaire, c'est la rapidité ». Contre un monde si « résolument moderne », il n'y aurait jamais plus que des combats d'arrière-garde. À moins que. À moins que certaines « nouvelles » valeurs ne soient encore plus stéréotypées et ringardes que les anciennes, parce que fermement reprises en main par le plus daté des nouveaux mondes : *dont'be evil...*

Lorsqu'elle se penche sur le schéma proposé par Page et Brin dans leur article inaugural relatif à leur moteur de recherche (Brin & Page, 1998), BC écrit (pp. 54-55) : « Et voici qu'ils retrouvent, au moins en partie, l'anatomie immémoriale du *logos* : parcourir, choisir-cueillir,

enranger, mettre en rapport, pour constituer, tonneau indexé après tonneau indexé, quelque chose comme un nouvel art de mémoire. » La référence se fait ici discrète à la rhétorique et à l'art de la mémoire que, à son acmé, la rhétorique portait en elle¹³. Ce n'est pas cette référence-là pourtant que BC développe : dans une longue note, que nous reproduisons ci-après, elle mêle, comme autant de valeurs culturelles et intellectuelles, Heidegger, le grec, le latin, les mathématiques.

Heidegger rend ainsi *logos* par « pose recueillante » et, se demandant « comment *legein*, dont le sens propre est “étendre”, en arrive à signifier dire et parler », passe par *legen* (all.), « rassembler, réunir », puis *lesen* (all.), « lire » mais aussi « rassembler, étendre devant », et, en composition, « glaner, vendanger, récolter, mettre à l'abri, conserver, rentrer » (« *Logos* [Héraclite, fragment 50] », *Essais et conférences*, trad. A. Préau, Gallimard, 1958⁶, p. 251-253). L'un des sens premiers de *logos* est « proportion » au sens mathématique du terme, « rapport », d'où le double sens de « parole » et de « raison » (*ratio et oratio*, comme traduisent les Latins).

Le recours à des connaissances indigènes de type linguistico-philosophico-philologiques semble parfois superfétatoire. C'est ainsi que, dans son chapitre III, qu'elle consacre à une analyse critique du slogan de Google “*Our mission is to organize...*” (sur lequel nous reviendrons), BC commente dans les termes suivants l'expression “*Our mission...*” (p. 82) : « “Notre”, l'adjectif possessif, marque l'appartenance à un “nous”. “Nous”, première personne du pluriel, inclut “je” dans une communauté. Mais “nous”, comme “je”, “ici” ou “maintenant”, change de référence selon celui qui parle. » Elle renvoie en ce point à une note de bas de page fort concise mais dont le contenu peut paraître inutilement ostentatoire : « Voir, sur le déictique, les pages de Hegel, *Phénoménologie de l'Esprit*, A 1. “La certitude sensible, ou le ceci et ma visée du ceci”. » Le lecteur peut voir là une curieuse parade, où l'on fait montre de ses armes en tirant des coups de feu en l'air, comme si des outils conceptuels et culturels forgés à d'autres fins étaient à tout coup employables sur des problèmes neufs : *Nihil novi sub sole*, rien de nouveau sous le soleil, semble-t-on suggérer.

Non pas organiser, mais faire système

De telles vignettes de pure culture, BC n'est pas avare. Mais son commentaire de l'autre membre de phrase, “... *is to organize*”, lui donne occasion, sur quelques pages (pp. 88-94), d'un morceau de bravoure en matière d'utilisation de « savoirs autochtones ». La source

essentielle à laquelle elle puise, elle la trouve dans une étude du philologue allemand Victor Klemperer que celui-ci met en forme après la guerre à partir de notes prises clandestinement entre 1933 et 1945 : publiée en 1947 sous le titre *LTI, Lingua Tertii Imperii. Notizbuch eines Philologen*, ce travail ne sera publié en français qu'en 1996 (chez Albin Michel) sous le titre *LTI, la langue du III^e Reich. Carnets d'un philologue*. On y découvre un tableau impressionnant et douloureux de la façon dont les Nazis se sont employés à changer la langue allemande pour en faire un outil de propagande.

La première expression que Klemperer ait ressentie comme véritablement nazie, confie-t-il, est celle de *Strafexpedition*, « expédition punitive ». Suivront, entre beaucoup d'autres, les néologismes *Untermenschentum* (sous-humanité), *entjuden* (déjudaïser), *arisieren* (arianiser), *aufnorden* (rendre plus nordique). La *LTI*, qui déferle sur l'Allemagne par la radio et les journaux, fera par exemple grand usage des préfixes *Welt* (mondial), *gross* (grand), ou *Volk*, mot qui se trouve à l'origine d'une autre création nazie, l'adjectif *völkisch* (« du peuple »).

Dans ce contexte de profonde perversion linguistique, le mot de *système*, explique Klemperer, en vient à être diabolisé : dans la langue nazie, il désigne en effet une réalité abhorrée, le *parlementarisme de Weimar*. Par contraste, le mot que la *LTI* valorise est celui d'*organisation* : « Conformément à son exigence de totalité, note Klemperer, le nazisme technicise et organise justement tout. » BC se pose alors en héritière d'analyses que, dans le chapitre 17 de son ouvrage intitulé « Système et organisation », Klemperer explicite longuement, en expliquant pourquoi le nazisme rejette le mot de *système* et promeut celui d'*organisation*. « *Organiser*, écrit alors BC, peut-être hâtivement (pp. 89-90), est l'un de ces mots qui font dresser l'oreille. Et j'en ai froid dans le dos qu'il ne fasse pas dresser l'oreille. »

Elle explicite rapidement le procès que Klemperer instruit contre le mot d'*organisation*, qui, dans la *LTI*, écarterait tout recours à l'entendement – lequel présuppose au contraire la notion de *système* – et promeut du même mouvement les forces mystérieuses à l'œuvre derrière le mot d'*organisation*.

BC cite en particulier ce passage de l'ouvrage de Klemperer (1947/1996, p. 140) : « Mais si le mot “système” est réprouvé, comment s'appelle donc le système gouvernemental des nazis ? Car ils ont bien un système, eux aussi, et ils sont même fiers que ce réseau embrasse absolument chaque manifestation et chaque situation de la vie ; raison pour laquelle “totalité” est une des poutres maîtresses de la LTI. Ils n'ont pas de “système”, ils ont une “organisation”, ils ne systématisent pas avec l'entendement, ils cherchent à entrer dans le secret de l'organique. » BC aurait pu aussi mentionner ce

passage, plus tragique encore, où Klemperer se réfère à l'ouvrage d'Alfred Rosenberg, Le Mythe du XX^e siècle (1930) : « À la place de la vérité une et universelle, censée exister pour une humanité universelle imaginaire, apparaît la “vérité organique” qui naît du sang d'une race et ne vaut que pour cette race. Cette vérité organique n'est pas pensée et développée par l'intellect, elle ne consiste pas dans un savoir rationnel, elle se trouve au “centre mystérieux de l'âme du peuple et de la race”, elle est, pour le Germain, donnée dès l'origine dans le sang nordique : “L'ultime ‘savoir’ d'une race est déjà contenu dans son premier mythe religieux.” » Alfred Rosenberg, né en 1893, fut un théoricien du nazisme reconnu responsable des massacres perpétrés dans l'est de l'Allemagne ; condamné à mort par le tribunal de Nuremberg, il fut pendu le 16 octobre 1946.

BC ne tire pas seulement de là une critique radicale de l'énoncé “*Our mission is...*” ; elle en infère aussi un principe en matière de conception et d'arrangement – pour ne pas dire... d'organisation – de toute bibliothèque possible (p. 91).

Si j'applique cela à Google, le duo-maître est « *organize all... in the world* ». D'où un conseil immédiat à toute Bibliothèque numérique faite ou à faire : systématisiez, n'organisez pas, faites la différence entre un « système », ensemble ordonné par l'entendement et l'intelligence, qui se tient debout ensemble (*sunistêmi*), et une « organisation » qui, c'est très étrange en grec même, lie l'organique et l'ustensilité...

On voit ainsi l'auteure « appliquer » un savoir qu'elle considère à l'évidence comme indigène, sans doute parce qu'il a été produit, dans certaines circonstances certes on ne peut plus tragiques, au sein d'une culture qu'elle tient fondamentalement pour sienne.

Pan et holon

Au-delà même de Klemperer, BC tire alors de son propre fond les outils d'une analyse que, pour une fois, nous reproduirons *in extenso* (pp. 91-92).

Il y a en grec deux manières de dire « tout », *pan* et *holon*, un tout composite ouvert ou une totalité close. Quand il s'agit d'information, le tout est nécessairement un *pan* branché sur l'infini, c'est-à-dire sur le « encore encore » – l'infini sous sa définition aristotélicienne, dépréciative, privative (ou mauvais infini pour la suite de l'histoire de la philosophie) : non pas ce en dehors de quoi il n'y a

plus rien, mais ce en dehors de quoi il y a toujours quelque chose. Une + une information, *ad infinitum*, par adjonction de chaque une dans le temps, dans un temps linéaire, sans dialectique ni éternel retour de l'identique, approprié à l'infini en expansion. Il n'y a pas d'*Aufhebung* (de « relève » ou de « dépassement » comme on traduit parfois) possible de l'information. La volonté d'information est toujours une volonté de plus d'information, par analogie avec la volonté de puissance. Google a une vocation panique, affine au Web comme univers en expansion. Peut-on tenter la différence suivante : le tout « panique » ne fait pas système, il est organiquement en expansion. Mais une bibliothèque « holiste » serait, non, pas l'organisation de toute l'information, mais un système provisoire de la culture tel qu'on en puisse, au moins partiellement, en (*sic*) systématiser les défauts et les manques.

Ce passage montre de façon exemplaire le travail et le jeu avec des connaissances empruntées à des savoirs sans nul doute indigènes pour BC. En premier lieu est citée l'opposition des deux manières de dire, en grec, « tout » : *pan* et *holon*. De cette opposition, le lecteur sera peut-être surpris de découvrir, quelques lignes plus loin, que BC dévalorise le premier terme, *pan* (parce qu'il renverrait à l'idée d'un « tout composite ouvert »), au bénéfice du second, *holon* (qui désigne une « totalité close »). Pour justifier cette dépréciation, BC recourt à Aristote lui-même, en sa conception des deux infinis, actuel et potentiel. En outre, elle souligne que l'addition indéfiniment répétable d'éléments toujours distincts, qui caractérise le « mauvais » *pan*, fait entrer dans un « temps linéaire », où aucune dialectique n'est à l'œuvre, où, selon une notion qui, dans la tradition intellectuelle où s'inscrit BC, renvoie au moins à Hegel, Freud et Lacan, aucune *Aufhebung* n'est opérante.

On doit renvoyer ici au Commentaire parlé sur la Verneigung de Freud de Jean Hyppolite (Lacan, 1966, pp. 879-887), qui précise le concept de Aufhebung en ces termes : « C'est le mot dialectique de Hegel, qui veut dire à la fois nier, supprimer et conserver, et foncièrement soulever. Dans la réalité, ce peut être l'Aufhebung d'une pierre, ou aussi bien la cessation de mon abonnement à un journal. Freud ici nous dit : "La dénégation est une Aufhebung du refoulement, mais non pour autant une acceptation du refoulé." » En anglais, le concept de Aufhebung a été rendu par sublation, terme auquel l'encyclopédie Wikipedia ("Sublation") consacre un article qui commence par les lignes suivantes : "Sublation is an English term used to translate Hegel's German term Aufhebung. The German word Aufhebung literally means 'out/up-lifting.' In Hegel, the term Aufhebung has the apparently contradictory implications of both preserving and changing. The tension between these senses suits

what Hegel is trying to talk about. In sublation, a term or concept is both preserved and changed through its dialectical interplay with another term or concept. Sublation is the motor by which the dialectic functions.”

La mention de l'éternel retour renvoie à Nietzsche, que BC sollicite plus encore pour tenter de justifier, « par analogie avec la volonté de puissance », cette assertion toute personnelle : « La volonté d'information est toujours une volonté de plus d'information... » Le postulat ouvre alors un développement porté par le mot de *panique*, dont le lecteur semble être invité à saisir la parenté avec le préfixe *pan*.

On retrouve ici les effets de connivence recherchés à propos de savoirs que l'on désigne ainsi, contre toute objectivité, comme « bien connus ». Le bruit d'une parenté entre le Dieu Pan – d'où dérive « panique » – et le préfixe pan, « tous », semble avoir couru dans la culture grecque depuis Homère au moins. L'article Pan (Mythology) de l'encyclopédie Wikipedia (en anglais) précise ainsi : “Probably the beginning of the linguistic misunderstanding is the Homeric Hymn to Pan, which describes him as delighting all the gods, and thus getting his name.”

« Google a une vocation panique », écrit donc BC ; c'est-à-dire qu'il ne saurait produire qu'une (mauvaise) organisation (*pan*), et non un (bon) système (*holon*). La mobilisation apparemment toute personnelle de ces éléments de savoir tenus pour sûrs, et la production par BC avec leur concours de ce qui serait un savoir *inédit* sur Google – Google crée la panique au lieu de « faire système » – méritent qu'on l'analyse plus avant, ne serait-ce que dans l'étroite perspective qui est la notre ici.

CHAPITRE 2

Ratés de l'enquête et contre-enquête avec Internet

Si on laisse de côté, ainsi que nous le ferons, tout jugement relatif à la répudiation par BC de toute totalité indéfiniment ouverte, qui ne ferait pas « système », on ne peut cependant qu'être frappé de l'isomorphie entre sa valorisation de l'idée de totalité close – dont on pourrait connaître « les défauts et les manques » – et son style didactique, qui mobilise un système de savoirs (Aristote, Hegel, Nietzsche, etc.) apparemment protégé contre toute invasion « barbare », et qui semble ainsi fonctionner comme une totalité close.

Par contraste, l'image de la totalité ouverte, *indéfiniment recommencée*, car potentiellement infinie, semble engendrer un vertige dont BC s'efforce d'empêcher la survenue. Nous serions portée à penser qu'il y a là l'expression inattendue d'un aspect banal, ou du moins largement partagé, du rapport *professoral* au savoir (et aux savoirs), qui répond subjectivement à une clause essentielle du contrat liant les professeurs à l'institution scolaire et à la société : faire entrer les élèves dans un monde de savoirs supposé *clos* (et non en expansion indéfinie), supposition qui fonde la prétention professorale (et l'exigence qui, de façon consubstantielle, pèse sur les professeurs) de *maîtriser* les savoirs à enseigner. Si notre lecture du texte de BC est juste, il nous faut conclure que celle-ci opère ici moins en chercheur qu'en professeur, sous la forme de « mises au point » magistrales, qui se prévalent d'une panoplie de savoirs faisant système. Pourtant toute mise en œuvre de savoirs *a priori* disponibles porte en elle, en bonne épistémologie, une interrogation indépassable : mobilise-t-on là des savoirs véritablement *pertinents* ? Ou ne recourt-on aux savoirs effectivement mis en jeu que parce qu'ils sont disponibles, et ressentis comme familiers, assurés, en quelque manière rassurants ?

De "In God We Trust" à "Our mission"

Lorsque BC s'attaque par exemple à "*Our mission is...*", lorsqu'elle cite la *Phénoménologie de l'Esprit*, lorsqu'elle rapproche le "*Our*" du "*We*" de la devise *In God We Trust* sur la monnaie états-unienne, lorsqu'elle appuie ce rapprochement d'une longue note, que nous

reproduisons ci-après ¹⁴, mobilise-t-elle des connaissances adéquates à la quête qu'elle poursuit ?

« *The trust of our people in God should be declared on our national coins* », écrit le secrétaire du Trésor, Salmon Chase, en 1861, après la guerre de Sécession, et la devise apparaît pour la première fois en 1864 sur la pièce de deux cents. Sa généralisation de fait n'a fait l'objet d'une loi, votée par le Congrès, qu'en 1956. Sans doute ne faut-il rien moins que le « *trust* » en Dieu, foi-confiance, pour garantir le « *trust* » en la monnaie, sa fiabilité et celle de l'économie américaine, surtout si « nous » ne vivons plus sous le régime de l'étalon-or, mais bien sous celui de la loi anti-trust.

Notons que toutes les informations données ici à propos de la devise des États-Unis apparaissent dans l'article de Wikipedia intitulé "In God We Trust". Si l'on admet – par une inférence certes incomplètement assurée... – que BC a procédé à une brève recherche sur Internet pour en savoir plus sur ce qui est, précise l'article de Wikipedia, "the official national motto of the United States and the U.S. state of Florida", on peut se demander si le savoir allogène ainsi localisé (et exploité alors dans une note de bas de page) ne vient pas prendre la place d'autres savoirs allogènes, que, faute sans doute de les avoir identifiés, BC n'envisage nullement comme constitutifs de l'outillage intellectuel propre à lui permettre une analyse solide de l'énoncé qu'elle s'affaire à déconstruire.

C'est en ce point qu'Internet pourra jouer un rôle d'outil sinon unique du moins essentiel, nous allons le voir. Que signifie, que recèle l'énoncé que BC appelle une « phrase-emblème » : "*Our mission is to organize...*" ? Pour répondre à cette question, BC s'engage dans une problématique qui, on l'a vu, la conduit à passer, sans intermédiation aucune, du plus *spécifique* – l'énoncé incriminé, censé manifester en chacun de ses constituants l'être même de Google, et regardé donc comme lui étant en tout point consubstantiel – à ce qui est pour l'auteure sans doute le niveau plus élevé de généricité « américaine » – la foi publiée partout en un Dieu gardien et garant de la monnaie. Or un geste de recherche des plus simples permet d'accéder à une vérité nue que ce grand écart dissimule ; un geste qui procède d'un questionnement « intermédiaire », de « moyenne portée » (*medium range*) ; un geste qui, humblement, de façon presque routinière, interroge la civilisation américaine dans son « matériel linguistique » même. Ce geste consiste ici à demander à... Google (par exemple) de nous en dire plus sur l'expression "*Our mission is...*". S'agit-il là d'une formulation *rarissime*, ce qui validerait l'idée d'y rechercher une parcelle de la spécificité googléenne ? S'agit-il, à l'opposé, d'une formulation commune, d'un syntagme figé, ou quasi tel ?

L'expression "Our mission is" vue à travers Google

Tentons d'éprouver cette dernière conjecture : l'expression "our mission is" serait-elle une formule figée, toute faite, canonique ? Pour cela, tapons dans la fenêtre de recherche de Google cette expression entre guillemets (voir ci-après).



Réalisée le 29 août 2007, cette opération livre un ensemble de résultats dont nous ne reproduirons ici que ceux figurant sur les *deux* premières pages, en mettant **en gras** les occurrences de "our mission is". On obtient ainsi le tableau suivant.

- <http://www.si.umich.edu/about-SI/mission.htm> :

Information professionals are playing an increasingly vital role in empowering individuals, communities, and organizations to capture the promise of the information age. The School of Information embraces a vision that harmonizes people, information systems, and organizations to improve the quality of life. **Our mission is** to discover the principles and concepts that will enable society to realize this vision, to design the technologies, systems, and practices that will substantiate the vision, and to educate new generations of professionals who will put that vision into practice.

- <http://www.learn4good.com/jobs/language/french/search/company/29091/> :

Thai Restaurant in CT. **Our mission is** servicing with warmth and hearth, Serving delicious authentic Thai cuisine. Satisfying our customer's pleasure, Surplusing "Healthy, tastety, happy"

- <http://geodiscover.cgdi.ca/gdp/search?action=listEntries&entryType=organization&language=fr&selecctionCriterion=i> :

I.E.S. Imagery Exploitation Services, Inc.

OUR MISSION IS TWO FOLD Our primary mission is to provide our clients with detailed accurate interpretation and intelligence analysis of Remotely Sensed Satellite and Airborne derived imagery data...

- http://www.neurosciences.asso.fr/Activites/colloques/SN07/framesEN/Biotech_frameEN.html :

Eurogentec was founded in 1985 as a spin-off from the University of Liège (Belgium). **Our mission is** to design and deliver reliable and innovative services and products for the Life-Science community. As a leading supplier in genomic and proteomic research, Eurogentec is able to offer its customers integrated solutions, whether they use DNA, antibodies, peptides or proteins as research tools. Eurogentec also offers research and development services for the biopharmaceutical industry.

- <http://www.microsoft.com/about/default.msp> :

Our mission is to help people reach their potential.

- http://www.idealists.org/idealists/fr/SiteIndex/AssetSearch/search?assetTypes=Org-NON_PROFIT_TYPE&country=Morocco&sid= :

Our mission is to assist community-based groups to acquire the necessary skills and information so as to be capable of analysing their needs and formulating projects to satisfy them.

- <http://magazine.mba-center.net/index.php?site=magazine§ion=us&page=art1> :

Harvard Business School: **Our Mission Is** to Educate Leaders

- <http://blog.jamendo.com/index.php/category/press-release/> :

Our mission is to help turn visions into realities by providing financing, thoughtful advice, relevant experience and deep industry relationships to our portfolio companies.

- <http://www.kapitec.com/Societe/en/index.html> :

Our mission is to provide the best software tools to streamline the development, testing, and deployment of e-business applications (Java and Web).

- http://www.diplomatie.gouv.fr/fr/actions-france_830/recherche-sciences_1029/programmes-cooperation_2609/laboratoire-europeen-biologie-moleculaire_2610/programme_2618/summary-of-presentation-about-the-european-bioinformatics-institute_4782.html :

Our mission is summarised below :

- To provide bioinformatics facilities and services for the scientific community
- To become a flagship laboratory for basic investigator-driven research in bioinformatics
- To provide advanced bioinformatics training to individual scientists at all levels, from PhD students to independent investigators
- To help disseminate cutting edge technologies.

- http://www.dailymail.co.uk/pages/live/articles/sport/football.html?in_article_id=478084&in_page_id=1779 :

Wenger: Stage one of **our mission** is accomplished

- http://europa.eu/eas/images/eas_leaflet_en.pdf :

Our mission is to:

- ⇒ promote cooperation among the institutions in the area of training,
- ⇒ support the spread of common values and good practice,
- ⇒ create synergies in the use of human and financial resources.

- <http://www.xbee.com/> :

Our mission is the preservation of the environment for the future generations and the improvement of the comfort of life of every person using fuels made from crude oil.

- <http://www.worldmap.org/> :

Our mission is comprehensive: To create maps of languages and people groups for each country of the world while linking appropriate missions related data.

- http://www.fencing.ca/foundational_documents.htm :

Our mission is to enrich the lives of Canadians through the sport of fencing in partnership with our stakeholders by building a common plan for success which : ...

- <http://www.coarc.org/index.php> :

Our Mission is to expand abilities, one person at a time, so individuals experiencing disabilities can achieve their individual goals.

- <http://fra.proz.com/profile/566703> :

Our Mission is “To build customer relationships and deliver accurate translations on time at a competitive price”.

- <http://iao.ish-lyon.cnrs.fr/Francais/villedura/site.html> :

→ **Our mission** is to be a leader in enhancing the quality of life in Calgary by working together to achieve better communities and a safe, livable environment.

→ [La DATAR a pour mission de préparer, d’impulser et de coordonner les décisions relatives à la politique d’aménagement du territoire conduite par l’État.]

→ **Our mission** is to advocate for sustainability in Seattle/King County.

- <http://blogs.msdn.com/jobblog/archive/2006/08/16/702586.aspx> :

Microsoft - Our Mission is My Reality!

Our mission at Microsoft is “To enable people and businesses throughout the world to realize their full potential.



- <http://iambeautiful.com/> :

Our mission is simple: to create a world in which every woman and girl can proudly proclaim, “I Am Beautiful!”

Ce simple échantillon suggère que la formule “our mission is”, et le fait même de parler de *mission*, sont choses *banales* dans l’univers de langue anglaise. Le lecteur attentif aura noté que si la Harvard Business School se fixe pour mission d’éduquer des *leaders*, on voit l’entraîneur de football (français) Arsène Wenger déclarer que la première étape de la *mission* du club dont il s’occupe – l’Arsenal FC de Londres – est accomplie¹⁵. La conclusion essentielle – bien que toujours conjecturale – de ce premier geste empirique est que “our mission is” est d’un emploi banal et même *obligé* dans certains contextes – où l’on est amené à déclarer l’objectif ou l’objet de son activité. Ainsi voit-on le géant Microsoft énoncer sa mission (“our mission at Microsoft is...”), à l’instar de Google, sans doute, mais aussi comme le fait ce restaurant Thaï de Boston (Connecticut), qui, en fait, recherche simplement un “Thai chef”.

“Our mission” *en français*

La formule utilisée en anglais, on le voit aussi, peut être reprise en français, où on peut penser que, cependant, elle n’est pas commune : on la voit ici figurer à propos de la DATAR, mais dans une forme non canonique (la DATAR « a pour mission... »). Elle apparaît encore dans le document concernant la Canadian Fencing Federation, la Fédération canadienne d’escrime, qui présente ses documents en deux langues, comme on le voit sur la copie d’écran ci-après.

<p>Bylaws </p> <p>Vision</p> <ul style="list-style-type: none"> • Our vision is to be the best sport system which provides national leadership and resources for growth and development supporting our fencers to be the best in the world. • We can and will be the best <p>Mission</p> <p>Our mission is to enrich the lives of Canadians through the sport of fencing in partnership with our stakeholders by building a common plan for success which :</p> <ul style="list-style-type: none"> • increases participation • improves the quality of fencing activities • develops resources and tools to provide an effective national fencing sport system • develops excellence in athlete performance <p>through advocacy, promotion and leadership.</p>	<p>Statuts </p> <p>Vision</p> <ul style="list-style-type: none"> • Notre vision est d’être le meilleur système de sport possible, qui procure à nos escrimeurs un leadership national et des ressources de croissance les aidant à devenir les meilleurs du monde. • Nous pouvons être et nous serons les meilleurs. <p>Mission</p> <p>Notre mission consiste à collaborer avec nos partenaires pour enrichir les vies des Canadiens grâce à l’escrime, en élaborant un plan commun visant le succès qui :</p> <ul style="list-style-type: none"> • augmente la participation; • améliore la qualité des activités d’escrime; • élabore des ressources et des outils nécessaires pour disposer d’un système sportif national efficace en escrime; et qui • développe d’excellentes performances sportives; <p>par le soutien, la promotion et le leadership.</p>
---	---

On peut imaginer que la traduction en français de “our mission is” par « notre mission *est* » a pu sembler rendre un son brutal pour des oreilles de francophones : de là peut-être le « notre mission *consiste à* » adopté par la Fédération canadienne de d’escrime ¹⁶. Il est assez facile de voir que, toutefois, l’emploi de « notre mission *est* » gagne du terrain dans le monde francophone, et pas seulement au Canada, quoique peut-être de façon préférentielle dans les activités socio-économiques les plus influencées par l’univers anglo-américain des affaires. C’est là une conjecture que l’on peut tenter de vérifier en interrogeant Google comme on le voit sur la copie d’écran qu’on trouvera ci-après : nous laisserons cela à la discrétion du lecteur.



Que disent les dictionnaires ? La notion de “mission statement”

On peut regarder comme accomplie une première étape de notre enquête sur l’expression “our mission is”. Une deuxième étape peut alors consister, banalement, à interroger des dictionnaires de la langue anglaise à propos du mot *mission*. Pour le commun des mortels, la chose aurait exigé autrefois, en France, beaucoup d’énergie – car qui y dispose chez soi de multiples dictionnaires de la langue anglaise ? Or il est possible aujourd’hui d’accéder en ligne à un grand nombre de tels dictionnaires – qu’on ne pouvait, jadis, même rêver de posséder physiquement dans sa bibliothèque ¹⁷.

Deux éléments de conclusion se dégagent d’un examen toujours partiel, toujours à contrôler. Tout d’abord, plusieurs dictionnaires ne font *pas* référence à l’emploi qui est fait de *mission* dans l’expression “our mission is”. C’est ainsi que le *Cambridge Dictionary of American English* fournit la définition illustrée d’exemples reproduite ci-après.

mission

[Show phonetics]

noun [C]

the action of sending someone to a place to do a particular job, esp. one for a government or religious organization, or the job they are sent to do

They were sent on a secret political mission to the Middle East.

(FIG.) *She's a woman on a mission (= She is strongly determined) to teach those children to read.*

A mission is also a group of people who are sent to another place to do a particular job or to represent their country, organization, or religion, or the place where they go to do this work: *A trade mission was sent to South Africa.*

The Methodist mission (= place for those who are poor and need help) is located near the railroad yards.

Quelques dictionnaires font cependant une place à l'acception qui nous intéresse. Ainsi en va-t-il du *Merriam-Webster's Online Dictionary* : l'acception que nous cherchons se trouve au point 4c de l'entrée *mission*, où l'on voit apparaître discrètement l'expression cruciale de "statement of the company's mission" (voir ci-après).

4 a : a specific task with which a person or a group is charged **b** (1) : a definite military, naval, or aerospace task <a bombing *mission*> <a space *mission*> (2) : a flight operation of an aircraft or spacecraft in the performance of a mission <a *mission* to Mars> **c** : a preestablished and often self-imposed objective or purpose <statement of the company's *mission*>

5 : CALLING, VOCATION

Un autre dictionnaire, le *Cambridge International Dictionary of English* propose pour *mission* plusieurs entrées¹⁸ : "mission (JOB)", "mission (PEOPLE)", "mission control", "mission statement". Le texte correspondant à la première entrée mérite d'être reproduit car il montre l'emploi de "Our/My/Her mission is" dans la langue courante et non dans les déclarations d'organisations ou d'entreprises.

mission (JOB)

noun [C]

1 an important job, especially a military one, that someone is sent somewhere to do:

Your mission is to isolate the enemy by destroying all the bridges across the river.

a peace/rescue/fact-finding mission

2 any work that someone believes it is their duty to do:

My mission in life is to educate the rich about the suffering of the poor.

She's a woman with a mission and she's absolutely determined to finish the project.

On notera toutefois une certaine emphase, que marque le terme de “*duty*”, devoir, adhésion – ici, adhésion *personnelle*. La quatrième entrée, “mission statement”, nous permettra de mettre le doigt sur un élément clé.

mission statement noun [C]

a short written description of the aims of a business, charity, government department or public organization

Nous découvrons ici, en effet, que les déclarations rencontrées en interrogeant Google à propos de l’expression “our mission is” n’étaient en vérité rien d’autres que des “*mission statements*” – ce qu’on peut rendre en français par l’expression de *déclarations de mission* ¹⁹. L’enquête menée jusqu’ici avec l’aide d’Internet – ou plutôt de Google – peut entrer alors dans une nouvelle étape : celle de l’exploration des usages et de la signification de l’expression “*mission statement*”.

Une autre recherche avec Google : “mission statement”

Lorsqu’on interroge Google à propos de l’expression “mission statement”, les résultats présentés sur les deux premières pages se rangent grosso modo en deux catégories. La première concerne simplement des institutions qui explicitent leur “mission statement”. Le tableau suivant en fait apparaître un petit nombre.

- <http://www.weather.gov/mission.shtml> :

National Weather Service Mission Statement

“The National Weather Service (NWS) provides weather, hydrologic, and climate forecasts and warnings for the United States, its territories, adjacent waters and ocean areas, for the protection of life and property and the enhancement of the national economy. NWS data and products form a national information database and infrastructure which can be used by other governmental agencies, the private sector, the public, and the global community.”

- http://www.yorku.ca/web/about_yorku/mission/ :

Mission Statement

The mission of York University is the pursuit, preservation, and dissemination of knowledge. We promise excellence in research and teaching in pure, applied and professional fields. We test the boundaries and structures of knowledge. We cultivate the critical intellect. York University is part of Toronto: we are dynamic, metropolitan and multi-cultural. York University is part of Canada: we

encourage bilingual study, we value tolerance and diversity. York University is open to the world: we explore global concerns. A community of faculty, students and staff committed to academic freedom, social justice, accessible education, and collegial self-governance, York University makes innovation its tradition.

Énoncé de Mandat

L'Université York a pour mandat d'accroître, de préserver et de diffuser le savoir. Nous nous engageons à ce que la recherche et l'enseignement, qu'ils soient purs, appliqués ou professionnels, se placent à un degré d'excellence élevé...

- <http://www.mcgill.ca/oro/mission/> :

Mission statement

THE OFFICE OF RESEARCH OPPORTUNITIES (ORO) HAS BEEN MANDATED WITH A MULTI-FACETED MISSION FOR MCGILL'S FACULTY OF ARTS. ITS PRINCIPAL OBJECTIVES ARE: ...

...

La mission confiée au Bureau des Initiatives de recherche (BIR) par la Faculté des arts de l'Université McGill comporte de multiples volets dont les objectifs principaux sont : ...

- <http://www.fda.gov/opacom/morechoices/mission.html> :

FDA's Mission Statement

The FDA is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing the public health by helping to speed innovations that make medicines and foods more effective, safer, and more affordable; and helping the public get the accurate, science-based information they need to use medicines and foods to improve their health.

- <http://www.emc.org/body.cfm?id=86> :

Mission Statement

Our Vision

...

Our Mission

Eisenhower Medical Center, a not-for-profit organization, exists to serve the changing health care needs of our region by providing excellence in patient care with supportive education and research.

- http://www.motel6.com/about/careers/mission_values.asp :

Accor North America Mission Statement And Values

OUR MISSION

To establish Accor as the best-managed, most profitable hotel and services company in the U.S., Canada and Mexico.

To execute strategies which balance the needs of our customers, employees, investors and the communities in which we do business while respecting their diversity.

To be resourceful in overcoming obstacles, while being practical, cost conscious and realistic.

To grow Accor's presence in North America.

OUR VALUES

...

NUESTRA MISIÓN

Hacer de Accor North America, la empresa mejor administrada y la más rentable de las compañías hoteleras y de servicio de Canadá, los Estados Unidos, y México.

Alcanzar las metas respetando el equilibrio entre nuestros clientes, colaboradores y accionistas.

Optimizar nuestros recursos, permaneciendo pragmáticos y realistas en las acciones que emprendemos.

Hacer crecer la presencia de Accor en América del Norte.

NUESTROS VALORES

...

NOTRE MISSION

Devenir l'entreprise la mieux dirigée et la plus profitable des sociétés hôtelières et de service du Canada, des Etats-Unis et du Mexique.

Atteindre nos objectifs dans un esprit d'équilibre entre nos clients, nos collaborateurs et nos actionnaires.

Optimiser nos ressources et être pragmatique, réaliste dans les actions que nous entreprenons.

Développer la présence de Accor en Amérique du Nord.

NOS VALEURS

...

- <http://www.mil.be/ops-trg/subject/index.asp?LAN=fr&ID=935> :

Mission Statement

FLEXIBLE, MOBILE AND SWIFT IN ACTION

VISION

...

MISSION

Qui sommes-nous ?

En tant que “core business” de la Défense, le Département Opérations & Entraînement est le principal responsable de la mise en condition, de l’emploi et de l’engagement des Forces armées.

Quelle est notre mission ?

En tout temps, engager nos capacités avec succès, tant à l’étranger que sur le territoire national ; dans un cadre international, pouvoir participer avec toutes nos capacités à des engagements du type opérations de soutien de la paix ;

avec certaines de nos capacités et dans des « niches » opérationnelles bien définies, pouvoir participer, dans un cadre international, à un engagement de haute intensité.

...

VALEURS

...

On aura noté que les universités canadiennes York (Toronto) et McGill (Québec) énoncent leur mission en anglais et en français. Dans le premier cas, l’expression “mission statement” est traduite par « énoncé de mandat » (la mission devient un mandat), ce qui ne semble pas être une traduction canonique. Dans le cas du groupe Accor, on aura observé que l’expression figure en trois langues, l’espagnol se rajoutant à l’anglais et au français. On n’aura garde d’oublier le dernier “mission statement” de ce petit tableau : rédigé dans un langage mixe, il émane du ministère de la défense belge.

Une seconde catégorie de références est constituée de documents qui précisent ce qu’est un “mission statement” et, pour certains d’entre eux, qui s’emploient à donner des conseils à qui souhaite rédiger un tel énoncé de mission. Sur les résultats fournis par Google apparaît ainsi, sur la première page, en deuxième position, un lien conduisant à l’article de l’encyclopédie Wikipedia intitulé “Mission statement” : nous en reproduisons ci-après l’essentiel.

Mission statement

From Wikipedia, the free encyclopedia

A **mission statement** is a brief statement of the purpose of a company or religious or other organization.

Companies sometimes use their mission statement as an advertising slogan, but the intention of a genuine mission statement is to keep members and users aware of the organization’s purpose. In the case of public commercial companies, the primary purpose must always be to uphold the interests of shareholders, whatever the mission statement.

Examples of mission statements

■ “Provide society with superior products and services by developing innovations and solutions that improve the quality of life and satisfy customer needs, and to provide employees with meaningful work and advancement opportunities, and investors with a superior rate of return.” – Merck

■ “A computer on every desktop and in every home, running Microsoft software” – Microsoft

■ “Organize the world’s information and make it universally accessible and useful.” – Google

■ “Salem Methodist Church seeks to promote the spiritual, physical and educational well-being of all in the community of Smallthorne and neighbouring areas.” [...]

■ “The mission of the National Association of Evangelicals is to extend the kingdom of God through a fellowship of member denominations, churches, organizations, and individuals, demonstrating the unity of the body of Christ by standing for biblical truth, speaking with a representative voice, and serving the evangelical community through united action, cooperative ministry, and strategic planning.”

Stakeholder Conflict Resolution

The mission statement can be used to resolve differences between different business stakeholders. Stakeholders include: employees including managers and executives, stockholders, board of directors, customers, suppliers, distributors, creditors, governments (local, state, federal, etc.), unions, competitors, NGO’s, and the general public. Stakeholders affect and are affected by the organization’s strategies.

On découvre ainsi que le “*motto*” mis en pièce par BC ne serait rien d’autre que le “mission statement” de Google : “*organize the world’s information*”²⁰. D’autres documents précisent ce que doit être – et comment concevoir et rédiger – un “mission statement”. Dans le tableau qui suit nous donnons des extraits de deux d’entre eux : le premier concerne les entreprises (*for-profit organizations*), le second les organisations sans but lucratif (*not-for-profit* ou *nonprofit organizations*).

- <http://www.businessplans.org/Mission.html> :

Mission Statement

The mission statement should be a clear and succinct representation of the enterprise’s purpose for existence. It should incorporate socially meaningful and measurable criteria addressing concepts such as the moral/ethical position of the enterprise, public image, the target market, products/services, the geographic domain and expectations of growth and profitability.

The intent of the Mission Statement should be the first consideration for any employee who is evaluating a strategic decision. The statement can range from a very simple to a very complex set of ideas.

How Specific Should You Be?

Normally, the Mission Statement should represent the broadest perspective of the enterprise's mission.

You may want to take the approach of being very specific...

Mission Statements of Well Known Enterprises

The following are some examples of mission statements from real enterprises.

3M

"To solve unsolved problems innovatively"

Mary Kay Cosmetics

"To give unlimited opportunity to women."

Merck

"To preserve and improve human life."

Wal-Mart

"To give ordinary folk the chance to buy the same thing as rich people."

Walt Disney

"To make people happy."

These are the 'one-liners', but each is supported by a set of values that set the performance standards and direct the implementation of the mission.

For example, Merck, a company that produces pharmaceutical products and provides insurance for pharmacy benefits, publicly states the following values.

- ◆ Corporate social responsibility
- ◆ Unequivocal excellence in all aspects of the company
- ◆ Science-based innovation
- ◆ Honesty & integrity
- ◆ Profit, but profit from work that benefits humanity

And Walt Disney, an entertainment business states their values as follows.

- ◆ No cynicism
- ◆ Nurturing and promulgation of "wholesome American values"
- ◆ Creativity, dreams and imagination
- ◆ Fanatical attention to consistency and detail
- ◆ Preservation and control of the Disney "magic"

Should Your Grasp Exceed Your Reach?

Many believe that the Mission Statement should have a grand scale, be socially meaningful and be measurable. The following are some examples of historical Mission Statements that were truly grand in scale.

Ford Motor Company (early 1900's)

"Ford will democratize the automobile"

Sony (early 1950's)

"Become the company most known for changing the worldwide poor-quality image of Japanese products"

Boeing (1950)

"Become the dominant player in commercial aircraft and bring the world into the jet age"

Wal-Mart (1990)

"Become a \$125 billion company by the year 2000"

Conclusion

So, when you are preparing your Mission Statement remember to make it clear and succinct, incorporating socially meaningful and measurable criteria and consider approaching it from a grand scale. As you create your Mission Statement consider including some or all of the following concepts.

- ◆ The moral/ethical position of the enterprise
- ◆ The desired public image
- ◆ The key strategic influence for the business
- ◆ A description of the target market
- ◆ A description of the products/services
- ◆ The geographic domain
- ◆ Expectations of growth and profitability

Copyright © 1994-2007 Business Resource Software, Inc.

- <http://www.tgci.com/magazine/How%20to%20Write%20a%20Mission%20Statement.pdf> :

How to Write a Mission Statement

Every organization has a mission, a purpose, a reason for being. Often the mission is why the organization was first created – to meet a need identified years ago. Sometimes, the same problems that the organization initially tried to address continue to haunt generation after generation. In that case, the organization's purpose doesn't change – although how it does business has probably evolved. Other times, even 10 or 20 years can change the landscape so markedly that the original mission must be updated, altered, or changed dramatically in order to address those new realities.

That your organization's mission is current, alive, and well, however, doesn't necessarily mean that the organization has translated that purpose into a clear, concise mission statement. A good mission statement should accurately explain why your organization exists and what it hopes to achieve in the future. It articulates the organization's essential nature, its values, and its work.

This should be accomplished in a brief paragraph that is free of jargon and "terms of art." In other words, it should avoid the kind of shorthand that you may be in the habit of swapping with others who work in the field, but is unfamiliar to anyone who is outside the organization or the field in which it works.

Another important consideration is how recently your mission statement was reviewed by board or staff members. If it has been more than five years, now is probably a good time to review and, if necessary, fine-tune or even rewrite your mission statement. All too often an organization's mission statement, which has been handed down over the years, loses relevance and ceases to speak to staff, board members, or supporters.

An effective mission statement must resonate with the people working in and for the organization, as well as with the different constituencies that the organization hopes to affect. It must express the organization's purpose in a way that inspires commitment, innovation, and courage – not an easy task!

At the very least, your organization's mission statement should answer three key questions:

- 1. What are the opportunities or needs that we exist to address?** (the purpose of the organization)
- 2. What are we doing to address these needs?** (the business of the organization)
- 3. What principles or beliefs guide our work?** (the values of the organization)

You can begin the process of drafting a mission statement by creating a worksheet based on these questions. Ask staff, volunteers, and constituents to list any words, phrases, or ideas that come to mind with respect to the organization and these various categories. Do not edit at this point. Give everyone a chance to be heard. Look for language and concepts that enjoy broad consensus.

Here are three mission statements that do attempt to answer these questions. Let's see what they have in common.

1. The mission of Big Brothers/Big Sisters of America is to make a positive difference in the lives of children and youth, primarily through a professionally supported, one to one relationship with a caring adult, and to assist them in achieving their highest potential as they grow to become confident, competent, and caring individuals, by providing committed volunteers, national leadership and standards of excellence.

The purpose: to make a positive difference in the lives of children and youth so that they'll achieve their highest potential

The business: providing and supporting committed volunteers who have one to one relationships with children and youth

The values: individuals who are confident, competent, and caring; leadership and standards of excellence

2. The National Conference, founded in 1927 as the National Conference of Christians and Jews, is a human relations organization dedicated to fighting bias, bigotry, and racism in America. The National Conference promotes understanding and respect among all races, religions and cultures through advocacy, conflict resolution, and education.

The purpose: to fight bias, bigotry, and racism in America

The business: advocacy, conflict resolution, and education

The values: understanding and respect among all races, religions, and cultures

3. Planet 3000 is committed to healing the earth. Using research into natural ecosystems, Planet 3000 develops policy recommendations and pilot projects that apply these underlying principles to human ecosystems that are in harmony with other life on the planet. By bringing the human social order into balance with ecological principles, diversity of all living things can be sustained and the evolutionary process that has guided and nurtured life on this planet for millions of years can continue unabated.

The purpose: to “heal” the planet

The business: advocacy, research, and demonstration projects

The values: ecological principles; protecting balance, diversity, the evolutionary process, and harmony with life on the planet

Your Mission Statement Should...

- express your organization’s purpose in a way that inspires support and ongoing commitment
- motivate those who are connected to the organization
- be articulated in a way that is convincing and easy to grasp
- use proactive verbs to describe what you do
- be free of jargon
- be short enough so that anyone connected to the organization can readily repeat it

Excerpted from Strategic Communications for Nonprofit Organizations: Seven Steps to Creating a Successful Plan. Reprinted by permission of the publisher, John Wiley & Sons, Inc.

Comme on le voit, un “mission statement” a une valeur d’orientation autant pour l’extérieur de l’organisation que pour ceux qui agissent en son sein. Les documents précédents ne

doivent cependant pas faire oublier que, à côté de l'idée de “*our mission*” – concernant une “*organization*” –, il existe aussi “*my mission*”. Certains sites aident à formuler un tel engagement personnel. L'organisation FranklinCovey, qui, dans son site en français, se présente comme « acteur international de la formation au leadership et du conseil en alignement stratégique », énonce sa « mission » dans les termes suivants : « Nous aidons par la formation et le conseil l'ensemble des hommes de l'entreprise à se concentrer sur les priorités stratégiques de son organisation et à les exécuter. » Son site international en anglais propose un “*mission statement builder*” comportant un lien “Create Your Mission” ; si l'on suit ce lien, on parvient à une présentation dont l'une des diapositives, ornée de portraits de Gandhi, d'Einstein, de Martin Luther King et de Mère Teresa, porte ce texte²¹ : “A Mission: An inner urge to pursue an activity or perform a service. A calling.”

Des outils d'étude pertinents

Lorsqu'on étudie une question, lorsqu'on enquête à son propos, des outils sont nécessaires, notamment des outils de savoir. Lorsqu'elle s'attaque à l'expression “*our mission*”, BC semble, en vérité, relativement démunie à cet égard. On a vu qu'elle commentait le “*our*” de “*our mission*” en se référant à la notion de déictique et à la *Phénoménologie de l'Esprit*. Ajoutons que, pour le reste, si l'on omet une longue note sur l'adjectif « moral » (note où apparaissent Kant, Aristote, Nietzsche et où BC oppose morale et éthique), BC ne se prévaut – avec véhémence, certes – que de quelques éléments de première prise concernant les États-Unis, le billet vert, les discours de George W. Bush, etc.²². Ce qui fait défaut dans ce qu'elle propose au lecteur, nous espérons l'avoir montré, ce sont notamment des connaissances *spécifiques* à propos du monde économique américain, de sa culture, de ses usages. Il est vrai sans doute que ces connaissances participent de savoirs peu diffusés dans la culture française « générale », et restent confinés en quelques îlots voués au monde de l'entreprise.

Une recherche rapide conduit à constater que nombre d'ouvrages pour le grand public touchant à la vie des entreprises (par exemple les multiples « vocabulaires » de l'anglais ou de l'américain des affaires) se placent essentiellement du point de vue de l'employé, du cadre, à qui des missions peuvent certes être confiées, et non du point de vue de l'entreprise comme totalité. Cela noté, la pénétration de l'idée de mission statement semble faible : même les organismes d'aide à la création d'entreprises semblent l'ignorer (pour un exemple, voir <http://www.creeruneentreprise.fr/>). Dans la collection « Pour les Nuls » publiée par les éditions First, l'index de l'ouvrage Le

management pour les nuls (*Nelson & Economy, 2007*) n'attribue au mot mission qu'une seule occurrence, où le mot se révèle avoir le sens de mission confiée à un membre de l'entreprise par son chef. Ce n'est que dans l'ouvrage *Business plans pour les nuls* (*Tiffany & Peterson, 2005*) que le sujet du mission statement est véritablement abordé – tout un chapitre lui est consacré. Sur le site de l'éditeur, la présentation de l'ouvrage commence significativement par ces mots : « *Quelle est la mission de votre entreprise, quels objectifs vous êtes-vous fixés ? Quelles activités souhaitez-vous développer ?...* »

Ajoutons que, à l'évidence, le choix ou l'exclusion *a priori* de savoirs potentiellement utiles tend à être soumis au jeu des sympathies et des antipathies, prégnantes, ou au moins présentes, dans la culture où l'on baigne. À cet égard, on peut aisément imaginer que, s'agissant de BC, son sentiment la porte plutôt du côté des savoirs établis de l'univers philologico-philosophique dont elle est familière que des savoirs, en apparence plus insaisissables, touchant au capitalisme états-unien. Par contraste, l'enquête sur "*our mission*" dont nous avons rendu compte précédemment illustre une problématique 1) potentiellement émancipée de tout sentiment positif et négatif à l'endroit de la question étudiée et des objets auxquels elle se rapporte, 2) ne faisant pas fond uniquement sur des savoirs préalablement établis, qu'ils soient indigènes ou allogènes, 3) produisant (ou découvrant) des connaissances utiles, pertinentes, à l'aide (en l'espèce) d'outils offerts par le monde de l'Internet.

Complément d'enquête : déclarer sa mission est-il obligatoire ?

Pour souligner un peu plus cette problématique de production de connaissances *avec* Internet, nous nous arrêterons ici un instant sur une question subsidiaire que l'enquête à propos de "*our mission*" suscite assez naturellement : « déclarer sa mission » est-il légalement obligatoire pour une entreprise aux États-Unis ? Nous verrons en l'espèce que c'est là une question qui ne reçoit pas aisément une réponse sûre (si l'on utilise seulement les outils et matériaux accessibles sur Internet).

Comment, déjà, solliciter Google là-dessus ? Telle est bien la première difficulté²³. Il faut s'exprimer en anglais, et dans un domaine – en gros le droit des affaires aux États-Unis, et qui plus est en anglais américain – que l'on supposera, ici, mal connu, ainsi qu'il peut l'être dans la culture courante française. Commençons par essayer la requête "**corporate law**". À cela, Google répond comme on le verra sur la copie d'écran ci-après. On voit donc d'abord que l'expression fait sens en anglais ; et que l'encyclopédie Wikipedia, en tête de liste ici, lui consacre même un article. Malheureusement, le mot de "mission" n'a pas d'occurrence dans

cet article. Il en est de même dans les pages auxquelles conduisent les autres résultats proposés sur la première page.

The screenshot shows a Google search interface. At the top, the Google logo is on the left, and navigation links for 'Web', 'Images', 'Groupes', 'Actualités', and 'plus »' are on the right. Below these, a search bar contains the text 'corporate law', followed by a 'Rechercher' button and links for 'Recherche avancée' and 'Préférences'. Under the search bar, there are radio buttons for 'Web' (selected), 'Pages francophones', and 'Pages : France'. The main results section is titled 'Web' and shows 'Résultats 1 - 10 sur un total d'environ 1 890 000'. A tip suggests searching in French. The first four results are listed, each with a title, a brief description, and a link to the full page.

Web Résultats 1 - 10 sur un total d'environ 1 890 000

Conseil : [Recherche pour résultats en Français uniquement](#). Vous pouvez indiquer votre langue de recherche dans [Préférences](#)

Corporate law - Wikipedia, the free encyclopedia - [[Traduire cette page](#)]
Corporate law (also corporations law or company law) refers to the law establishing separate legal entities known as the company or corporation and governs ...
[en.wikipedia.org/wiki/Company_\(law\)](http://en.wikipedia.org/wiki/Company_(law)) - 74k - [En cache](#) - [Pages similaires](#)

Corporate Law - Guide to Corporate and Business Law - [[Traduire cette page](#)]
 Business can be undertaken in various forms. In many instances, ways of doing business are governed by law.
www.hg.org/corp.html - 60k - [En cache](#) - [Pages similaires](#)

Corporations - Wex - [[Traduire cette page](#)]
Corporate law has important intersections with contracts and commercial transactions ...
 Center for **Corporate Law** - University of Cincinnati College of Law ...
www.law.cornell.edu/topics/corporations.html - 23k - [En cache](#) - [Pages similaires](#)

Welcome to the Securities Lawyer's Deskbook - [[Traduire cette page](#)]
Corporate Law Center, University of Cincinnati College of Law - » Recent Posts from the Securities Law Prof Blog, Official Blog of the **Corporate Law** Center ...
www.law.uc.edu/CCL/ - 19k - [En cache](#) - [Pages similaires](#)

On peut alors tenter la requête suivante : "**corporate law**" AND "**mission statement**"²⁴. Celle-ci engendre, aux premières places, des résultats relatifs à des organisations ayant à voir avec la "corporate law", mais qui, simplement, précisent là leur "mission statement", ce qui explique la présence de cette dernière expression sur la page. Ainsi le premier lien renvoie-t-il au "mission statement" du *Center in Law, Economics and Organization* (CLEO) de la *School of Law* de l'Université "of Southern California".

The screenshot shows a list of search results. The first result is from the 'Center for Corporate Law' and includes a link to their 'Mission Statement'. The second result is from 'Corplaw' and also includes a link to their 'Mission Statement'. The third result is from the 'USC Gould School of Law' and includes a link to their 'Mission Statement'. The fourth result is from the 'Commercial Law Association of Australia' and includes a link to their 'Mission Statement'. Each result includes a title, a brief description, and a link to the full page.

Center for Corporate Law - [[Traduire cette page](#)]
 Center for **Corporate Law**. **Mission Statement** To develop programs relating to the education and training of law students in the field of **corporate law** and to ...
www.law.uc.edu/corplaw/index.html - 8k - [En cache](#) - [Pages similaires](#)

Corplaw Mission Statement - [[Traduire cette page](#)]
MISSION STATEMENT. The **Corporate Law** Society (CLS) is committed to preparing its members for successful careers in fields such as start-ups, general counsel ...
www.kentlaw.edu/student_orgs/corplaw/missionstatement.html - 7k - [En cache](#) - [Pages similaires](#)

USC Gould School of Law - [[Traduire cette page](#)]
CLEO Mission Statement. Interdisciplinary study is nothing new to the USC Law School ... including one on **corporate law** and another on behavioral economics, ...
law.usc.edu/academics/centers/cleo/mission.cfm - 21k - [En cache](#) - [Pages similaires](#)

Commercial Law Association of Australia - [[Traduire cette page](#)]
Mission Statement. The Mission of the Commercial Law Association of Australia Ltd is ... of knowledge of commercial and **corporate law** usages and practices; ...
www.cla.org.au/aboutus/mission.aspx - 26k - [En cache](#) - [Pages similaires](#)

Une petite enquête sur la manière d'exprimer l'*obligation* en anglais conduit ensuite à essayer la requête "**mission statement is mandatory**". Situation peu ordinaire, cette requête ne ramène que 5 résultats : on a reproduit ci-après la présentation qu'en donne l'unique page de résultats.

[Writing a Meaningful Mission Statement](#) - [[Traduire cette page](#)]
Scores of business planning and strategic experts will state a **mission statement is mandatory** for your company direction and fund raising. ...
[sbinformation.about.com/cs/businessplans/a/mission.htm](#) - 24k - [En cache](#) - [Pages similaires](#)

[tiger - IceRocket Tag](#) - [[Traduire cette page](#)]
Scores of business planning and strategic experts will state a **mission statement is mandatory** for your company ... 1 day 8 hours ago · 400 words · The ...
[blogs.icerocket.com/tag/tiger](#) - 72k - [En cache](#) - [Pages similaires](#)

[\[PDF\] Microsoft PowerPoint - JBR0506](#)
Format de fichier: PDF/Adobe Acrobat - [Version HTML](#)
Inclusion of **mission statement is mandatory**. • The more specific or constraining the corporate purposes (as restrictions in the ...
[www.carters.ca/pub/seminar/charity/2005/jbr0506pp.pdf](#) - [Pages similaires](#)

[\[PDF\] Care of civilians during military operations](#)
Format de fichier: PDF/Adobe Acrobat
A clear **mission statement is mandatory** to avoid mission expansion. There must be no ambiguity for Role/Echelon 3 regarding who is entitled to care and ...
[taylorandfrancis.metapress.com/index/G65324Q3P7722264.pdf](#) - [Pages similaires](#)

[Mission Statement Template](#) - [[Traduire cette page](#)]
A Corporate / Organizational **Mission Statement is mandatory**, from a sound business practice perspective, to inform employees, customers, and business ...
[www.sound-business-practices.com/rationale/mission-statements.php](#) - 16k - [En cache](#) - [Pages similaires](#)

Comme cette copie d'écran le suggère, le caractère « obligatoire » semble davantage moral que légal. Le premier lien²⁵ conduit au texte reproduit ci-dessous, qui confirme notre conjecture : certains experts pensent qu'un énoncé de mission est indispensable, d'autres que c'est inutile, l'auteur prenant parti – pour des raisons en partie intéressées... – en faveur de l'existence d'un énoncé de mission.

Writing a Meaningful Mission Statement

From Darrell Zahorsky,

Your Guide to Small Business Information.

FREE Newsletter. Sign Up Now!

Scores of business planning and strategic experts will state a mission statement is mandatory for your company direction and fund raising. Other advisors suggest writing a mission statement becomes a meaningless few sentences collecting dust somewhere in your office. Is a personal and corporate mission statement necessary for success in today's hostile business climate?

Do You Need A Mission?

The answer depends on whether or not the mission statement you compose has significant meaning to you, or is just another corporate exercise in futility. A mission statement can guide your company in good times and bad. A meaningful mission can act as a moral and corporate compass. It can help you make decisions aligning with your values and goals.

Speaker and author, Laurie Beth Jones of “The Path: Creating Your Mission Statement for Work and for Life” states, “It is the key to finding your path in life and identifying the mission you choose to follow. Having a clearly articulated mission statement gives one a template of purpose that can be used to initiate, evaluate, and refine all of one’s activities.”

Le deuxième résultat²⁶ conduit à retrouver le texte précédent. Le troisième résultat est une présentation PowerPoint (mise au format pdf) concernant la législation canadienne (et non pas états-unienne). La quatrième référence ne conduit pas, nous semble-t-il, à ce qui est annoncé²⁷. Le cinquième et dernier lien, en revanche, vaut d’être examiné plus attentivement. On a reproduit ci-après une partie du texte, ainsi qu’une image, figurant sur la page à laquelle il conduit²⁸.

Mission Statement Template - Mission Statement Formulation Package

Have you ever thought about what a *professional* difference it would make to your business, and / or your team, and / or your website, from a perception, a success, and a *sales* standpoint, if you had a **Mission Statement** to guide you?

Another way of looking at them is that they are part two of a two-part description of the overall business intent. Vision Statements, part one, describe the future: where you are going or where you want to go. While Mission Statements, part two, describe today: why you exist today and / or what you are doing to pursue your vision of the future. Together they provide direction for the business by focusing your attention on doing things day-to-day to accomplish your mission, while taking steps to pursue your vision of the future – your long-term business intent.

Note: From a chronological standpoint it is usually created first – out of need. The Vision Statement will usually be formulated after the business has progressed to the point that it’s viable and profitable. However, from a Business Planning or Business Model perspective the Vision Statement is usually listed first.

Without them organizations will flounder and pull in different directions. Employees must be clear about business intent if you want a team that’s focused on the priorities.

Our package contains several example Mission Statements, is easily customizable in almost any word processing software, and most of all, there is no waiting! This package is available via *download...* and you could be printing and publishing one in a matter of minutes.

Some of the positives that can result from establishing one are:

- ★ Greatly improved business focus
- ★ Everyone has the same corporate / organizational image - no misunderstandings
- ★ Enhances the professional perception of your business
- ★ Fosters a team oriented environment
- ★ Enhances employee morale
- ★ Helps attract and retain the best people

A Corporate / Organizational Mission Statement is mandatory, from a sound business practice perspective, to inform employees, customers, and business associates of who you are and what you do. Our package and set of examples is easily adaptable to literally any business type, and if you'd like some personal customization, i.e., your logo, your company name, you want us to prepare yours from scratch, etc. please see the custom service that we provide for this item... Mission Statement Service, and we'll get back to you, *OTHERWISE*, continue on...

Note:

1. Many of our visitors who *want* the Mission Statement Package come back to *get* the Vision Statement Package too, and vice versa. If you think you are that *type* of customer... we have a Mission/Vision Combo package where you can get them *both* at an *overall substantial reduction* in cost.

2. Depending on your *needs*, you might want to *take-a-look* at our Business Model Bundle which contains a *bundle* of Business Planning *necessities*, including this item. It will enable you to *broaden* your Business Planning thought process, and *lessen* the elapsed time for you to get your Business Model, which includes a **Mission Statement**, done.


Otherwise, to enable the above mentioned positive aspects for your business, there's no reason not to purchase this package today, and get the solution to *enable* the quick creation of your business Mission Statement.

Mission Statement Template - Mission Statement Formulation Package



Ce texte, qui vante les mérites d'un produit – le “*mission formulation package*”, ainsi d'ailleurs qu'un “*vision formulation package*” –, développe un argumentaire où il n'y a pas trace (et en tout cas pas d'utilisation commerciale) d'une quelconque législation contraignante. On peut conclure de tout cela que le fait de rédiger et de rendre public un “*mission statement*” est regardé, ici, comme relevant des “*sound business practices*”, mais nullement d'une obligation légale. C'est même le contraire qui paraît évident, au grand dam des vendeurs de semblables “*mission statement formulation packages*” ! On notera ainsi qu'on trouve dans le commerce – offert en telle grande librairie parisienne²⁹ – un ouvrage proposant “*301 Corporate Mission Statement from America's Top Companies*” (voir ci-après). On pourrait s'amuser des aspects à la fois cocasses et mercantiles de ces références. En tant qu'éléments de preuve, seul point de vue qui nous importe ici, elles concourent à alimenter et à soutenir notre conjecture : l'existence d'un “*mission statement*” n'a pas de caractère d'obligation légale au États-Unis, même si la pression à la fois commerciale et « éthique » peut être forte pour faire adopter une telle “*sound business practice*” supposée.

Notice



Prix indicatif
41.39 €

pour commander

Mission Statement Book: 301 Corporate Mission Statements from America's Top Companies (2nd Ed.)

Auteur(s) : ABRAHAM Jeffrey

Date de parution: 10-1999

Langue : ANGLAIS

486p. 22.2x15.2 Paperback

Etat : Disponible chez l'éditeur (délai de livraison : 14 jours)

Thèmes :

- ◆ Economie et gestion des entreprises / Organisation de l'entreprise / Ressources humaines

© Lavoisier 2000-2007

> Nouvelle recherche

> Voir le panier

Nous terminerons cette courte enquête par un document non moins révélateur³⁰ : on y verra le PDG de Yahoo! avouer qu'il *ignore* si l'entreprise qu'il dirige possède un énoncé de mission !

Analysis from FORTUNE: Plugged In Column archive

Yahoo's mission quest

It's not as memorable as 'Don't be evil,' but Yahoo's new mission statement shows it's ready to leverage its unique strengths, says Fortune's Adam Lashinsky.

FORTUNE Magazine

By Adam Lashinsky, Fortune senior writer

February 2 2007: 7:17 AM EST

NEW YORK (Fortune) -- Mission statements are a funny thing. They seem terribly important to the corporate pooh-bahs who craft them and then force their minions to memorize and swear to live by them. The rest of us rarely pay attention, no matter how often we're told about a given company's supposed credo.

And yet, now and again it's worth listening, either when a company changes its mission statement, or, worse, when one of the pooh-bahs can't necessarily articulate what the mission is.

Take the case of Yahoo (Charts), the Internet's most successful punching bag. The only way to think of Yahoo as less than one of the great media companies of our day is in relation to Google (Charts), which is truly one of the great media companies of our day. Yahoo's strategy, execution and stock price have been stuck for months, a victim of its poorly played competition with Google. Part of Yahoo's problem has been that for all its lucrative online advertising businesses, it botched search, the best online business of them all for several years running.

In an interview last summer with Yahoo CEO Terry Semel, I cited to him his arch-foe's disarmingly simple motto, "Don't be evil," and asked what Yahoo's motto is. After an uncomfortably long pause, Semel replied: "I don't know that we have a motto. Well, the mission of the company is, Deliver great value to our consumers and, basically, value them."

Wrong. As it happened, Yahoo did have a mission statement, even if Semel, chief executive for five years, didn't know it: "Our mission is to be the most essential global Internet service for consumers and businesses." That was broad stuff, and not bad as far as mission statements go. That Google had become far more essential for one of the most common Internet services consumers and business people seek -- search -- sort of derailed Yahoo's mission. Clearly it needed a new one.

Investors groan as Google grows

Now Yahoo has one. The company invited journalists to its Sunnyvale headquarters Tuesday for a presentation by three executives from its Santa Monica, Calif.-based media group. That's the outfit formerly run by ex-television executive Lloyd Braun and now run by Jeff Weiner, who has expanded his responsibilities from being Yahoo's top search executive.

At the beginning of his presentation, Scott Moore, Yahoo's head of news and information, flashed a slide of Yahoo's mission statement, which, it turns out, was quietly rolled out internally around the time the company restructured its management late last year.

The new line is equally broad. But it might fit a bit better than the old one. “Yahoo’s mission,” it reads, “is to connect people to their passions, their communities, and the world’s knowledge. To ensure this, Yahoo offers a broad and deep array of products and services to create unique and differentiated user experiences and consumer insights by leveraging connections, data, and user participation.”

As with seemingly everything else, Yahoo’s mission statement will suffer in comparison to Google’s, which is to “organize the world’s information and make it universally accessible and useful.” What’s good about Yahoo’s new message is that it clarifies the differences between Google and Yahoo. The geeks in Mountain View may be organizing the world’s information. The cool kids in Sunnyvale are connecting people to their passions. In other words, they’re in the entertainment business, a time-tested and durable way of making money.

The “deep array” of gizmos means Yahoo still tries basically anything that works online – as increasingly does Google as well. The line about leveraging connections, data and user participation, however, truly says something about how Yahoo sees itself. Its e-mail program is still the most popular online, it is increasingly willing to use its data as a come-on for advertisers, and by user participation, Yahoo is referring to its Yahoo Answers, Flickr and other social-networking services that give the 10-plus-year-old company some true hipness in a Web 2.0 world.

Wall Street’s hope for Yahoo is that if it can deliver with its new search-advertising platform, Panama, it can at least hold its own against Google and begin to leverage its unique strengths. If Yahoo’s articulation of its own new mission statement is any indication, it might have half a chance.

CHAPITRE 3

Égarements et approximations de l'enquête

Les contre-enquêtes précédentes montrent que, même si la notion de “mission statement” porte une certaine charge eschatologique (d’origine religieuse très certainement), le fait de *déclarer sa mission* – “*our mission is...*” – est aujourd’hui *banal* de la part des « organisations » états-uniennes. À cette exigence sinon légale, du moins morale, Google répond donc en définissant sa mission ainsi : “*Organize the world’s information and make it universally accessible and useful*”.

En philologue de profession et de passion, BC s’attaque – après “*our mission is...*” – à l’utilisation du verbe « organiser », *to organize*³¹. Ici, nous l’avons vu, se produit un événement sans doute inattendu pour le lecteur qui, par rapport à la « tribu » de BC, demeure résolument un *outsider* : BC fait brusquement appel à un « savoir » produit, à propos du verbe « organiser » (*organisieren*) par le philologue allemand Victor Klemperer dans des circonstances exceptionnellement tragiques, et à vrai dire dans un tout autre cadre historique, culturel et politique. Ce geste de transfert présente un risque certain : celui d’exporter dans un contexte bien différent les résultats d’un travail philologique qui porte moins sur les mots *organiser* et *organisation* eux-mêmes que sur l’emploi qui en fut fait dans le jargon proliférant de l’Allemagne nazie, cadre surdéterminé dans lequel, on l’a vu, Klemperer oppose *organisation* et *système*.

Organiser, dites-vous ?

« Chaque mot est un préjugé », remarquait Nietzsche, qui voyait en cela le « danger du langage pour la liberté de l’esprit »³². Certes. Mais les préjugés que portent en eux les mots varient. La détestation exprimée par BC à l’encontre du mot d’organisation va sans doute au-delà des analyses à chaud de Klemperer lui-même. Surtout, il ne semble pas qu’une telle détestation, hier justifiée, soit aujourd’hui partagée par ceux-là mêmes qui portent sur Google les regards les plus critiques. C’est ainsi que, dans son livre (que BC ne fait que mentionner en passant), Jean-Noël Jeanneney utilise le verbe *organiser* et le substantif *organisation* sans

façon, comme si de rien n'était, ce qui a peut-être déterminé en partie la discrétion³³ de BC à propos d'un livre dont le chapitre 6 est précisément intitulé « Organiser le savoir » et s'ouvre par ces mots³⁴ : « Dans mon esprit, il s'agit moins de rêver à une exhaustivité, toujours utopique, que d'aspirer à la plus riche, la plus intelligente, la mieux *organisée*, la plus accessible des sélections possibles. » (Jeanneney, 2006, p. 115). Ce même auteur, pour qui le « vrac » est le « danger absolu », écrit un peu plus loin (p. 121) : « Le progrès de la civilisation peut être analysé – entre autres – comme la réduction des forces du hasard au profit d'une réflexion qui sache s'enrichir d'une connaissance *organisée*. » Il faut donc, ajoute-t-il (p. 124), « améliorer les principes *organiseurs* des corpus ou les chemins qui permettent de s'y promener ». Quelques lignes plus loin (p. 125), il évoque ensuite son souhait d'un « système d'*organisation* de la connaissance », qu'il veut, certes, non « immuable ». À cet égard, il évoquera élogieusement le moteur de recherche français Kartoo.com, qui, précise-t-il, permet à l'utilisateur de « *réorganiser* autour [d'un "concept"] une série de sites qui s'y rattachent » (p. 125). Le projet qu'il esquisse va cependant bien au-delà, comme il appert dans les lignes suivantes (p. 126).

... il faudra trouver le moyen que soient fournies à l'enquête non seulement, comme on l'a dit, les informations *organisées*, mais même des indications sur leurs limites et simultanément des pistes conduisant à d'autres richesses, celles des fonds traditionnels, notamment grâce aux catalogues qui autorisent qu'on s'y aventure utilement.

Dans le chapitre qui suit (le chapitre 7, intitulé « Projet culturel, projet industriel »), Jean-Noël Jeanneney énonce alors cette « question primordiale » : « comment mettre à disposition des richesses intelligemment choisies et utilement *organisées* en corpus ? » (p. 130). Cet emploi désinhibé de ressources lexicales que BC proscriit ne s'arrête pas, dans le texte de Jean-Noël Jeanneney, à l'organisation de la connaissance (ou de l'information). C'est ainsi qu'il assigne aux politiques le devoir de « définir une *organisation* propre à réguler le système, à en canaliser l'inventivité tout en la défendant contre des dérives » (pp. 130-131). (On aura noté, ici, la co-présence solidaire des notions d'*organisation* et de *système*, que Klemperer distingue et que BC oppose.) De même, Jeanneney écrira quelques pages plus loin (p. 138) : « On *organiserait* de la façon la plus souple possible, en évitant à tout prix de mettre en place des organismes pesants, la coordination de ces réflexions, pour fixer les grandes lignes d'une stratégie collective... ». En même temps, la question de l'organisation *de la connaissance* demeure un thème récurrent. Revenant sur le danger qu'il pressent dans l'évolution actuelle des choses, l'auteur condense en quelques mots le cœur de la doctrine qu'il défend (p. 138) :

« Ainsi débattrait-on de concert de l'*organisation* en corpus dont j'ai développé plus haut la philosophie – dans l'obsession de répondre à cette inquiétude lancinante du n'importe-quoi, de la dispersion du savoir en poudre dont nous sommes menacés. »

Ainsi que nous l'avons fait à propos de "*our mission is...*", on peut effectuer une recherche sur Internet pour apprécier la fréquence d'apparition de la formulation "*Our mission is to organize...*". Adressée à Google à la date du 1^{er} septembre 2007, la requête correspondante suscite une page de résultats dont les premiers ont trait à la « mission » de... Google lui-même (voir ci-après).

[Google-moi : la deuxième mission de l'Amérique: un nouveau livre ...](#)
Our mission is to organize all the information in the world ("Notre mission est d'organiser toute l'information dans le monde") ; Don't be evil ("Ne sois ...
[www.trackbusters.fr/actualites/20070323-livre-sur-google-barbara-cassin-google-moi.html](#) - 32k
- [En cache](#) - [Pages similaires](#)

[Pascal Rossini: Commissions versées par Google Adsense aux ...](#)
Our mission is to organize the world's information and make it universally accessible and useful. We believe that the most effective, and ultimately the ...
[pascalrossini.blogspot.com/2006/01/commissions-verses-par-google-adsense.html](#) - 39k -
[En cache](#) - [Pages similaires](#)

[Google Video Viewer Free Download- Our mission is to organize the ...](#) - [[Traduire cette page](#)]
Google Video Viewer Google Video Viewer allows you to view videos from Google Video Search. **Our mission is to organize** the world's information and make it ...
[www.softs.info/download/37/2302/Google_Video_Viewer.html](#) - 46k -
[En cache](#) - [Pages similaires](#)

[Christian Fauré » Blog Archive » La mission de Google](#)
"Our mission is to organize the world's information and make it universally accessible and useful" "Organiser l'information à l'échelle mondiale", ...
[www.christian-faure.net/2005/11/06/la-mission-de-google/](#) - 18k - [En cache](#) - [Pages similaires](#)

[Christian Fauré » Blog Archive » Ce que veulent les consommateurs](#)
La mission de Google; "Our mission is to organize the world's information and make it universally accessible and useful" "Organiser l'information à ...
[www.christian-faure.net/2007/01/05/ce-que-veulent-les-consommateurs/](#) - 16k -
[En cache](#) - [Pages similaires](#)
[[Autres résultats, domaine www.christian-faure.net](#)]

[GlobalWarmingSolution.Org: Our Network](#) - [[Traduire cette page](#)]
Our mission is to organize and maintain a diverse grassroots organization dedicated to the improvement and preservation of the environment through education ...
[www.globalwarmingsolution.org/about%20us/info.html](#) - 28k - [En cache](#) - [Pages similaires](#)

La première référence se rapporte au livre de BC. Les quatre références suivantes renvoient au "*mission statement*" de Google, que seule la cinquième référence ne concerne pas : elle renvoie à un site créé en 2003, *GlobalWarmingSolution.Org*. La formulation recherchée n'apparaît pas dans l'énoncé de mission de cette organisation³⁵ ; en revanche, elle apparaît bien sur la page à laquelle on arrive en cliquant sur le lien proposé³⁶, où figure les URL et une présentation succincte de diverses organisations écologistes, dont l'*Ohio Valley Environmental Coalition*, qui énonce sa mission dans les termes suivants.

The Ohio Valley Environmental Coalition, formed in 1987, is a nonprofit organization. Our mission is to *organize* and maintain a diverse grassroots organization dedicated to the improvement and preservation of the environment through education, grassroots organizing and coalition building, leadership development and media outreach. Our work encompasses much of West Virginia and portions of southern Ohio and eastern Kentucky (<http://www.ohvec.org>).

Sur la première page affichée par Google figurent encore quatre résultats (non reproduits ci-dessus). Le deuxième et le quatrième d'entre eux concernent Google. Le troisième envoie sur une page du site de la *Secular Student Alliance*, où on lit ³⁷ : “Our mission is to *organize*, unite, educate and serve students and student communities that promote the ideals of scientific rationality, secularism, democracy, and human based ethics.” Le premier de ces quatre résultats ³⁸ correspond au site d’une “nonprofit corporation” de Floride animée par un couple à l’allure sympathique, Miles et Mimi Coon, qui précise : “part of our mission is to *organize* educational outreach programs that bring the pleasures of poetry to our community. We are a proud member of the Palm Beach Country Cultural Council.”

Pour obtenir plus encore d’autres références que celle à Google, on peut proposer la requête **"our mission is to organize" -Google**. Sur la première page de résultats affichés, les deux premiers nous sont déjà connus : ils renvoient au site de *GlobalWarmingSolution.Org* puis à celui du *Palm Beach Poetry Festival*. Voici alors les énoncés de mission non encore rencontrés jusqu’ici et qui figurent sur les deux premières pages affichées par Google (dans ce qui suit, c’est nous qui soulignons l’expression “*Our mission is to organize*”).

- *Our mission is to organize* and develop resources to help Elders who are being abused in any way. (<http://www.elderlyabuse.com/missionstmt.htm>)
- *Our mission is to organize* all workers, educate all workers and raise the standards of collective bargaining in our nation. (<http://local137.com/> : site de la “Local Union 137” de la “Sheet Metal Workers International Association”)
- *Our mission is to organize* a annual conference to celebrate and recognize Vietnamese American leaders throughout the country. (<http://lindavunguyen.com/?cat=3> : personne privée qui fait connaître l’action du *Vietnamese American National Gala* (VANG), dont le site a pour adresse <http://www.vangusa.org/>)
- In the NTEU Mission Statement, *our mission is* “to *organize* federal employees to work together to ensure that every federal employee is treated with dignity and respect.” (<http://www.nteu.org/ExternalOrganizing/> : la NTEU est la “National Treasury Employees Union”).

- *Our mission is to organize* volunteers at Huron who will support other Food Not Bombs activists in reducing waste and sharing free food and information in protest of war and poverty. (<http://huronglobe.net/committees/committees.pdf>)
- *Our mission is to organize* and implement all SWTC youth activities in order to promote real-life events, which inform, encourage, instruct, excite and enable the making of disciples. (<http://www.reachingyouth.org/ccym.html> : site reachingyouth.org de la “Southwest Texas Conference” [SWTC] de la “United Methodist Church”).
- *Our Mission is to organize* a foundation of fathers as role models in applying positive actions for safer neighborhoods and communities free from drugs. (<http://www.volunteersolutions.org/org/dir-d-1.html>)
- *Our mission is to organize* a volunteer force of Kentuckians who want to be a part of improving care in nursing homes, and who want to help educate the public about the problems in nursing homes, mainly the shortage of front-line caregivers. (<http://www.kynursinghomereform.org/mission.html>)
- *Our mission is to organize*, encourage, and educate mental health consumers throughout Texas. Our vision is the integration of individuals diagnosed with mental illness into the community as productive, functional, valued, citizens. (<http://www.tnhc.org/>)
- *Our mission is to organize*, represent and mentor students preparing for initial licensure as registered nurses, as well as those nurses enrolled in baccalaureate completion programs; to promote development of skills needed to be responsible and accountable members of the nursing profession; to promote community service through organized activities; and to advocate high quality health care. (<http://www.gmu.edu/org/nursingsa/mission.htm> : site de la “Student Nurses’ Association” de la George Mason University à Fairfax en Virginie.)
- *Our mission is to organize* and conduct fundraising events driven by children and to donate these funds to organizations directly involved in the research, prevention or treatment of serious childhood diseases, such as diabetes, cancer, and AIDS. (<http://www.kids-kount.org/> : Kids Kount se présente comme “a charitable foundation of kids caring for kids”).
- *Our mission is to organize* seminars, workshops for pupils/students in the schools, higher institutions and the public. To organize and execute a national and international yearly conference and to acquire and build facilities for training and research on peace. To educate and work with the general public on and about dangers/disadvantages in conflicts, crisis, demonstrations and rebellious acts. (<http://www.nicped.interconnection.org/mission/> : site du “Nigerian Center for Peace Education” [NICPED])

- *Our mission is to organize* small groups of interested owners into single partnerships in order to maximize and diversify everyone's investments which will create the maximum earning potential possible for all involved. (<http://www.limespyder.com/page/page/3257254.htm> : site du "Lime Spyder Racing Team", qui se décrit comme "the perfect opportunity for those who have always dreamed of becoming an owner involved in the exciting world of top class harness racing" ["harness racing" : trot attelé ou, parfois, trot monté].)
- *Our mission is to organize* grassroots non-profit and non-governmental organizations from a diversity of disciplines, and to support and coordinate these organizations' efforts to eliminate poverty, its causes and symptoms, in developing mountain communities around the world. (http://www.mountainfund.org/html_site/ : site du "Mountain Fund", qui se présente comme "a U.S. I.R.S. 501(c)(3) Charitable Organization".)
- *Our mission is to organize*, train, effectively represent and elevate all Carpenters. (<http://www.ubc247.com/front.htm> : site de l'union locale 247 de la "United Brotherhood of Carpenters and Joiners of America".)
- *Our mission is to organize*, promote and administer interdisciplinary research among faculty and students on the social, political, historical, cultural, and economic experiences of people of African descent. (<http://research.ucsb.edu/cbs/xsite/annual/ar00/directors.html>)
- *Our mission is to organize* fun and fulfilling scientific and social gatherings for students. (<http://uwaaps.blogspot.com/> : site du "AAPS Student Chapter", University of Wisconsin at Madison, School of Pharmacy ; l'AAPS est l'"American Association of Pharmaceutical Scientists".)

Nous ne poursuivrons pas plus avant cette amorce d'enquête, qui suffit à montrer que l'expression "*our mission is to organize*" n'est en rien l'apanage de Google.

Pour approfondir cette conclusion, on peut encore se demander si le verbe *to organize* est ou non fréquemment employé après l'expression "*our mission is*". Pour tenter de répondre, on a procédé de la façon suivante. On a d'abord présenté au moteur de recherche Google la requête "**our mission is to**" -Google. On a alors relevé les verbes apparaissant après l'expression "*our mission is to*" – en se limitant aux dix premières pages de résultats affichées par Google. Parmi ces verbes, il est notable que, en l'espèce, *to organize* n'apparaît jamais à la suite de "*our mission is*". Pour chacun des verbes qui apparaissent, on a alors fait afficher les résultats et leur effectif (tel que le fournit Google) dans lesquels apparaît le verbe en question après l'expression "*our mission is*". On obtient alors les résultats consignés dans le tableau ci-après. L'effectif médian se situe au 24^e rang : il est obtenu par "*encourage*". On voit immédiatement que "*organize*", qui vient au 23^e rang, est assez bien classé : en aucun

cas on ne peut donc dire – à suivre les indications chiffrées données par Google ³⁹ – que ce verbe serait *rare* dans le lexique des énoncés de mission.

<i>Our mission is to...</i>					
1	... subvert	3	25	... foster	37 600
2	... honor	519	26	... enable	38 300
3	... extend	533	27	... increase	39 400
4	... cultivate	641	28	... enhance	58 800
5	... demystify	648	29	... empower	61 800
6	... sell	748	30	... meet	64 500
7	... represent	773	31	... give	68 000
8	... conceptualize	852	32	... assist	70 600
9	... guide	920	33	... educate	72 100
10	... surprise	2 640	34	... support	75 600
11	... liberate	2 820	35	... ensure	96 000
12	... surround	3 400	36	... develop	109 000
13	... devise	4 040	37	... serve	115 000
14	... fly	10 000	38	... offer	137 000
15	... enrich	15 000	39	... find	150 000
16	... design	16 700	40	... deliver	151 000
17	... keep	19 700	41	... promote	192 000
18	... use	20 300	42	... create	229 000
19	... prepare	23 600	43	... be	243 000
20	... welcome	26 800	44	... help	413 000
21	... preserve	29 900	45	... build	779 000
22	... inspire	32 300	46	... improve	1 250 000
23	... organize	32 700	47	... provide	2 080 000
24	... encourage	37 300			

Qu'en conclure ? Il est sûr que la *LTI* a fait un usage ubiquitaire d'*organisieren* et que, à suivre Klemperer, cela ne fut pas un tic de langage inspiré par le seul hasard, le caractère « organique » (*organisch*) constituant à l'évidence une valeur sûre au panthéon de l'idéologie du III^e Reich (Klemperer, 1947/1996, p. 141). Il est non moins certain que la diffusion du mot, sa présence récurrente dans l'usage quotidien ont dû entretenir, pour les idéologues du

régime, un lien fort avec la doctrine nazie. Ce lien était-il évident pour le citoyen ordinaire ? La chose n'est pas claire ; car tous les locuteurs n'avaient sans doute pas la sensibilité et la sagacité linguistiques d'un Klemperer. Bien des années après les tragiques événements auxquels Klemperer survécut, un témoignage l'illustre parfaitement : dans son blog, à la date du 15 janvier 2005 et sous le titre *Les mots des déportés*, Pierre Assouline écrit ces lignes qui rappellent l'horreur d'alors.

À quoi reconnaît-on un ancien déporté quels que soient ses origines, son engagement ou son itinéraire ? Pas à son numéro tatoué car la majorité n'en ont pas. Alors ? À son vocabulaire. Aux mots qu'il continue à employer soixante ans après. Un ancien déporté dira encore *läger* pour « camp », *revier* pour « infirmerie », *sonderkommando* pour « équipe spéciale », *musulmans* pour « morts-vivants »...

J'y repensais l'autre soir en écoutant le témoignage de Simone Veil à la télévision. Elle a employé à quatre ou cinq reprises le verbe « organiser » dans un sens inintelligible pour les téléspectateurs. Car dans l'argot du camp, ce terme essentiel de leur vie quotidienne signifiait « voler », « se débrouiller »...

Dans ce symptôme linguistique dont Simone Veil demeure marquée si longtemps après, on doit certainement entendre d'abord un reste inaltéré de l'abjection nazie⁴⁰. Mais on va voir qu'on peut aussi y entendre un fonctionnement « ordinaire » de certaines langues européennes⁴¹.

Élargir l'enquête...

À nouveau, une enquête sur la langue anglaise permettrait de mettre au jour certaines permanences dont on peut douter qu'elles soient le simple écho ou le pur vestige de l'ordre hitlérien. Dans le dictionnaire *Robert & Collins* (Atkins et al., 1987), on trouve ainsi l'exemple suivant, qui ne semble pas correspondre à un usage courant, en français, du verbe *organiser* : “*can you organize some food for us?*” Cette formulation, les rédacteurs du dictionnaire ont d'abord tenté de la rendre en français par « est-ce que vous pouvez nous organiser de quoi manger ? », avant de proposer au lecteur, non plus « nous organiser de quoi manger », mais « nous arranger une collation ». Ici, au reste, l'occurrence de *organize* est suivie d'un astérisque dont le dictionnaire cité commente l'usage en ces termes (p. XX) :

* indicates that the expression, while not forming part of standard language, is used by all educated speakers in a relaxed situation but would not be used in a formal essay or letter, or on an occasion when the speaker wishes to impress.

Y aurait-il une circulation souterraine de certains emplois marginaux d'*organiser*, de *organisieren*, ou de *organize* ? Rappelons une anecdote rapportée par Klemperer, qui semble l'avoir marqué, et que BC ne manque pas de citer dans son livre (Klemperer, 1947/1996, p. 143 ; Cassin, 2005, p. 93) : « En 1936 déjà, un jeune mécanicien qui, à lui seul, était venu à bout d'une réparation délicate et urgente sur mon carburateur me dit ceci : "N'ai-je pas bien organisé ça ?" » Klemperer propose de cet épisode une interprétation dont nous n'avons pas de raison de douter, s'agissant du contexte global où l'épisode rapporté s'est produit (Klemperer, 1947/1996, pp. 143-144) ; à propos du « jeune mécanicien », il note ceci.

Il avait tellement dans l'oreille les mots « organisation » et « organiser », il avait été tellement gavé de l'idée que tout travail devait d'abord être « organisé », c'est-à-dire réparti par un coordonnateur entre les membres d'un groupe discipliné, que pour la tâche qui lui était propre et qu'il avait achevée tout seul, aucun des mots simples et pertinents, comme « travailler » ou « accomplir » ou « exécuter » ou tout simplement « faire », ne lui étaient venus à l'esprit.

Il est pourtant curieux de constater l'apparition de tels emplois d'*organiser* dans des contextes européens forts différents. Le dictionnaire *Robert & Collins* (Atkins et al., 1987) propose ainsi, à la suite de l'exemple déjà mentionné, cet exemple tout proche : *can you organize the food for us?* – « est-ce que vous pouvez vous occuper de la nourriture ? » Ce qui émerge là, c'est que le fait « d'organiser » suppose un effort, une capacité à résoudre des problèmes, au fond une certaine *débrouillardise*. Le *Cambridge Advanced Learner's Dictionary* (Walter et al., 2006) propose, à l'entrée *organize*, l'exemple suivant (précédé de la mention *UK*, qui semble indiquer un usage tout spécialement britannique) : *She had organized a car to meet me at the airport*. De même, tel dictionnaire accessible en ligne⁴² propose ceci : "I'll organize coffee and sandwiches for the delegates." Tel autre⁴³ précise à propos du verbe *to organize* : "To arrange, provide or prepare • *will organize a meal* • *organized the tickets*." De la simple *débrouillardise*, dans un contexte ordinaire de la vie, on imagine alors qu'on puisse passer aisément à un usage plus trouble de l'art d'organiser le monde autour de soi. C'est ainsi que, écho inattendu aux propos que rapportait Pierre Assouline, la consultation du *Harrap's New Standard French and English dictionary* en quatre volumes (Mansion et al., 1980) livre

significativement ceci : “*to organize a bottle of rum*, s’emparer d’une bouteille de rhum ; dénicher, *esp.* voler, une bouteille de rhum.”

Que conclure des développements précédents ? Lorsqu’on tente d’élucider une difficulté – ici, clarifier ce que porte en elle, éventuellement, la formulation “*our mission is to organize...*” –, on doit recourir parfois à des outils de savoir non triviaux, et non trivialement disponibles. Dans le cas que l’on vient d’examiner, BC a cru pouvoir faire fond sur un savoir produit dans un contexte très spécifique par l’un de ses prédécesseurs en philologie, Victor Klemperer. Or il n’est pas évident que l’enquête et l’analyse dus à Klemperer ne doivent pas être reprises *à nouveaux frais* à propos de l’énoncé de mission de la firme Google – reprise dont BC, s’agissant de “*organize*” tout au moins, fait purement et simplement l’économie. Les quelques pas que nous avons faits dans cette direction nous portent à penser, au contraire, qu’une telle reprise n’est pas vaine et que, dans sa réalisation, le secours des ressources de l’Internet peut jouer un rôle pertinent et, à certains égards, *crucial*⁴⁴. À titre d’illustration complémentaire, nous nous arrêterons maintenant sur ce qui, dans l’énoncé de mission de Google, vient après “*our mission is to organize...*”, et que BC vitupère plus encore peut-être que “*our mission*” et “*organize*”.

“All the information in the world” ou les aléas d’une enquête

Le chapitre III de l’ouvrage de BC, sur lequel nous nous sommes arrêtée dans les développements qui précèdent, porte un titre que l’auteure met entre guillemets pour signifier que – selon elle – il s’agirait d’une citation : “*our mission is to organize all the information in the world*”. Ce chapitre commence d’ailleurs par les lignes suivantes (Cassin, 2005, p. 81).

“*Our mission is to organize all the information in the world.*” “*Don’t be evil.*” Les deux grands secrets de Google sont quant à leur principe aussi visibles que possible, puisqu’ils s’incarnent dans les deux mots d’ordre : organiser et faire le bien.

Tout cela pose problème. La formule “*don’t be evil*” assignée à Google est, elle, incontestablement authentique. Dans une célèbre interview donnée en 2004 au magazine *Playboy* par Sergey Brin et Larry Page, on trouve par exemple l’échange reproduit ci-après.

PLAYBOY: Is your company motto really “Don’t be evil”?

BRIN: Yes, it’s real.

PLAYBOY: Is it a written code?

BRIN: Yes. We have other rules, too.

PAGE: We allow dogs, for example.

BRIN: As for “Don’t be evil,” we have tried to define precisely what it means to be a force for good –always do the right, ethical thing. Ultimately, “Don’t be evil” seems the easiest way to summarize it.

PAGE: Apparently people like it better than “Be good.”

BRIN: It’s not enough not to be evil. We also actively try to be good.

PLAYBOY: Who ultimately decides what is evil? Eric Schmidt, your CEO, once said, “Evil is whatever Sergey decides is evil.”

PAGE: That was not one of his best quotes, though it’s memorable.

Pour sa part, dans son livre, John Battelle raconte ainsi l’origine de cette devise (Battelle, 2006, pp. 138-139).

On July 19, 2001, about a dozen early employees met to mull over the founders’ directive. Joan Braddi, now vice president of search services, was there, as were David Krane, director of corporate communications; and Amit Patel, an engineer and employee number seven. Sullivan moderated the discussion, which began with the assembled group listing the core principles that they believed represented what Google was all about. The meeting soon became cluttered with the kind of easy and safe corporate clichés that everyone can support, but that carry little impact: Treat Everyone with Respect, for example, or Be on Time for Meetings.

The engineers in the room were rolling their eyes. Patel recalls: “Some of us were very anticorporate, and we didn’t like the idea of all these specific rules. And engineers in general like efficiency – there had to be a way to say all these things in one statement, as opposed to being so specific.”

That’s when Paul Buchheit, another engineer in the group, blurted out what would become the most important three words in Google’s corporate history. “Paul said, ‘All of these things can be covered by just saying, Don’t Be Evil,’ ” Patel recalls. “And it just kind of stuck.”

It more than stuck; it became a cultural rallying call at Google, initially for how Googlers should treat each other, but quickly for how Google should behave in the world as well. It helped that in the months after the meeting, Patel scribbled “Don’t Be Evil” in the corner of nearly every whiteboard in the Company. For an organization consisting mainly of engineers, whiteboards served as the corporate equivalent of the water cooler. The message spread, and it was embraced, especially by Page and Brin. “The phrase captured what we all inherently felt was already true about the Company,” Krane recalls. “It was the lyrics written over a melody that already existed.”

“I think it’s much better than Be Good or something,” Page jokes. “When you are making decisions, it causes you to think. I think that’s good.”

Qu’en est-il alors de ce qui n’est pas un *motto* (une devise) mais bien, on l’a vu, le *mission statement* même de la firme ? Dans le document “*filed with the Securities and Exchange Commission on August 13, 2004*”, jour où Google “*goes public*” (entre en bourse), on lit par deux fois la formule consacrée (Google, 2004) : “*Our mission is to organize the world’s information and make it universally accessible and useful.*” Dans leur « lettre à Wall Street » intitulée *Letter from the Founders. “An Owner’s Manual” for Google’s Shareholders*, Brin et Page écrivent encore⁴⁵ : “We also believed that searching and organizing *all the world’s information* was an unusually important task that should be carried out by a company that is trustworthy and interested in the public good.” La formulation “all the information in the world” a-t-elle jamais été employée par la firme Google ou par quelques-uns de ses membres les plus éminents ?

Deux points sont sûrs. Tout d’abord, l’énoncé de mission (*mission statement*) de Google *n’est pas*, formellement, ce que BC semble dire : ainsi qu’on vient encore de le voir, cet énoncé de mission se réfère simplement à “*the world’s information*” là où BC mentionne “*all the information in the world*”⁴⁶. Ensuite, il est peu douteux que les leaders de la firme Google ont effectivement eu à l’esprit, au moins dans une première période, l’idée d’organiser *toute* l’information du monde : il y a bien, pour eux, équivalence entre “*the world’s information*” et “*all the world’s information*”. Pour ne prendre ici que quelques exemples complémentaires, dans un article paru le 9 août 2006 sur le site ZDNet, l’auteure, Donna Bogatin, rapporte des propos qu’auraient tenus lors d’une réunion publique Eric Schmidt, qui préside aux destinées de Google ; elle écrit notamment ceci : “The Google mission to ‘organize *all the world’s information* and make it universally accessible’ informs all that Google does, according to Schmidt.” De même donne-t-elle à son article le titre suivant : “Google CEO Eric Schmidt: ‘*all world’s information* includes personal information’.” Dans un article plus récent, signé de Jeffrey Toobin et paru dans *The New Yorker* sous le titre “Google’s Moon Shot. The quest for the universal library”, Marissa Mayer, vice-présidente de Google en charge du projet de numérisation de livres, déclare : “Google has become known for providing access to *all of the world’s knowledge*, and if we provide access to books we are going to get much higher-quality and much more reliable information. We are moving up the food chain.” Plus loin dans cet article, on voit Dan Clancy, ingénieur en chef responsable de la numérisation de livres⁴⁷, tenir à l’auteur de l’article le propos suivant : “I hope this world evolves so that there exists a time where somebody sitting at a terminal can access *all the*

world's information.” Si tous ne le disent pas toujours, apparemment tout le monde y pense : l'idée originaire était de réunir toute l'information du monde.

Cela noté, on peut se demander pourquoi BC fait référence à la formule “*all the information in the world*” comme si elle était authentique. Cette expression, parfaitement correcte au demeurant, semble à première vue employée dans certains commentaires qui donnent assez souvent l'impression de marquer leur distance vis-à-vis de Google en ne reprenant pas la formule canonique “(*all*) *the world's information*”. C'est ainsi que, dans l'article de l'encyclopédie Wikipédia francophone consacré aux moteurs de recherche, on lisait ceci début septembre 2007⁴⁸ : « Google : Moteur généraliste avec services associés (slogan : “Our mission is to organize all the information in the world” et “Don't be evil” ... ayez confiance...). » Présentée à Google le 26 août 2007, la requête “**all the information in the world**” fait d'abord apparaître un article canular publié par le magazine humoristique *The Onion* le 31 août 2005, intitulé “*Google announces plan to destroy all information it can't index*”⁴⁹. Le deuxième résultat affiché renvoie à un article publié en novembre 2005 par *Forbes.com*, article que son auteur, Quentin Hardy, a intitulé “*Google thinks small*”. On y lit ceci⁵⁰ : “[Google] has huge sales, prodigious profits and a bid to log *all the information in the world* – one tiny project at a time”. Le troisième résultat est un bref compte rendu du livre de BC sur le site *Trackbusters*, site dont l'objet est de proposer « un haut niveau de positionnement web pour les résultats organiques (résultats naturels) des moteurs de recherche tels que Google, Yahoo!, MSN Live Search... » : la présence de la formule recherchée (“*all the information in the world*”) est due simplement au fait que ce compte rendu emploie les formules mêmes usitées par BC. Le quatrième document est un article du numéro spécial de juin 2007 de la revue *Library Philosophy and Practice* sur le thème “*Libraries and Google*”. Sous le titre *Unclear on the Context: Refocusing on Information Literacy's Evaluative Component in the Age of Google*, l'auteure, Genevieve Willams, y formule la conclusion suivante.

Visions of what the library in general and information literacy in particular should be ought to be informed by a greater understanding, not of what we think our students need, but by what they do need: the tools to empower themselves in an age when access to *all the information in the world* is through a narrow rectangular box on Google's spare white page.

Le cinquième résultat renvoie à un bref article qui mérite d'être reproduit intégralement : si la formule “*all the information in the world*” y apparaît bien, en effet, on y verra que le recueil « complet » de cette information serait différé *sine die*.

We'll index the world by 2310, says Google

By Lucy Sherriff

Published Monday 10th October 2005 15:42 GMT

Good news for people who like making lists: Google's Eric Schmidt says that it will take another three hundred years before *all the information in the world* is neatly indexed and searchable.

Schmidt was speaking at the US 'Association of National Advertisers' annual conference in Phoenix when a member of the audience asked how long it is likely to take before Google fulfilled its mission (<http://www.google.com/corporate/>).

"We did a math exercise and the answer was 300 years," CNet quotes Schmidt as saying. "The answer is it's going to be a very long time."

He estimated that out of the five million terabytes of information (no, we don't know how he counted it all up, either) in the world, a mere 170 terabytes have been indexed so far.

Schmidt did not elaborate on what kind of maths had been deployed to answer this riddle, so we can't say for sure whether the calculation only accounts for information that exists right now, or whether it includes new information that will be generated over the next three hundred years.

Presumably, however, by the time these questions are closer to being resolved, we puny humans will have ascended to become beings of pure information, all happily indexing and searching ourselves into digital oblivion.

Le sixième résultat renvoie à un texte intitulé "*The Internet is changing Higher Education*" qui comporte le passage suivant : "*All the information in the world* will be of no avail, unless we can use it intelligently and wisely." Le sous-titre précise qu'il s'agit de "*Remarks by Harvard University President Neil Rudenstine at the Harvard Conference on the Internet and Society, May 29, 1996*" : ces remarques datent donc d'un temps où la « mission » de Google... n'avait pas encore été formulée.

Pour approfondir cette enquête, on peut réduire la requête à un domaine particulier, en l'espèce à *Google.com*, ainsi que le permet le dispositif dit de « recherche avancée » du moteur de recherche. La recherche ne ramène alors que 5 résultats (voir les deux copies d'écran ci-après). Ces résultats contiendraient-ils des déclarations googléennes officielles utilisant l'expression "*all the information in the world*" ? Les deux résultats « chauds » à cet égard sont situés en deuxième et troisième positions, comme on le verra. Le troisième résultat renvoie à une lettre extraite du rapport annuel 2005 de la firme Google (Page & Brin, 2005), où on lit notamment ceci ⁵¹.

I mentioned earlier how we are striving to make Google really understand your query and *all the information in the world*. To do that, we will have to make Google smart, and that requires artificial intelligence. We are particular believers in large-scale AI that involves both a lot of computation and a lot of data. We're looking to build the best center for this kind of work in the world.

[The Google Podium: transcript](#) - [[Traduire cette page](#)]
 When the value was there, suddenly **all the information in the world** was in HTML. When we all make this vision real for health care, suddenly everyone will ...
www.google.com/press/podium/markle.html - 17k - [En cache](#) - [Pages similaires](#)

[Google Press Center: Press Release](#) - [[Traduire cette page](#)]
 "A perfect search engine will process and understand **all the information in the world**," said Sergey Brin, Google president and co-founder of Google. ...
www.google.com/press/pressrel/pressrelease1.html - 11k - [En cache](#) - [Pages similaires](#)

[Google Investor Relations](#) - [[Traduire cette page](#)]
 I mentioned earlier how we are striving to make Google really understand your query and **all the information in the world**. To do that, we will have to make ...
investor.google.com/2005_founders_letter.html - 39k - [En cache](#) - [Pages similaires](#)

[Google Answers: Law enforcement](#) - [[Traduire cette page](#)]
 GA does not archive **all the information in the world** and have it at our disposal to simply look up and present to customers. Researchers here merely "find" ...
answers.google.com/answers/threadview?id=610732 - 22k - [En cache](#) - [Pages similaires](#)

[Google Labs - Google Suggest | Google Groups](#) - [[Traduire cette page](#)]
 life is to make **all the information in the world** (universe?) accessible, and have it served on a cracker. This could be that ...
groups.google.com/group/Google-Labs-Google-Suggest/tree/browse_frm/month/2005-04?_done=%2Fgroup%2FGoogle-... - 114k - [En cache](#) - [Pages similaires](#)

On aura noté que, ici, il s'agit moins d'*organiser* toute l'information du monde que de la *comprendre vraiment* ("*really understand*"). On notera aussi un trait constant de la problématique de Brin et Page, celle du "*large scale*", du plus grand et du meilleur, que le

deuxième résultat va nous faire à nouveau rencontrer. Le lien qu'il propose renvoie en effet à une page contenant un texte daté du 7 juin 1999, où l'on trouve ceci ⁵².

Google employs several key technologies to generate search results of unprecedented accuracy and quality. These technologies extend Stanford University research into large-scale data mining of the Web. "A perfect search engine will process and understand *all the information in the world*," said Sergey Brin, Google president and co-founder of Google. "That is where Google is headed."

Ici, l'emploi de l'expression est attribué à Brin, mais toujours à propos du fait de *comprendre* : "A perfect search engine will process and *understand all the information in the world*." Quelle est donc la nature du document où apparaît cette citation ? Il s'agit apparemment d'un communiqué figurant sur le *Google Press Center*, c'est-à-dire en principe d'un document *quasi* officiel, validé par les dirigeants de Google. Pour confirmer cette interprétation, on peut demander à Google de rechercher les occurrences de "**A perfect search engine will process and understand all the information in the world**". L'examen des premiers résultats ⁵³ conduit alors à retrouver la citation en question à la page 68 d'un livre que BC mentionne quasiment à l'égal de celui de John Battelle ⁵⁴, en le présentant comme « très informé mais plus apologétique » : l'ouvrage de David Vise et Mark Malseed, *The Google Story* (Vise & Malseed, 2006). On y lit ce qui suit ⁵⁵.

On June 7, 1999, less than one year after they took leave from Stanford, Brin and Page issued a press release announcing that Kleiner Perkins and Sequoia Capital had agreed to invest \$25 million in Google Inc. On Stanford's campus and around Palo Alto, jaws dropped. It was a huge sum. These two venture capital firms didn't coinvest, yet Doerr and Moritz were both joining Google's board of directors. And somehow, the two guys, who had always seemed a bit too confident to some of their classmates, had managed to extract an enormous amount of money seemingly without giving up anything in return. All indications were that the Google Guys had a dream deal: the money they needed to build their beloved search engine, and the control and power they needed to call the shots. "We are delighted to have venture capitalists of this caliber help us build the company," Page said in a formal statement. "We plan to aggressively grow the company and the technology so we can continue to provide the best search experience on the Web."

Brin added an appropriately grandiose comment, his confidence higher than ever. "A perfect search engine will process and understand all the information in the world," he said. "That is where Google is headed."

Le document trouvé est donc tout simplement un “*press release*”, un communiqué de presse émanant de Brin et Page et annonçant la bonne nouvelle du financement à hauteur de vingt-cinq millions de dollars de la firme Google. On peut en conclure que les fondateurs de Google en ont approuvé les termes, même si l’on ignore où et quand, par exemple, Brin aurait réellement déclaré : “A perfect search engine will process and understand *all the information in the world*.” On a reproduit ci-dessous les premiers paragraphes de ce communiqué de presse.

Google Receives \$25 Million in Equity Funding

Sequoia Capital and Kleiner Perkins Lead Investment; General Partners Michael Moritz and John Doerr Join Board

Palo Alto, Calif. -- June 7, 1999 -- Google, a start-up dedicated to providing the best search experience on the web, today announced it has completed a \$25 million round of equity funding led by Sequoia Capital and Kleiner Perkins Caufield & Byers.

Google also announced that Michael Moritz, general partner of Sequoia Capital, and John Doerr, general partner of Kleiner Perkins Caufield & Buyers, have joined its board of directors. Michael Moritz is currently a director of numerous companies, including Yahoo, eToys, Quote.com, eGroups, PlanetRx, Flextronics, and WebVan. John Doerr was a co-founder of @Home and is a director of several high growth internet companies, including Amazon.com, Drugstore.com, Handspring, Healtheon/WebMD, Homeshop.com, Intuit, and Sun Microsystems.

"We are delighted to have venture capitalists of this caliber help us build the company," said Larry Page, CEO and co-founder of Google. "We plan to aggressively grow the company and the technology so we can continue to provide the best search experience on the web."

Google employs several key technologies to generate search results of unprecedented accuracy and quality. These technologies extend Stanford University research into large-scale data mining of the Web. "A perfect search engine will process and understand all the information in the world," said Sergey Brin, Google president and co-founder of Google. "That is where Google is headed."

On peut alors imaginer que BC ait emprunté au livre de Vise et Malseed la formulation “*all the information in the world*”⁵⁶. Qu’elle l’ait substituée à la formule “*the world’s information*” dans le *mission statement* de Google est en revanche beaucoup plus étonnant de la part d’une philologue qu’on pourrait supposer, par vocation, attentive au choix des mots et à leur organisation.

CHAPITRE 4

Le cas de PageRank

L'enquête que BC conduit, non sur la devise (*motto*), mais sur le *mission statement* de Google suppose, comme toute enquête, de ne pas tenir pour indépassables des savoirs dont les conditions de production sont si fortement éloignées de celles sous lesquelles on est amené à travailler. La référence à Klemperer à propos de *to organize* est peut être efficace au plan polémique ; elle ne nous semble guère l'être, après enquête, au plan épistémologique de la recherche de la simple vérité. Au reste, on aimerait en savoir plus, et plus nettement, sur ce que l'auteure évoque et invoque à ce propos. La mauvaise part qu'elle croit trouver en l'idée d'organisation est-elle là dès l'origine dans les langues européennes, ou du moins dans certaines d'entre elles ? Comment se fait-il que les usages de *organisieren* que Klemperer stigmatise – non d'ailleurs sans autodérision⁵⁷ – se retrouve apparemment en anglais ? Les deux langues communiqueraient-elles souterrainement ? Laquelle des deux, éventuellement, serait-elle l'inspiratrice de l'autre ? Autant de questions que la courte étude menée précédemment conduirait à soulever, et qui exigeraient un travail *philologique* adéquat⁵⁸.

Dans le chapitre III du livre de BC consacré au *mission statement* de Google, l'auteure met en relation le souci googléen d'organiser l'information avec l'algorithme PageRank : « À l'organisation, écrit-elle (p. 81), correspond l'algorithme PageRank qui permet de classer les réponses pertinentes. » En fait les références à PageRank ne manquent pas depuis les premières pages du livre de BC. Au passage, BC rectifie elle-même l'affirmation précédente, légèrement simplificatrice⁵⁹ : dans une présentation profuse de l'anatomie du « gros engin » (ainsi nomme-t-elle le moteur de recherche Google), elle résume dans les termes suivants le principe du classement par Google des réponses apportées à une requête déterminée⁶⁰.

Tout moteur de recherche classe, plus ou moins bien, en fonction de la coïncidence avec les mots de la requête, en pesant la proximité de tous les mots, puis de quelques mots, dans la réponse à une requête multiple. Google utilise, outre un « *IR score* » de ce type, un algorithme de classement de la pertinence en fonction des liens, à savoir PageRank, et pondère l'un par l'autre pour déterminer le classement. C'est de là que vient la qualité de ses réponses.

Reprenant alors un exemple donné par John Battelle dans son livre à propos de la comparaison entre Google et Yahoo! ⁶¹, elle conclut (p. 63) : « Bref, PageRank, que nous analyserons en détail, est l'une des clefs du secret Google. »

Google : une image ambivalente

Pourquoi BC s'intéresse-t-elle à l'algorithme de PageRank au point de vouloir l'analyser « en détail » ? Le rôle assigné à PageRank par BC dans sa vision personnelle de Google paraît en vérité central : il s'agit d'un objet ambivalent qui, d'un côté, ferait l'originalité de Google, sa force d'intelligence pure, et, de l'autre, aurait partie liée avec son caractère diabolique, que BC ne cesse de dénoncer. Google, ce serait en quelque sort le dieu Pan, d'abord protecteur attentif des bergers du Net et de leurs troupeaux de pages Web, ensuite dieu mauvais de la foule hystérique des « informations en panique ».

Que peut-on cependant regarder comme prétendument bon chez Google/PageRank ? Par rapport aux autres moteurs de recherche existant vers 1999, raconte BC (p. 69), « Google fait la différence grâce à PageRank ». Comment cela ? Parce que, répond BC, il « ne mélange pas les résultats de la recherche (*organic results*) avec la pub et produit du coup de meilleures réponses ». Mais en quoi PageRank est-il lié à la séparation des résultats « *organiques* » (ou « *naturels* ») et de la publicité qui s'affiche à l'écran ? Vraisemblablement, l'auteure veut rappeler ici que les résultats « organiques » ne sont pas influencés dans leur classement par le fait qu'un organisme se serait payé de la publicité chez Google (en lui achetant des mots clés). Et s'ils ne sont pas influencés, semble-t-elle penser, c'est bien parce que le mode de classement des résultats est défini par un algorithme qui ne prend pas en compte les éventuelles annonces publicitaires payées à la firme Google. Elle écrit d'ailleurs (p. 69) : « Avec Google, la morale rapporte, et c'est en somme l'organisation même du moteur qui est morale. » Elle écrira de même plus loin (p. 140) : « ... PageRank intervient seul pour classer les réponses pertinentes. Personne ne peut acheter son classement. Le *motto* moral est, au choix, une condition ou un sous-produit de PageRank. » À l'appui de cette dernière affirmation, BC reproduit alors le passage suivant, que l'on trouve sur le site en français de Google à propos de l'« intégrité » de ce moteur de recherche ⁶².

Comme nous l'indiquons clairement dans nos listes de résultat, certains sites peuvent être associés à une publicité "Sponsored Link". Toutefois, Google ne pratique pas la vente des positions dans ces résultats ; autrement dit, il n'est pas possible d'acheter une valeur PageRank supérieure à la réalité

du Web. Avec la recherche Google, vous disposez d'une solution simple, rapide, honnête et objective pour trouver des sites Web de la plus haute qualité et dont les informations répondent parfaitement à vos besoins.

Pourrait-on se contenter, en ce point, de ce niveau d'analyse ? Ce que l'enquête conduite par BC semble livrer pour le moment est ceci : les résultats affichés par le moteur de recherche Google à une requête donnée ne dépendent que d'un certain algorithme qui tient compte des pages Web existantes et de la requête de l'utilisateur, toute tentative pour corrompre ce mécanisme étant sévèrement réprimée par la firme Google⁶³. En savoir plus sur l'algorithme de classement ne peut servir *a priori* qu'à *vérifier* que l'indépendance proclamée existe bien. Pour cela sans doute, l'enquête sur cet algorithme mérite de se poursuivre.

Poursuivons donc avec BC. Ainsi qu'on l'a vu, elle signale que le PageRank n'est pas le seul critère de classement. « Il y a deux composantes de classement dans la page », note-t-elle, avant d'ajouter (p. 105) : « la première, PageRank, correspond donc à ce que Google appelle la “réalité du Web”, c'est la composante objective. La seconde, c'est “vous”, la composante subjective ». Les qualificatifs « objectif » et « subjectif » peuvent sans doute être contestés ; mais, comme nous le verrons, il y a bien dans l'algorithme de Google une part qui peut être dite *query-independent*, en ce qu'elle résulte de la structure « objective » du Web et est donc indépendante des demandes des usagers (mais non bien sûr de ceux qui « font » le Web en mettant des contenus en ligne) ; et il y a une part qu'on dit en anglais *query-dependent*, qui tient compte de la requête formulée par l'utilisateur⁶⁴, en la rapportant (par le truchement du *IR score*) aux informations détenues par le moteur de recherche sur les pages visitées, analysées et référencées dans ses index. À cette deuxième composante, BC fera une place un peu plus loin. Mais elle va consacrer beaucoup plus d'attention au fonctionnement de la partie *query-independent*, dont l'outil historiquement essentiel est l'algorithme PageRank.

PageRank, mathématiques et bibliométrie

Nous examinerons d'abord les résultats de l'investigation que BC conduit à propos de PageRank, ainsi que les conclusions qu'elle en tire. L'auteure consacre à cette question un développement de quelque huit pages⁶⁵, la référence à PageRank étant par ailleurs fréquente dans l'ensemble de son ouvrage⁶⁶. C'est ainsi que, dès la page 14, elle précise ceci.

Ce moteur de recherche [Google] est basé sur un algorithme, appelé PageRank (parce qu'il donne un rang aux pages Web – mais peut-être aussi parce que c'est Page qui en fut le principal artisan et que

l'humour fait partie de la culture de firme). Cet algorithme, on ne le sait pas assez, est la propriété de l'Université de Stanford conféré sous licence exclusive à Google, mais simplement jusqu'en 2011.

L'auteur précisera un peu plus loin (p. 67) : « Le 4 septembre 2001, Google obtient enfin la validation de son brevet concernant PageRank. C'est un moment clef. » Quelques pages avant, elle avait été amenée à préciser un fait crucial quant à la structure de PageRank, fait qu'elle présente en l'espèce de manière fort concise ⁶⁷.

La grande originalité de Google consiste à indexer non seulement les liens qui partent d'une page, mais les liens qui y mènent. L'indexation génère ainsi une base de données de liens qui sont des paires de documents identifiés, avec les « textes d'ancrage » d'où partent ces liens. C'est à partir de cette base de données, pondérée, qu'est calculée PageRank.

Quel intérêt peut avoir la précision apportée ici ? Ce que BC affiche comme le point de départ de son intérêt pour PageRank est au fond très simple : écoutons-la un peu plus longuement.

On sait que 80 % des recherches s'arrêtent à la 1^{re} page et qu'on les poursuit très rarement après la 3^e page. « Demain, ce qui ne sera pas disponible en ligne risque de devenir invisible à l'échelle du monde », dit Jacques Chirac [à l'occasion des « Vœux aux forces vives », le 5 janvier 2006] ; mais, aujourd'hui, ce qui n'apparaît pas dans les premiers résultats de Google a très peu d'existence. PageRank est l'opérateur de classement dont, il vaut la peine de le répéter, la licence exclusive n'est accordée à Google par l'Université de Stanford que jusqu'en 2011. Formule secrète, aussi secrète paraît-il que celle du Coca-Cola, qui comporterait 500 millions de variables et plus de 2 milliards de termes [C'est du moins l'information que donne Google sur son propre site (« Informations sur la société », au chapitre « Technologie »)]. (p. 95)


On aura observé l'insistance de BC sur la date d'expiration de la licence accordée à Google par l'université de Stanford : il semble que l'on doive interpréter cette répétition comme un indice de l'importance qu'elle attribue à PageRank dans le système Google, donnant à entendre que, sans PageRank, le moteur de recherche Google serait promis à la ruine. Notons ensuite cette autre apparente énigme : alors que, à propos de PageRank, elle parle de « formule secrète » à l'instar de celle du *Coca Cola*, BC reproduit dans la suite de son propos une formule *mathématique* qu'elle présente comme celle-là même de l'algorithme en question ! La formule proposée est manifestement empruntée à une présentation due à Brin et Page ⁶⁸ dont nous reproduisons ci-dessous la diapositive appropriée.

Actual PageRank Model

$$W_j = (1-d) + d \sum_{i=1, i \neq j}^N l_{i,j} \frac{W_i}{n_i}$$

- Calculated using iterative method
- Makes graph fully connected
- d parameter approximates “random surfer” path lengths
- Personalization and Spam

page@stanford.edu
<http://www.stanford.edu/>
14



Slide 14 of 23

La petite énigme ne sera pas davantage élucidée : manifestement, c’est *chacun* des deux termes de l’énigme – PageRank secret, PageRank mathématiquement réglé – qui intéresse BC, plutôt que la résolution de la contradiction qu’il pourrait y avoir entre eux : il n’y aura pas d’*Aufhebung* !

BC effectue cependant une plongée dans la partie technique de PageRank. Nous n’en retiendrons d’abord, ici, que les éléments qui paraissent essentiels pour notre propos, et tout d’abord l’idée même à l’origine de l’algorithme PageRank, que BC présente dans les termes suivants.

L’idée de base de Bryn (*sic*) et Page, que Vise qualifie de « *one-shot idea* [Vise, p. 38 (...)] », pour classer la pertinence des réponses et les hiérarchiser, c’est de prendre en compte non seulement les mots-clés qui déterminent un site, mais les liens qui pointent vers lui. L’anatomie du gros engin s’adosse en effet, on l’a vu, à l’idée de *backrub*, de *backlinks*, c’est-à-dire de liens qui mènent au site et non qui partent de lui... (pp. 97-98)

Notons que le mot *BackRub* qui apparaît dans ce passage n’est nullement synonyme de *backlink* : il s’agit simplement du nom donné par Page à une version primitive du moteur de recherche Google. Le site de la firme Google précise ainsi ⁶⁹ que, “by January of 1996, Larry and Sergey had begun collaboration on a search engine called BackRub, named for its unique ability to analyze the ‘back links’ pointing to a given website”. De même Vise et Malseed écrivent-ils de leur côté (2006, p. 38) : “By early 1997, Page had developed a primitive search engine that he named ‘BackRub’ because it dealt with the incoming – or ‘back’ – links to Web pages. Ever thrifty, Page put his left hand on a scanner, converted the image to black and



white, and the new BackRub site had its logo.” L’image en question ⁷⁰ (ci-contre) permet de comprendre ainsi le choix du verbe *to rub* (qui signifie « frictionner », « frotter », « bouchonner ») : la paume de la main repère les points d’impact des liens vers la page examinée (*inlinks*) ; un mouvement inverse de la main (vers l’arrière, *back*) permet de remonter jusqu’aux pages où figurent ces liens.

Lorsqu’un lien pointe vers une page Web, ce lien est regardé métaphoriquement, par Page et Brin, comme une *citation* de la page Web « cible ». Le texte d’ancrage (*anchor text*), qui sert de support au lien (dans la page où se trouve le lien-citation) est regardé comme une *annotation*. L’idée de base est donc celle-ci : plus une page Web est citée, c’est-à-dire plus elle reçoit de liens (que Page et Brin nomment des “*backlinks*” mais qu’on peut aussi nommer des “*inlinks*”), et plus cette page recevra un « rang » important. *Mais toutes les citations ne se valent pas* : être cité par une page qui cite beaucoup a moins de valeur que le fait d’être cité par une page qui cite peu. Ces principes, intégrés mathématiquement dans l’algorithme PageRank ⁷¹, témoignent du fait que Google participe, à l’origine, du monde académique d’où viennent Brin et Page, qui sont encore doctorants à Stanford lorsqu’ils écrivent leur article sur l’anatomie du « gros engin » ⁷². BC rassemble toutes ces données dans le développement que voici.

Dans l’« Anatomie », les deux doctorants donnent une excellente description des concepts « académiques » (au sens anglo-saxon de « scolaire » et « universitaire ») à la base de PageRank : « Les moteurs de recherche ont migré du domaine académique au domaine commercial. [...] Avec Google, nous avons l’objectif fort d’impulser plus de développement et de compréhension dans le royaume académique. » Page part de l’idée que le Web tout entier a comme prémisses la citation : c’est le lien ; et l’annotation : c’est la description du lien. Le rapport entre un moteur de recherche et le monde de la recherche est trouvé, en même temps qu’une définition ô combien « académique » du monde de la recherche : citation, annotation, autorité, évaluation par les pairs, classement. Avec même l’idée de ce qu’est un mauvais devoir, comme effet pervers du système éducatif : un mauvais devoir est truffé de citations, qui ont d’autant moins de chances d’être de bonnes indications qu’elles sont plus nombreuses : plus un site fait de liens, moins un lien émanant de ce site sera donc coté. (pp. 98-99)

On aura remarqué l’usage que fait ici l’auteure d’un savoir des plus indigènes, à propos du « mauvais devoir » : mauvais parce qu’il serait truffé de citations ! Ce rapprochement, qui l’a

peut-être aidé à comprendre (comme elle espère à son tour que cela aidera son lecteur), ne paraît pourtant pas entièrement justifié ⁷³. Peut-être a-t-il pour effet principal de suggérer que l'idée de base de Brin et Page ne serait pas véritablement une “*one shot idea*”, une idée décisive et qu'on n'a qu'une fois, mais bien une resucée d'un critère fort commun dans le monde scolaire et universitaire. BC le dit d'ailleurs explicitement dans ce commentaire (p. 99) : « Ce n'est pas vraiment nouveau, mais ce qui est nouveau est de prendre ce taureau-là par les cornes. » Faisant référence un peu plus loin à des pratiques classiques en bibliométrie, elle conclut (p. 100) : « Avec PageRank, Google ne fait là encore que montrer ce qui nous arrive à tous et partout. La hiérarchie dépend de la manifestation objective de la considération. » À nouveau, elle recourt en ce point à un savoir indigène pour tout chercheur, en rapprochant les principes « bibliométriques » dont s'inspire Google, d'une part du classement de Shanghai des universités, d'autre part du système qu'elle désigne erronément ⁷⁴ par l'expression de “*Quotation index*”, en protestant au motif que, en cette façon, « la recherche devient ce que les revues disent que la recherche est ».

PageRank contre la démocratie et la culture ?

Sous le titre « La qualité comme propriété émergente de la quantité », BC ouvre alors une nouvelle section de son chapitre en rappelant le principe de PageRank tel que Google l'expose officiellement ⁷⁵.

« PageRank est un champion de la démocratie : il profite des innombrables liens du Web pour évaluer le contenu des pages Web – et leur pertinence vis-à-vis des requêtes exprimées. Le principe de PageRank est simple : tout lien pointant de la page A à la page B est considéré comme un vote de la page A en faveur de la page B. Toutefois, Google ne limite pas son évaluation au nombre de “votes” (liens) reçus par la page ; il procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par Google ont plus de “poids” et contribuent ainsi à “élire” d'autres pages. »

À l'image de la “*citation*” se substitue ici la métaphore du *vote* et de la *démocratie*. C'est l'occasion pour BC d'exprimer son hostilité à ce qu'elle regarde, en fin de compte, comme le triomphe de la *doxa*, de l'opinion.

L'originalité, l'atypie, le génie, le caractère singulier et intempestif de la vérité n'entrent pas dans le système tant qu'ils ne sont pas banalisés : il n'y a pas d'autre de la *doxa*. C'est l'« opinion » qui sert

de point de départ et de point d'arrivée, d'unité de mesure et de critère. Elle définit le statut ontologique des objets qui sont sur la toile et du classement qu'en fait Google. (p. 104)

Contre ce modèle « populiste »⁷⁶, BC évoque d'autres manières de hiérarchiser les œuvres de la culture, sans proposer encore de projet clairement dessiné. Ainsi en va-t-il dans les lignes suivantes⁷⁷.

La hiérarchie ne provient pas du dehors, telle une hiérarchie platonicienne par devoir-être, avec un philosophe-roi pour l'imposer à la masse, ce n'est pas non plus une hiérarchie démocratique par *agôn*, discussion ouverte et dissensus-consensus. Elle est immanente, parce que, personne d'autre que « nous » ne la fait, tout en demeurant opaque parce que mécanique et robotique, mathématique et systémique.

BC retrouve en cela l'opposition de l'*organisation* et du *système*. « L'organisation », écrivait-elle quelques pages plus haut (p. 93), « c'est le technique qui se fait passer pour du naturel, la naturalisation de la technique. » Et elle concluait déjà : « Telle est l'exacte définition de PageRank qui dit la "réalité du réseau" : le secret de l'organique, dans l'organisation Google, c'est le secret du technique... » Ce « naturel organisé » qui prétend imiter la démocratie n'en est qu'une contrefaçon. Plus haut encore dans son ouvrage (p. 84), BC avait persiflé, en passant, l'algorithme PageRank avec cette formule assassine⁷⁸ : « tous les liens sont des votes, mais certains sont plus égaux que d'autres. »

Ainsi qu'on l'aura peut-être remarqué, le positionnement anti-Google de BC ne vient pas comme *conclusion* d'une enquête dont le lecteur serait fait le témoin haletant : il s'exprime en un leitmotiv coextensif à son ouvrage. C'est ainsi, que dès la page 29, on lit ceci.

... à quelque item que l'on ouvre [Wikipedia], c'est utile quand on ne connaît rien et souvent à pleurer dès qu'on connaît un peu (demandez « Platon »). On touche à l'un des problèmes-clefs, que Google gère de front avec l'algorithme PageRank : la quantité produit-elle la qualité ? Le « chacun », le tous un par un, est-ce une garantie d'universalité, est-ce une garantie de démocratie ?

Dans ce registre, la répétition semble la règle. Ainsi retrouve-t-on plus loin dans l'ouvrage (p. 102) l'analyse dépréciative déjà rencontrée (p. 84).

« Page Rank est un champion de la démocratie », la procédure est assimilée au suffrage universel, mais pondéré par une aristocratie de l'importance au moyen d'une analyse de contenu, selon un modèle mixte de démocratie éclairée ou bien tempérée...

Au nom de quoi BC réproouve-t-elle ainsi le principe même des moteurs de recherche et plus particulièrement celui de Google ? La réponse nous est déjà en partie connue et peut encore s'énoncer ainsi (pp. 171-172) : « l'information, travail de journaliste, n'est pas simplement un amas de nouvelles organisées par un algorithme en fonction de ce qui est disponible sur le Web selon le principe de PageRank... » Vers la fin de son livre, elle explicitera son point de vue de manière toujours polémique mais dans une visée sensiblement plus positive.

Pour passer d'un entrepôt numérique à une sélection utilisable, on peut évidemment envisager plusieurs médiations différentes. PageRank est une organisation in(com)parable dont on a compris les limites – doxico-américaines. Ces limites constituent certainement des biais redoutables pour les œuvres et pour la « culture » comme telles, et Jean-Noël Jeanneney a raison d'attirer l'attention sur la « tête de gondole » googléenne, si « réalité du Web » soit-elle, qui ne correspondra jamais à la ou aux structururations qu'une bibliothèque, une université, un musée, un ministère de l'éducation nationale ou de la culture d'un État-nation, ou le commissariat d'une union européenne, seraient en droit de vouloir non seulement pour les siens, mais pour tous. (p. 212)

Évoquant plus loin une « structuration » *qui ne serait pas l'« organisation »* googléenne, elle énonce cette ambition (p. 214) : « La qualité, avec d'autres algorithmes que PageRank, ne sera plus propriété émergente de la quantité, mais affaire d'histoire, d'expertise, de diversité, de cultures et de culture ». On a là une opposition déjà notée, mais apparemment indurée, entre le vrac apparent dont PageRank prétendrait exprimer l'intelligibilité constitutive (grâce à la structure des liens entre pages Web), et la structure construite et imposée par des « experts » que l'on invoque plutôt qu'on ne les définit. Cette opposition n'était pas moins marquée chez Jean-Noël Jeanneney (dont nous avons, plus haut dans ce chapitre, interrogé l'ouvrage à propos du concept d'organisation) : son choix, consonant à celui de BC, est, certes, des plus fermes ; mais cela ne l'empêche nullement d'avouer que, à l'instar sans doute de beaucoup d'intellectuels, il connaît les plaisirs furtifs que le vrac procure parfois, comme il le suggère dans ce passage, où tentation et rémission se côtoient un court instant (Jeanneney, 2006, p. 121).

Certes, nous savons les plaisirs que peuvent procurer les mystérieuses rencontres qu'un livre feuilleté, une émission trouvée en « zappant » et désormais une errance sur la Toile offrent à l'improvisiste. On peut s'y enrichir au petit bonheur la chance. Mais ces menus bonheurs-là n'ont leur prix que s'ils tombent sur un terreau fertile, celui d'une pensée construite, que s'ils passent dans l'étamine d'une raison sage, d'une culture bien constituée. Sinon, puisque l'on est impuissant à situer chaque donnée, ce n'est qu'anarchie pour l'analyse et chausse-trapes pour l'action. Je parlais en commençant de l'espoir que la Toile puisse diminuer, à hauteur planétaire, les inégalités devant le savoir. Or voici une évidence : le vrac, s'il domine, les accroîtra au contraire.

Méprises et malentendus

La prise de position de BC, dont nous ne discuterons pas davantage ici, la conduit sans doute à une méprise partielle quant au projet conçu et mis en œuvre par Page et Brin. Ces derniers, en effet, se placent dans le cas du Web *tel qu'il existe* et non dans une uchronie où tout serait toujours déjà structuré en ce que les “Google Boys” appellent des “well controlled collections”. Or, à la page 99 de son livre, BC affirme ce qui suit.

Le problème est au fond le suivant, diagnostiqué dans l'« Anatomie » : comment utiliser le Web, qui est « une vaste collection de documents complètement hétérogènes » comme si c'était une ou plusieurs « collections bien contrôlées » ? C'est un problème de bibliothécaire, de bibliométrie...

En réalité, ce n'est pas cela que disent Brin et Page dans leur article fondateur. Ils affirment en effet que “the web is a vast collection of completely uncontrolled heterogeneous documents”, et qu'une “big difference between the web and traditional well controlled collections is that there is virtually no control over what people can put on the web”. Mais ils ajoutent alors très précisément ceci.

Couple this flexibility to publish anything with the enormous influence of search engines to route traffic and companies which deliberately manipulating (*sic*) search engines for profit become a serious problem. This problem that (*sic*) has not been addressed in traditional closed information retrieval systems.

Le problème que se posent Page et Brin n'est pas de transformer un jour le Web en un système de collections bien contrôlées (ce dont rêvent, à l'évidence, BC et Jean-Noël

Jeanneney) ; mais, prenant le Web tel qu'il est, il est de permettre aux usagers d'accéder à ses ressources d'une façon « idoine », laquelle, bien entendu, se discute, mais doit d'abord avoir le mérite d'exister un tant soit peu, fût-ce à travers PageRank.

La construction du malentendu se nourrit de diverses confusions, dont la plus énigmatique peut-être tient en deux mots, “*click*” et “*clickstream*”, « clic » et « flux de clics ». BC en effet associe l'algorithme PageRank avec ce qui serait la mémorisation par Google des séries de clics d'un usager utilisant son moteur de recherche. Ainsi écrit-elle par exemple ceci.

Google, par robot interposé, voit tout, et vous fait les yeux doux en ne vous proposant que ce qui vous intéresse – c'est sa manière de flirter, avec chacun de vous, en vous donnant l'impression d'être unique. On y reviendra avec l'algorithme PageRank et la « BigBrotherisation ». (pp. 78-79)

À propos de l'un de ses thèmes favoris, elle écrira de même un peu plus loin ceci ⁷⁹.

Cette transformation de la quantité en qualité, qui vaut pour la manière dont marche un « ordinateur », vaut aussi, on l'a vu, pour l'équation constitutive de Google, PageRank. *Le nombre de liens et le nombre de clics*, quantités discrètes, sont au fondement de la qualité ou de la valeur de l'information : ils informent l'information.

L'affirmation paraît opaque : de quels clics s'agit-il, et en quoi de tels clics interviendraient-ils dans le classement des pages de résultats affichées dans Google ?

Avant de répondre, une précision quant au vocabulaire employé s'impose. Le terme « clic », précise le Dictionnaire culturel en langue française (Rey et al., 2005), entre dans cette langue au XV^e siècle comme simple onomatopée. Au XIX^e siècle, il reçoit de l'anglais une valeur nouvelle, spécialisée, celle de « phonème articulé par une double occlusion du passage buccal sans participation de la respiration », comme il en va du « son noté ts-ts, en français » (le hottentot est un exemple de « langue à clics »). Enfin, une troisième valeur, venant de l'anglo-américain, s'ajoute à la fin du XX^e siècle, qui correspond au clic de l'utilisateur d'Internet utilisant la souris de son ordinateur. Notons que, en français, le « clic », c'est, à proprement parler, le bruit entendu lorsque se produit un déclic.

Il semble que, pour BC – c’est là une conjecture que nous faisons à son propos –, les résultats que tel usager (supposé identifié par Google grâce aux témoins de connexion, les *cookies*, installés subrepticement sur son ordinateur) verrait s’afficher sur l’écran de son ordinateur après une requête adressée à Google, tiendraient compte de certains traits de son « comportement digital » tel que Google les aurait espionnés à l’occasion de requêtes précédentes. Cela augmenterait donc les principes déjà mentionnés d’un troisième critère de classement des résultats affichés : à la congruence avec la requête de l’usager des pages indiquées par Google, à la structure des liens du Web tel que la reflète l’algorithme PageRank (ou du moins l’algorithme effectivement utilisé aujourd’hui), s’ajouterait ainsi l’expression de la « personnalité » de l’usager, concrétisée par ses recherches, récentes et anciennes, sur le Web à l’aide de Google.

Une telle hypothèse, notons-le expressément, n’est évoquée par John Battelle dans son livre (sur lequel BC s’appuie fortement) que comme une vision d’un futur encore incertain, ainsi que le peint le passage que voici ⁸⁰.

Clickstreams are the seeds that will grow into our culture’s own memex – a new ecology of potentiel knowledge – and search will be the spade that turns the Internet’s soil. Engines that leverage clickstreams will make link analysis-based search (nearly all of commercial search today) look like something out of the Precambrian era. The first fish with feet are all around us – nearly every search engine now supports search history, and dozens of interesting tools have recently come to market that attempt to make sense of the patterns we searchers are leaving upon the Internet’s corpus. We have yet to aggregate the critical mass of clickstreams upon which a next-generation engine might be built, and it will not necessarily be built with our tacit consent, as I pointed out in Chapter 8. But regardless of our trepidation, we’re already pouring its foundations.

Dans le tableau que peint BC, le présent semble se mêler sans distinction nette avec un futur contingent. Dans le domaine qui nous occupe, bien des commentateurs, il est vrai, s’emploient fréquemment à brouiller cette incertaine frontière. Pourtant, savoir où l’on en est aujourd’hui, ou, plus précisément, connaître la *distribution statistique des usages* (plutôt que des usagers), voilà sans doute un savoir qui pourrait être pertinent pour une utilisation raisonnée des moyens d’accéder aux ressources du Web. De la même façon, savoir que les données transmises par des *cookies* sur les pratiques de tel usager ne sont pas nécessairement utilisées pour connaître les habitudes *personnelles* possibles de cet usager-là mais pour être agrégées à une description statistique *d’ensemble* des usages, cela en effet peut permettre de se former une vision des choses moins nécessairement paranoïaque, même si elle n’est pas

pour autant approbatrice. Bien d'autres aspects élémentaires devraient, dans cette perspective, être clarifiés chez l'utilisateur ou le futur usager. Déjà, dans leur article fondateur (1998), Brin et Page écrivaient par exemple ceci.

Because of the vast number of people coming on line, there are always those who do not know what a crawler is, because this is the first one they have seen. Almost daily, we receive an email something like, "Wow, you looked at a lot of pages from my web site. How did you like it?"

La longue citation que nous ferons encore, ci-après, du livre de John Battelle donne à voir un phénomène semblable, qui a eu un retentissement bien plus important, à propos du service de messagerie électronique *Gmail* lancé par la firme Google le 1^{er} avril 2004 ⁸¹.

... Gmail sparked the company's first full-blown PR crisis. The reason? Privacy. Gmail used Google's AdWords technology to place advertisements alongside users' e-mail messages. Now, the idea of placing ads in e-mail is certainly not new – Yahoo and Microsoft both did it, and Web mail users were accustomed to seeing ads – they were the quid pro quo of having a free service. But somehow Gmail pushed the boundaries – its ads were simply *too relevant*. When Mom sends an e-mail about apple pie to her son, and her son sees ads for apple pie recipes alongside her e-mail – well, for some, that crosses the line into spooky. It's a transgression of the public-private line – it's as if someone at Google was really reading Mom's e-mail, then choosing the ads that should accompany it.

The initial reaction was negative. "Search is one category; your e-mail is quite another. Do you really want Google snooping so close to home?" wrote Charles Cooper, a commentator on CNET.com, an industry news site. "The company says it is not going to read the contents of anyone's in-box. Still, you don't need to be a privacy extremist to realize that this fundamentally remains a bad idea."

Of course, Google's computers were not actually *reading* the e-mail; instead, they were simply parsing text for matches with the AdWords network. And that's the difference between Google's approach and that of Yahoo or Microsoft: Google used e-mail as a distribution system for its massive network of advertisers. Since there were so many possible ads for any given phrase, the chance that one matching an otherwise innocuous line in e-mail ("apple pie," for example,) would come up was quite high, compared with the more primitive approach taken by others e-mail providers.

To most of the world, it appeared Google was indeed reading your e-mail. Now, to be clear, only human beings can actually read, but that distinction was for the most part lost in the ensuing debate.

And there were larger issues at play. Privacy advocates such as Daniel Brandt of Google Watch.org pointed out that now that Google had your e-mail address, it could potentially tie your IP address (a unique number that is used by browsers to identify your computer) to your identity, creating an opening for all sorts of potential privacy abuses. Theoretically anyway, Google could now track your entire Web usage, not just your e-mail. (p. 195)

Un appel de note est appendu à l'affirmation que "only human beings can actually read" ; la note correspondante précise ceci : "This may change if search indeed becomes the place where true artificial intelligence arises, as outlined in Chapter 1." On notera, là encore, le souci de distinguer la réalité présente de celle qui *pourrait advenir* mais qui n'est encore qu'une utopie – et pour certains un mauvais rêve.

CHAPITRE 5

Scilicet ou Qu'est-il permis de savoir ?

Dans son ouvrage, Barbara Cassin mentionne deux fois un même article signé de Jacques Lacan et figurant dans le numéro 4 de la revue *Scilicet*. Dans la présentation de cette revue, parue dans le numéro 1 en 1968, Lacan proposait du mot *scilicet* cette traduction : « tu peux savoir ». Le mot *scilicet* résulte en effet de l'accrochage des mots latins *scire* (savoir) et *licet* (il est permis de) : *il est permis de savoir* ; ou, comme le disent les dictionnaires de l'anglais, langue où ce mot a pris la signification de *à savoir* (et est un équivalent de “namely”, ou de “to wit”) : “one is permitted to know”. Il y a donc ce qu'il est permis de savoir, et notamment aujourd'hui ce qu'il est permis de savoir *grâce aux ressources du Web*.

La question directrice de notre travail est celle-ci : *qu'y aurait-il à savoir et que peut-on en savoir* pour faire un usage raisonné, optimal sous tel ensemble de contraintes, des moyens d'accès aux ressources du Web ? Un tel usage suppose à l'évidence une *technique* et une *technologie*, les mots « technique » et « technologie » étant pris ici au sens que leur donne la théorie anthropologique du didactique. Comme le prévoit la TAD, ces composants praxéologiques sont accompagnés d'éléments d'une *théorie*, dont, comme pour la technique et la technologie, il conviendra constamment de faire l'analyse critique, en la déconstruisant pour tenter de la reconstruire de manière contrôlée. Dans l'exposé de BC, par exemple, affluent fréquemment des éléments théoriques *lourds*, qui pèsent manifestement sur les composants technologiques et sur les techniques utilisées par BC. La théorie, ici, indubitablement, fait sienne une appréciation extrêmement positive de certains objets – ces « bons objets » que sont apparemment, pour BC, la philosophie grecque, celle de Hegel, la philologie de Klemperer, et ainsi de suite – en même temps qu'elle profère une appréciation très négative d'autres objets – ces « mauvais objets » que sont, en dépit de certaines ambivalences plus en moins assumées, Google, les États-Unis, George W. Bush, etc. – contre lesquels, pour paraphraser une formule connue, tout semble pouvoir être retenu.

C'est cette réalité théorique où s'inscrit le propos de BC que l'on rencontrera encore et encore à travers les savoirs et, plus largement, les œuvres de la culture que BC fait défiler dans son ouvrage, parfois en se contentant de les mentionner à la façon dont on prononce une

formule conjuratoire, parfois en essayant de leur donner un rôle opérationnel dans l'offensive qu'elle développe à l'encontre du petit monde googléen en expansion apparemment indéfinie.

Savoirs et folklore de l'Internet

L'un des grands problèmes auxquels se heurte qui entend explorer la cartographie des savoirs éventuellement utiles pour un bon usage du Web – c'est-à-dire *des moyens pour accéder aux ressources du Web* – est précisément celui du *degré de pertinence* de ces savoirs, ou plutôt celui de l'identification des ensembles de conditions et de contraintes assez fréquemment réalisés et sous lesquels une telle pertinence apparaît fondée. Pour commencer ici d'aborder ce problème tout en poursuivant encore un peu avec BC, nous examinerons rapidement les sept encadrés qu'elle place en fin de chacun des chapitres I à V de son livre en les introduisant par ce commentaire⁸² : « J'ai choisi de proposer en encadré à la fin de chaque chapitre les digressions qui m'ont paru personnellement nécessaires pour mieux comprendre ce dont il s'agit avec Google. »

Le chapitre I, intitulé « Google révélateur du Net », comporte deux encadrés, A et B. Le premier des deux constitue une petite synthèse que BC destine sans doute d'abord à elle-même sur des questions que le titre révèle sans façon : « Ce que j'ai toujours voulu savoir sans jamais oser le demander : l'Internet, le Web, un peu d'histoire immédiate. » Comme pour l'ensemble des encadrés, nous aborderons ce texte avec trois questions en tête, auxquelles nous ne répondrons que très partiellement : tout d'abord, quelles informations apporte-t-il et à propos de quels objets ? Ensuite, d'où BC tire-t-elle ces informations ? Enfin, ces informations comportent-elles des éléments erronés ou incertains ?

Le premier paragraphe de l'encadré A fournit un bref historique de l'Internet en précisant certains aspects de sa structure et en évoquant la situation géopolitique que son existence met en jeu ; le voici.

L'Internet, abréviation de « *Interconnected Networks* », est le réseau mondial qui interconnecte tous les réseaux, auquel tous les ordinateurs du monde peuvent se connecter. Son ancêtre est l'Arpanet créé en 1969 à l'initiative du ministère de la Défense américain. Le point de départ est tout pragmatique : comment consulter avec un terminal unique plusieurs centres de calcul ? Mais le résultat est que la communication ne dépend plus d'un centre unique stratégiquement vulnérable : avec le réseau, il y a une multiplicité de couches hétéroclites (chacune a son protocole de communication, son langage et son canal – ligne téléphonique, fibre optique, câble, satellite) et une

multiplicité de centres (50 % du trafic mondial d'Internet passent toujours par l'État de Virginie).
L'Internet permet d'interconnecter le monde entier, et la France y est reliée depuis 1988. (pp. 31-32)

Les informations données dans le passage précédent proviennent apparemment, en grande partie, d'un livre auquel renvoie un appel de note situé en fin de passage ⁸³.

Voir le petit livre déjà ancien de Christian Huitema, *Et Dieu créa l'Internet...*, Eyrolles, 1995, qui raconte l'histoire du Net à travers sa propre expérience au CNET (Centre national d'études sur les télécommunications), à l'INRIA (Institut national de recherche en informatique et automatique) et à la présidence de l'IAB (*Internet Activities Board* et, depuis 1993, *Internet Architecture Board*). (p. 32, note 11)

En vérité, les informations données peuvent être regardées aujourd'hui comme du « folklore », masse mouvante de connaissances prétendues, où tout n'est pas forcément exact, recopiées à satiété par une foule d'historiens ou de commentateurs improvisés pour qui présenter l'objet « Internet » est d'abord un geste rhétorique de courtoisie, sans pour autant que cela soit pour eux une nécessité fonctionnelle de leur travail.

Précisons ici le sens à donner au mot folklore tel que nous l'utilisons dans ce travail. Le mot a été créé en anglais au milieu du XIX^e siècle à partir de folk, le peuple, et de lore, le savoir. Il signifie donc de façon littérale « savoir du peuple ». Il a aujourd'hui en anglais trois orientations sémantiques proches mais distinctes (les citations qui suivent sont toutes empruntées au OneLook Dictionary). Tout d'abord, il désigne les “traditional customs, tales, sayings, dances, or art forms preserved among a people”, les “legends, traditions, customs, beliefs, and the like, of a racial, geographical, or national group, handed down through the generations”, les “traditional beliefs, legends, customs, etc., of a people”, bref, ce qui constitue “the lore of a people”. Ensuite, il étiquette “a branch of knowledge that deals with folklore”, “the study of traditional stories, music, and customs”, “the comparative study of folk knowledge and culture”, qui est “also called folkloristics”. Enfin, le même terme se rapporte à “an often unsupported notion, story, or saying that is widely circulated”, “a body of widely accepted but usually specious notions about a place, a group, or an institution”, “a popular but unfounded belief”, “a body of widely held but false or unsubstantiated beliefs”. C'est dans un sens proche de cette dernière acception que nous utiliserons, en français, le mot de folklore, en subsumant sous ce concept des assertions réputées bien

connues dans un certain milieu humain où elles sont généralement tenues pour vraies sans toutefois avoir été récemment éprouvées, statut propice à la propagation d'erreurs ou d'à-peu-près. (Sur le sens que nous faisons nôtre ici, on pourra se reporter notamment à l'article "Mathematical Folklore" de l'encyclopédie Wikipedia.)

La première information dont la philologue Barbara Cassin se soucie est l'origine du mot « Internet ». L'indication qu'elle donne à cet égard – le mot dériverait de l'expression "interconnected networks" – ne semble pas garantie : plusieurs sources font du mot une simple abréviation de *internetwork*, ce qu'il faut lire *inter-network*, contraction qui aurait été créée par le ministère américain de la Défense vers 1986.

La référence à l'Arpanet appartient elle-même au folklore du sujet. Voici un minuscule florilège constitué à partir de quelques ouvrages, choisis de façon erratique, adressés généralement aux "*dummies*" – aux « nuls », comme le dit la version française d'une célèbre collection d'ouvrages destinés à satisfaire la curiosité du grand public ⁸⁴.

• **Gilles Bauche, *Tout savoir sur Internet*, Arléa, 1996, p. 11**

Arpanet : *Advanced research project administration Network* – ancêtre du réseau Internet développé par le DoD (Department of Defense) américain au début des années 70.

• **Angus J. Kennedy, *Internet et le Web facile*, Éditions Mille et une nuits, 1997, p. 11**

Le réseau Internet fut conçu en 1969 aux États-Unis, pour les besoins du Département américain de la Défense (ARPAnet). Son objectif était de pallier (*sic*) une éventuelle attaque nucléaire. Des scientifiques et des militaires avaient ainsi commencé à échanger des informations. Mais cela est une vieille histoire.

• **John R. Levine, Arnold Reinhold & Margaret Levine Young, *The Internet for Dummies*, Wiley Publishing, Inc., 8^e édition, 2002, p. 183.**

ARPANET: The original ancestor of the Internet, funded by the U.S. Department of Defense.

• **James Benenson & Brigitte Januals, *Guide bilingue de la communication électronique. Electronic Communications Bilingual Guide*, Pocket, 2003, p. 12**

Le premier logiciel de courrier électronique fut conçu en 1972 par l'Américain Ray Tomlinson dans la firme de consultants BBN (Bolt, Beranek & Newman), qui travaillait pour l'ARPA (*Advanced Research Project Agency*). Il était destiné à être utilisé sur le réseau *Arpanet* de l'agence ARPA, créée en 1957 (l'année du lancement du satellite soviétique *Sputnik*) par le Ministère de la Défense américain pour des projets de recherche scientifique avancée pouvant être utilisés à des fins militaires.

On notera quelques imprécisions ou incorrections vénielles : l'ARPA est l'*Advanced Research Projects Agency*, où *Projects* s'écrit avec un *s* marquant le pluriel (contrairement à ce qu'écrivent les auteurs du *Guide bilingue de la communication électronique*, qui en outre orthographient à l'américaine le nom du Spoutnik soviétique), et a le statut d'*agency* et non d'*administration* (contrairement à ce qu'indique Gilles Bauche). Voici maintenant, pour le contraste, un extrait d'un ouvrage avancé, dû à Romualdo Pastor-Satorras et Alessandro Vespignani, intitulé *Internet. Structure et évolution* (2004, p. 20).

En 1966, Lawrence G. Roberts, chercheur au MIT, a commencé pour l'ARPA la conception de l'ARPANET, un réseau conçu initialement pour relier quatre unités centrales majeures situées dans des universités du sud-ouest des Etats-Unis : la *University of California* à Los Angeles (UCLA), la *University of California* à Santa Barbara, le *Stanford Research Institute* (SRI), et la *University of Utah*. Les spécifications de l'ARPANET ont été déposées en 1968, et elles étaient fondées sur la technologie IMP. Le premier message d'hôte-vers-hôte a été envoyé depuis l'UCLA vers le SRI en octobre 1969. Les quatre unités centrales furent reliées à la fin de 1969 et formèrent l'ARPANET, le précurseur de l'Internet actuel.

Dans la période 1970-1971, l'ARPANET a grandi jusqu'à contenir 23 nœuds, tandis que le travail pour concevoir un protocole fonctionnel de communication d'hôte-vers-hôte appelé *Network Control Protocol* (NCP) avançait rapidement. Ce protocole a été totalement implémenté en 1971-1972. Après avoir établi un protocole fiable, les chercheurs ont commencé à développer des applications, et en 1972 Ray Tomlison du BBN a écrit le logiciel de courrier électronique élémentaire, qui est devenu au cours du temps l'application probablement la plus utilisée sur le réseau. Le *Internetworking Working Group* (INWG) est devenu la même année la première entité d'établissement de standards qui gouvernait le réseau croissant. Les premiers nœuds internationaux ont été établis en Angleterre et en Norvège en 1973. En 1977, l'ARPANET contenait 107 nœuds. Sa croissance était lente mais constante, et la communauté scientifique reconnaissait que ce nouveau réseau de communication allait devenir quelque chose de plus grand que tout ce qu'ils avaient imaginé. Cela appelait des plans et des technologies conçus pour un réseau plus grand, et l'Internet a démarré.

On aura observé que le caractère plus savant de ce texte n'empêche pas les bévues : ainsi INWG n'est-il pas l'acronyme de *Internetworking Working Group* mais de *InterNetwork Working Group*. Cela noté, l'indication donnée par BC que, au point de départ de la création de l'Arpanet, se trouve la volonté d'avoir un terminal unique permettant de consulter plusieurs *centres de calcul* est ambiguë et gomme un élément important de l'histoire. Si, en

effet, il s'agissait bien de relier entre eux plusieurs "computer centers", s'il est vrai aussi que ceux-ci étaient regardés à l'origine plutôt comme des "computing centers" (centres de calcul), le mouvement ayant amené la création de l'Arpanet semble avoir consisté à se déprendre d'abord, non sans difficulté, d'une conception de l'ordinateur – le *computer* – comme « machine arithmétique » (*arithmetic engine*), pour y voir de plus en plus un dispositif permettant la *communication*. Dans un texte de l'ARPA, le rédacteur de l'époque souligne en effet ceci ⁸⁵ : "The ARPA theme is that the promise offered by the computer as a *communication medium* between people, dwarfs into relative insignificance the historical beginnings of the computer as an arithmetic engine." Il y a là quasiment l'indice d'une *rupture épistémologique*, dont le même document indique qu'elle s'opère alors non sans rencontrer des résistances, et cela tant dans le monde de l'industrie informatique que dans le monde académique. À propos du premier, le même document indique par exemple ceci.

The computer industry, in the main, still thinks of the computer as an arithmetic engine. Their heritage is reflected even in current designs of their communication systems. They have an economic and psychological commitment to the arithmetic engine model, and it can die only slowly...

À propos du monde universitaire, la transition vers un autre modèle de *computer* se heurte à une semblable inertie, comme le même document le souligne encore : "... furthermore, it is a view that is still reinforced by most of the nation's computer science programs. Even universities, or at least parts of them, are held in the grasp of the arithmetic engine concept..." Bien entendu on ne saurait reprocher à la très courte présentation que s'accorde BC de ne pas mettre en évidence cette rupture essentielle dans la conceptualisation qui conduira à l'Internet. On notera cependant qu'elle donne une visibilité à l'« ancien », à savoir ici la référence au calcul, et gomme le « nouveau », c'est-à-dire l'idée de *communication* entre *computers*. Même dans un compte rendu qui se présente comme factuel, le poids des engagements pour et contre se fait ainsi sentir.

Politique et savoirs de l'Internet

Nous reproduirons encore un paragraphe de l'encadré A qui illustre bien le mélange – au reste parfaitement légitime – d'informations « techniques » et de critiques politiques qui semble consubstantiel à l'écriture de BC.

Son autre est l'intranet (ou plutôt les intranets), réseaux non pas globaux mais à chaque fois limités, *intra*, à une entreprise ou à une université par exemple. Un intranet peut évidemment être relié à l'Internet. On peut aussi dresser des barrières à l'intérieur de l'Internet : la nouvelle muraille de Chine que le gouvernement chinois est en train de construire au sein de l'Internet en Chine, censurant les sites d'opposants ou les informations et les images qui ne lui conviennent pas, réduit l'Internet chinois à un vaste intranet politiquement correct. Google, après Yahoo! vient de se plier à la censure comme condition du marché [12. Voir *Libération* du 15 février 2006 et, sur la relation entre Google et les États, p. 170-178.] Comme toute censure réussie (toujours le vieux principe selon lequel l'art suprême, c'est de cacher l'art), elle efface ses traces pour que l'utilisateur ignore qu'il y a censure : Google avait choisi dès sa première installation en Chine, comme obéissant à l'injonction d'un *Patriot Act* chinois, que les adresses interdites ne génèrent même pas de message d'erreur. (pp. 32-33)

Le propos introduit la notion d'intranet et l'illustre aussitôt par l'épisode – à résonances mondiales – qui a marqué entre 2004 et 2006 les relations de la firme Google avec les autorités chinoises. On notera ici, d'abord, à propos du couple internet-intranet, une différence formelle dans l'usage de ces mots. Alors que « intranet » accepte le pluriel – il est de l'essence de la notion d'intranet de renvoyer à une pluralité de réalisations –, « internet », notamment lorsqu'on l'écrit avec une majuscule à l'initiale (Internet), y est hostile : il n'y aurait qu'un Internet, « le réseau global ». Lorsqu'on saisit "**internets**" dans la boîte de recherche de Google (ou de Yahoo!, ou de MSN) on voit s'afficher comme premier résultat un lien renvoyant à l'article éponyme de l'encyclopédie Wikipedia, où ce pluriel est présenté et commenté comme l'un des plus célèbres *bushisms*. Georges W. Bush avait en effet déclaré ceci, le 17 octobre 2000, lors d'un débat présidentiel avec Al Gore ⁸⁶.

We can have filters on Internets where public money is spent. There ought to be filters in public libraries and filters in public schools so if kids get on the Internet, there is not going to be pornography or violence coming in.

Il réitérera ce lapsus le 8 octobre 2004, lors du second débat présidentiel de la nouvelle campagne, déclarant en réponse à une question sur la possibilité d'une nouvelle conscription (*draft*) : "I hear there's rumors on the, uh, internets [pause] that we are going to have a draft." Selon l'article cité ici, la transcription officielle de ce propos présidentiel fait suivre le mot "internets" de la mention "(sic)", en même temps qu'il gomme le "uh" qui le précède. Cela ne suffira pas à désarmer les médias satiriques, qui s'emparent aussitôt de l'affaire : dès le 9

octobre, une émission de télévision parodiera le propos présidentiel en lui mettant dans la bouche la réponse fictive suivante.

I hear there's rumors on the Internets that we're going to have a draft. I don't know how many of these Internets are carrying these rumors, but they're just wrong. I think the problem here may be more of a question of getting rid of the bad Internets and keeping the good Internets. You know, 'cause I think we can all agree... there're just too many Internets.

Le même article de Wikipedia précise toutefois ce qu'il nomme un "older use of 'internets'" en reconstituant ainsi une évolution à laquelle les mésaventures de Bush semblent alors mettre un point final, du moins pour le grand public américain.

The singular "Internet" (usually capitalized, but see Internet for discussion) results from the connection of many smaller networks. In the early years of internetworking, an "internet" (uncapitalized) represented the connection of any two separate networks, and this word can be properly pluralized. This usage has not been common since the 1980s; today, systems of internetworked computers apart from (or as a subset of) the larger Internet are referred to as intranets. The term "intranet" is also used to refer to collections of Internet-technology applications (i.e. private web pages, portals) that are not visible on the public Internet.

Sans doute les *intranets* que mentionne BC auraient-ils naguère été appelés des *internets*, pluriel plus ou moins tombé en désuétude aujourd'hui. Le pluriel d'« internet » continue cependant d'avoir une vie hors de la scène politique américaine. C'est ainsi que l'auteur d'un ouvrage intitulé *Computer Networks and Internets* (4^e édition, 2004), Douglas E. Comer (qui enseigne – ou enseignait – la *computer science* à Purdue University), le présente comme "an introduction to networking and internetworking", et ajoute : "the text answers the basic question 'how do computer networks and internets operate?' in the broadest sense." Dans la perspective où nous nous inscrivons ici, le type de savoirs évoqué à propos de cet ouvrage fait contraste avec un autre type de savoirs, ceux évoqués par BC à propos de ce que Vise et Malseed nomment, dans leur ouvrage, *The China Syndrome*, le syndrome chinois – tel est le titre du chapitre 25 de leur livre (pp. 269-280) –, soit les tribulations de Google en Chine, qui l'ont finalement amené à accepter une censure qui se traduit, pour certains mots clés désignés par le gouvernement chinois, par le fait que Google.cn affiche en bas de page une indication dont la traduction en français serait ⁸⁷ : « Conformément aux règles, directives et lois locales, une partie du contenu ne peut être montrée. » Devant ce qui apparaît à nombre

de commentateurs comme une violation de la célèbre devise de Google – “don’t be evil” –, la firme de Mountain View a tenté de se défendre par un argument que l’un de ses conseillers, Andrew McLaughlin, aurait énoncé ainsi⁸⁸ : “While removing search results is inconsistent with Google’s mission, providing no information (or a heavily degraded user experience that amounts to no information) is more inconsistent with our mission.”

Ce qui nous importe, ici, tient dans ce fait que l’évocation par BC d’un problème tout à la fois économique, idéologique et politique mobilise des savoirs en même temps qu’il les perturbe. On a vu que BC égalait « l’Internet chinois à un vaste intranet politiquement correct ». Il semble que cette image soit empruntée au compte rendu, paru dans le quotidien *Libération* du mercredi 15 février 2006, d’un entretien entre une journaliste, Frédérique Rousset, et Julien Pain, responsable du bureau « Internet et Libertés » de l’organisation non gouvernementale *Reporters sans frontières*. Le texte publié prête à ce dernier le propos suivant (Pain avec Roussel, 2006) : « Les Chinois sont en train d’inventer un modèle alternatif : des intranets nationaux, et un Internet structuré comme une somme d’intranets. Le modèle est en train de se répandre parmi les dictatures de la planète. » L’emploi d’« intranets » dans ce contexte est d’abord métaphorique et la métaphore est clairement *politique*. Tout se passe, semble-t-on nous dire, comme si l’Internet chinois fonctionnait comme un intranet à l’échelle du pays. L’image est sans doute efficace dans le registre où elle est utilisée. En même temps, elle gomme ce fait de base qu’un intranet relié à l’Internet est une partie de l’Internet et non un réseau isolé ou quasi isolé. Si, dans une famille, les parents installent un filtre sur l’ordinateur familial pour empêcher leurs enfants de visiter tel ou tel site Web, cet ordinateur n’en sera pas moins connecté à l’Internet, et on ne dira évidemment pas qu’il constitue à lui tout seul un intranet ! Si les enfants disposent de plusieurs ordinateurs en réseau, et si un et un seul de ces ordinateurs, muni d’un filtre, permet à chacun des autres d’accéder à Internet, on aura bien là un intranet relié à l’Internet, sans pour autant que ce soit l’interdit pesant sur l’accès à certains sites Web qui définissent l’existence d’un intranet. En d’autres termes, le fait d’avoir des interdits sur l’accès à l’Internet n’est pas ce qui permet de qualifier d’intranet le système formé par les ordinateurs concernés, quand bien même l’interdit existerait à l’échelle de tout un pays.

Ces remarques soulèvent deux questions. Tout d’abord, quand dira-t-on qu’il y a un intranet, y compris pour désigner un internet qui ne serait pas l’Internet, et notamment ce qu’on nomme parfois un *extranet* ? Ensuite, et surtout, en quoi consiste, phénoménologiquement, le fait pour un ordinateur connecté à l’Internet de n’avoir pas un accès libre à tous les sites Web par exemple ? Sans doute y a-t-il, derrière cette dernière question, un ou des savoirs potentiellement utiles à la recherche d’informations sur Internet –

que cela se fasse avec Google ou avec un autre moteur de recherche, et que l'auteur de la requête se trouve en Chine ou ailleurs dans le monde.

À l'automne 2002, précise John Batelle (2006, p. 205), le gouvernement chinois engagea une opération à l'encontre de Google et d'autres moteurs de recherche. Bien que l'on ne sache pas exactement ce que fut alors l'objet des négociations subséquentes entre Google et les autorités chinoises, on peut dire tout de même que, après cette épisode, "when Google users in China search for something that might return banned results, they see the links, but when they click on one, they are redirected to a government-approved site." Sans doute est-ce là ce que BC veut signifier en écrivant que Google « avait choisi dès sa première installation en Chine [...] que les adresses interdites ne génèrent même pas de message d'erreur ». En fait, dans le numéro du quotidien *Libération* déjà cité, auquel BC fait référence, deux journalistes se livrent à un exercice pratique : comparer les résultats proposés par Google en interrogeant le moteur de recherche en français à Paris et en chinois à Pékin. Ce qu'ils observent n'est pas que certains résultats seraient proposés qui ne renverraient pas à la page annoncée, mais simplement que les résultats indésirables n'apparaissent plus (Grangereau & Guenneugues, 2006). En vérité, selon Batelle (2006, p. 206), c'est dès le mois de février 2004 que Google, lançant une version en chinois de *Google News*, accepte de composer avec les autorités chinoises en purgeant l'index utilisé des pages regardées comme inacceptables par ces autorités. Une déclaration de la firme Google précisait alors ceci.

Google has decided that in order to create the best possible search experience for our mainland China users we will not include sites whose content is not accessible, as their inclusion does not provide a good experience for our News users who are looking for information.

Cela ne suffit évidemment pas à désamorcer la polémique à propos du comportement de Google en Chine. Aussi, par l'intermédiaire de son blog, Google fit une déclaration, fin septembre 2004, pour tenter à nouveau de se justifier. On y lisait notamment ceci (Google Team, 2004).

For Internet users in China, Google remains the only major search engine that does not censor any web pages. However, it's clear that search results deemed to be sensitive for political or other reasons are inaccessible within China. There is nothing Google can do about this.

For last week's launch of the Chinese-language edition of Google News, we had to decide whether sources that cannot be viewed in China should be included for Google News users inside the PRC. Naturally, we want to present as broad a range of news sources as possible. For every edition of

Google News, in every language, we attempt to select news sources without regard to political viewpoint or ideology. For Internet users in China, we had to consider the fact that some sources are entirely blocked. Leaving aside the politics, that presents us with a serious user experience problem. Google News does not show news stories, but rather links to news stories. So links to stories published by blocked news sources would not work for users inside the PRC -- if they clicked on a headline from a blocked source, they would get an error page. It is possible that there would be some small user value to just seeing the headlines. However, simply showing these headlines would likely result in Google News being blocked altogether in China.

We also considered the amount of information that would be omitted. In this case it is less than two percent of Chinese news sources. On balance we believe that having a service with links that work and omits a fractional number is better than having a service that is not available at all. It was a difficult tradeoff for us to make, but the one we felt ultimately serves the best interests of our users located in China.

Il semble donc que la gestion de situations du type « syndrome chinois » conduise à envisager trois possibilités dans le cas de liens conduisant à des sites « prohibés » : soit le lien est affiché, mais, lorsqu'on tente de l'actionner, on obtient un message d'erreur, et cela alors qu'on a pu voir le texte portant le lien, ce qui, selon Google, peut frustrer l'utilisateur ; soit le fait d'actionner le lien conduit bien à une page d'un site, mais d'un site autre que celui auquel le lien renverrait normalement ; soit enfin, pour épargner à l'utilisateur les frustrations engendrées par les deux situations précédentes, le lien n'apparaît pas sur la page de résultats. Du point de vue des savoirs utiles au bon usage des moyens d'accès aux ressources du Web, cette conclusion, tirée ici de l'examen d'une situation dramatique – celle de l'Internet en Chine –, doit sans doute être augmentée d'autres « cas de figure » qui peuvent survenir dans des contextes beaucoup plus ordinaires – tout utilisateur du Web connaît ainsi la page d'erreur 404, et chacun a un jour ou l'autre été l'innocente victime d'une « redirection », le lien proposé ne conduisant plus à la page annoncée ni même au site proposant cette page⁸⁹. Sans doute cela n'épuise-t-il pas les mystères de la recherche d'information sur Internet. Mais il y a là un point de départ possible – parmi d'autres – pour une enquête sur les savoirs utiles à la formation raisonnée à ce type d'activité.

De l'Internet au Web : invasion technique et déroute de la culture

La suite de l'encadré A comporte encore environ sept pages en petits caractères. Nous les parcourons rapidement, pour dresser un inventaire concis des objets que BC y a fait figurer,

un peu à la manière d'un dictionnaire qui fournirait ça et là des commentaires parfois un peu inhabituels⁹⁰. Le parcours commence avec les *protocoles TCP/IP*, qui gouvernent la transmission des données sur Internet. Ceux-ci sont suivis de la mention de l'ICANN, l'*Internet Corporation for Assigned Names and Numbers*, organisme mondial de gestion de l'Internet. Puis vient le *DNS (Domain Name System)*, qui gouverne le système des noms de domaine (avec notamment les extensions telles .org, .com, .fr, etc.), et qui inspire à BC une digression bien sentie à propos de l'attribution de l'extension .eu et des pratiques douteuses de la part de personnes ou d'institutions (y compris des villes ou des pays) désireuses d'acquérir des noms de domaines en .eu⁹¹. La puissance de l'ICANN est soulignée par BC à propos du refus de cet organisme, susceptible, dit-elle, d'« irréaliser » virtuellement un pays », d'attribuer « un nom de domaine aux sites “pour adultes seulement” et de faire place à un quartier rouge sur la toile » (p. 34), formulation qui comporte au reste une légère inexactitude⁹². BC note toutefois que ce genre de décisions, « prises aux environs de Washington », semblent aujourd'hui moins décisives qu'autrefois, dans la mesure où un moteur de recherche comme Google « permet par exemple [à l'utilisateur] de se passer du nom de domaine pour trouver une adresse ». Elle enchaîne ensuite sur l'*adresse IP* d'un ordinateur en soulignant alors que celle-ci peut susciter la crainte d'une surveillance à la *Big Brother* des activités digitales de tout un chacun.

L'encadré A se réfère ensuite aux services que l'Internet rend accessibles au public, en s'arrêtant essentiellement sur le *World Wide Web*, « la toile d'araignée mondiale ».

Telle est l'expression par laquelle la Commission générale de terminologie et de néologie a rendu en français l'expression anglaise de World Wide Web, où, pourtant, web ne renvoie pas d'abord à l'idée de toile d'araignée (en anglais spider's web ou cobweb). Dans le récit des débuts du Web qu'il fait le 2 novembre 1995 à l'occasion du lancement de la branche européenne du World Wide Web Consortium, Robert Cailliau, co-inventeur du Web avec Tim Berners-Lee, indique pour l'année 1990 : “During some sessions in the CERN cafeteria, Tim and I try to find a catching name for the system. I was determined that the name should not yet again be taken from Greek mythology. Tim proposes ‘World-Wide Web’. I like this very much, except that it is difficult to pronounce in French...” Pour ne prendre ici qu'un exemple, The Advanced Learner's Dictionary of Current English (Hornby et al., 1963) dans son édition de 1963 (Sir Timothy est né en 1955) propose d'abord, à l'entrée web, les deux acceptions suivantes : “1. Network (usu. fig.): a web of lies (deceit, intrigue). 2. sth. made of threads by a spider or other spinning creature: a spider's web (see cobweb).” (Le mot

désigne encore la palmure des canards et autres palmipèdes.) Dans son *Dictionary of Word Origins* (1994), John Ayto écrit à l'entrée web : "A web is etymologically something 'woven'. In common with Dutch web, Swedish väf, and Danish væv, the word goes back to a prehistoric Germanic *wabjam. This was derived from the base *wab-, *web-, which also produced English weave 'make cloth'." L'idée de toile d'araignée est donc seconde – d'après le *Online Etymology Dictionary* (Harper, 2001), elle apparaîtrait en anglais vers 1220.

BC rappelle d'abord le pastiche de l'appellation *World Wide Web* par l'expression *World Wild Web*, qui semble avoir été utilisée à l'origine pour signifier que, au plan économique et financier, le *Web* est très vite apparu aux firmes qui tentaient d'y faire connaître leur produits comme un monde « sauvage », où tant le développement explosif que l'écroulement sans appel pouvaient être rapides, en même temps que fluctuaient les prix et les cotes boursières des institutions concernées. La deuxième notation concerne *Timothy Berners-Lee*, dont le nom est étroitement associé à la saga du *Web* et que BC mentionne en relation avec les travaux sur la notion de « web sémantique » (question sur laquelle nous reviendrons). Elle souligne le rôle de l'*hypertexte*, mentionne le langage *HTML* (*HyperText Markup Language*), analyse la structure d'une adresse *Web* et évoque l'existence des *browsers* (logiciels de navigation).

Après cette rapide notule sur le *Web*, l'encadré A revient à l'Internet lui-même pour souligner l'existence de liens physiques, matériels ou immatériels, qui permettent la circulation des *bits*, mot dont BC note que, s'il apparaît en anglais comme la contraction de *binary digits*, il a aussi un sens propre, celui de *bouts* ou de *morceaux*. La référence philosophique ici n'est pas loin, avec l'idée leibnizienne de caractéristique universelle, partiellement réalisée par George Boole avec son algèbre en 0 et 1. Mais l'auteure revient alors à un autre fait fondateur : l'Internet est « historiquement américain ». Reprenant à ce propos des éléments de présentation et d'analyse empruntés à Huitema (1996), elle mentionne l'*IAB* (*Internet Architecture Board*), dont elle semble persifler la volonté de fonctionnement démocratique (fortement soulignée par Christian Huitema dans son livre), et conclut en donnant la parole à Louis Pouzin, tel qu'un article du quotidien *Le Monde* en brosse le portrait (Foucart, 2006)⁹³, celui d'un ingénieur français de très haut niveau, né en 1931, qui, après avoir inventé ce qui aurait pu devenir l'Internet – le réseau Cyclades –, travaille aujourd'hui à promouvoir, contre l'hégémonie de l'anglais, le multilinguisme sur l'Internet (Pouzin avec Bellin, 2007).

L'encadré B, intitulé « Alexandrie, Alex(andr)a, ou capitalisme et schizophrénie », se réfère à l'*Internet Archive* de Brewster Kahle⁹⁴, qui s'efforce, par des copies journalières du *Web*, de donner une réalité à la notion ambiguë de *Total Recall*⁹⁵. La comparaison est ici avec la bibliothèque d'Alexandrie : l'*Internet Archive* vise, précise BC, à « préserver l'héritage de l'humanité et [à] assurer l'accès à cet héritage », à ceci près cependant que chacun peut demander à n'être pas visité par le robot qui alimente l'*Archive*, échappant ainsi à cette thésaurisation d'ambition universelle⁹⁶. Le fondateur de l'*Internet Archive*, Brewster Kahle, crée la même année (avec Bruce Gilliat qui en est le PDG) la société Alexa, rachetée trois ans plus tard par la société Amazon.com. BC, qui rappelle la chose, s'empresse de noter que, selon un mélange typiquement américain, la philanthropie va ici main dans la main avec le souci du « business ». Elle évoque à ce propos le rôle de l'éthique protestante dans le développement du capitalisme – référence (implicite) à Max Weber –, puis revient momentanément à Google pour rappeler que cette firme « tire 99 % de ses revenus de la publicité » (p. 41). Mais alors que, par l'effet d'une schizophrénie bien ordonnée, Alexa est distincte de l'*Internet Archive* comme firme commerciale, Google rassemble sur *une même page* les résultats « philanthropiques » et les résultats « intéressés ». Plutôt qu'une véritable schizophrénie, semble s'interroger BC, ne serait-ce pas là, de façon typique, une manifestation de la *mêtis*⁹⁷ américaine ? Cette question posée, elle s'emploie à illustrer la complexité supposée de la « meule de foin » qu'est l'*Internet Archive* par un exemple qui, curieusement, n'a pas trait à une version ancienne de tel ou tel site Web, mais à un flux RSS à propos d'une chanson, dont, au demeurant, le texte (difficilement lisible) que BC reproduit et propose à son lecteur précise toutes les caractéristiques avec un sérieux documentaire radical pour le profane en la matière. Touchant *in fine* au thème (qu'elle développera longuement plus loin dans son livre) de la mort de l'imprimé (*death of print*), BC note que Brewster Kahle a lancé « récemment » (en octobre 2005) le projet d'une *Open Content Alliance* (OCA), qui se présente comme une réponse à l'initiative du *Google Book search*. Avec Tim Berners-Lee et le Web sémantique, note-t-elle, Brewster Kahle apparaît ainsi comme « une personnalité et un entrepreneur déterminants pour la suite des événements » (p. 43).

Avant de revenir à la suite des encadrés, observons que BC apporte fort peu d'informations sur le Web lui-même. De la même façon, son lecteur ne saura pas exactement ce qu'est *Alexa*, « rejeton commercial [de l'*Internet Archive*], qui a été acheté 250 millions de dollars en 1999 par Amazon.com ». En règle générale, le style est allusif et les gloses, commentaires et autres digressions l'emportent souvent sur la présentation élémentaire des objets évoqués. Le rapport à ces objets que le texte de BC peut induire chez le lecteur participe d'un premier repérage, parfois expéditif, d'éléments dont le rôle dans l'enquête de

l’auteure ne paraît pas toujours clairement intelligible. L’encadré C, ainsi, s’intitule un peu obscurément « Google et James Joyce : les yeux doux ». Avec l’aide de l’angliciste Jean Dixsaut (dont, nous l’avons vu, le nom figure dans la liste de remerciements), BC y traque le mot de *google* dans le *Finnegan’s Wake* de James Joyce, où le mot apparaît plusieurs fois (comme verbe). La structure du texte joycien tend à envahir l’exposé de la philologue, qui multiplie les associations verbales avant d’en venir enfin au verbe “to ogle”, bien vivant en anglais, où il désigne un regard amoureux, voire concupiscent, et est apparenté étymologiquement à “ogen”, « mirer » en néerlandais (langue dans laquelle le mot “oog” signifie « œil »). S’adonnant alors à une espèce de tachisme philosophique, elle évoque le verbe “to goggle” (qu’on peut définir rapidement comme signifiant “to look with amazement”, voire “to look stupidly”), précisant par exemple que *goggle-box* désigne en argot un poste de télévision. Le thème de la *bigbrotherisation* revient à nouveau en fin d’encadré, à propos d’un éditorial paru dans le *San Jose Mercury News* du 16 avril 2004, intitulé “If Google ogles your e-mail, will Ashcroft be far behind?”⁹⁸. Sur ce point comme sur quelques autres, l’enquête apparaît partielle, peut-être parce qu’elle est voulue partielle. Alors par exemple que BC tire sa référence à l’éditorial accusateur de la fameuse interview donnée à *Playboy* par les « Google Guys », elle ignore entièrement les réponses que ceux-ci apportent à l’interrogation du journaliste, et par exemple celle-ci, due à Sergey Brin (Google, 2004) : “Any web mail service will scan your e-mail. It scans it in order to show it to you; it scans it for spam. All I can say is that we are very up-front about it. That’s an important principle of ours.”

Le recours aux savoirs « indigènes » : à propos de doxa

L’encadré D est appendu au chapitre III intitulé « *Our mission is to organize all the information in the world* » ; il s’intitule lui-même *Qu’est-ce que la doxa ?* Bien entendu, on peut, par méthode, se demander, là encore, s’il est utile ou non d’approfondir cette notion, son histoire, ses déclinaisons, afin de mieux répondre à la question qui est l’alpha et l’oméga de la quête dont nous avons suivi jusqu’ici les tribulations : comment vivre avec Google ? L’auteure se trouve, en l’espèce, dans son domaine de prédilection, la philologie grecque⁹⁹. Le mot *doxa*, qu’elle emploie 28 fois dans son ouvrage, est, précise-t-elle, « l’un de mots grecs pour nous les plus polysémiques », que « la traduction française courante par “opinion” ne saurait », en conséquence, rendre exactement (p. 126). Elle retrace donc, sommairement, l’histoire de cette conceptualisation buissonnante : Parménide, « avec qui tout se met en place » ; Platon, qui oppose la science (*epistêmê*) à l’opinion, les philosophes aux philodoxes ;

Aristote, qui définit les *endoxa* comme « ce qui est reçu [*ta dokounta*] par tous ou par le plus grand nombre, ou par les savants [les sages, *sophois*], et, parmi ces derniers, ou par tous ou par le plus grand nombre, ou par les plus renommés et les plus réputés [*endoxois*] ». (La référence insistante aux experts, aux autorités, bref à « l’opinion des hommes illustres », fera florès dans certains discours.) Ce rapide parcours à propos de la notion de *doxa* conduit alors à une conclusion où l’auteure semble désespérer de la dialectique de la *doxa* et de l’*alêtheia* (la vérité).

Le savoir humain, en tant qu’il est information véhiculée et utilisée, qu’il sert de point de départ aux constructions communes et persuasives, est de l’ordre de l’opinion et non pas de la vérité. Tel est, de fait, le statut « ontologique » du classement des informations qu’on trouve sur Google, affine avec celui de l’information tout court. (p. 128)

Notons cependant que le système d’oppositions dans lequel est prise la notion de *doxa* semble plus ouvert, moins désespérant, en particulier chez Platon, du moins si l’on suit l’exposé d’Ivan Gobry dans son *Vocabulaire grec de la philosophie* (Gobry, 2000), où on lit par exemple ceci.

Chez Platon, le mot **doxa** revêt deux sens différents, dont le premier est original et le second classique. C’est dans le *Ménon* (97b-100b) qu’on trouve le premier sens : l’opinion est le premier degré de la vertu, celui de l’homme ordinaire, du non-philosophe, qui n’a pas encore conquis la vertu contemplative du sage ; elle est alors une divination spontanée du bien à faire dans le monde sensible. Dans sa nature, elle est une faveur divine (**theia moïra**). Elle n’a donc pas besoin d’être enseignée, contrairement à la doctrine socratique répercutée dans le *Protagoras* (357d) : “Elle n’est ni un don de la nature, ni le fruit de l’enseignement.” (*Ménon*, 99e). (p. 38)

Il y aurait ainsi une vision pessimiste, noire et une vision optimiste de la *doxa* : entre les deux, en tout cas, le choix de BC semble fait.

Traduction, trahison ? Une autre contre-enquête

Les encadrés E et F closent le chapitre IV, intitulé « *Don’t be Evil* » – la devise de Google. Le premier porte un titre de prime abord étonnant : *Le bras philanthropique*. Il s’agit en réalité de la traduction française, obtenue par l’auteure à l’aide du traducteur automatique de Google, d’un texte en anglais présent sur Google.org, *The philanthropic arm of Google*. L’exercice

auquel BC s'est ainsi livrée est classique, sans risques et, au reste, amusant, même s'il fait bon marché des problèmes de la traduction automatique : on trouvera ci-après un échantillon de ce à quoi il peut conduire ¹⁰⁰.

Welcome to google.org – the philanthropic arm of Google

Google.org aspires to use the power of information to help people better their lives. We are an experiment in active philanthropy. In addition to financial resources, we are fortunate to be able to engage Google's entire family of people and partners, information technologies and other resources to address three major growing global problems: [climate change](#), [global public health](#), and [economic development and poverty](#).

We know these are major concerns for a large number of individuals and institutions throughout the world, and google.org joins a community of like minded groups working to make the planet and population healthier and more equitable.

Bienvenue sur google.org, le bras philanthropique de Google

Google.org aspire à utiliser le pouvoir de l'information pour aider les gens à améliorer leur existence. Nous sommes actifs une expérience de la philanthropie. En plus des ressources financières, nous avons la chance de pouvoir engager Google ensemble de la famille des personnes et des partenaires, les technologies de l'information et d'autres ressources pour répondre à trois grands problèmes mondiaux de plus en plus : les changements climatiques, [la santé publique mondiale](#), et le développement économique [et la pauvreté](#).

Nous savons que ce sont des soucis majeurs pour un grand nombre de personnes et d'institutions à travers le monde, et se joint à google.org comme une communauté de vues des groupes de travail à faire et de la population de la planète plus saine et plus équitable.

La manœuvre cache toutefois une donnée essentielle : des systèmes de traduction automatique nettement plus évolués ¹⁰¹ sont couramment employés, depuis plusieurs décennies, dans les organismes internationaux pour fournir une première version des textes à traduire, version sur laquelle opèrent alors des correcteurs spécialisés. Mais enchaînons avec l'encadré F : intitulé *Déni de garanties*, il prolonge un développement du texte principal dans lequel BC identifie ironiquement Google à Dieu, un Dieu qui, « simultanément, se donne à lui-même le droit de changer de nature et de définition, voire de disparaître et de nous planter là », qui s'autorise sans préavis « à changer ses lois, à abandonner son organisation et ses organisés » (p. 152). L'encadré F reproduit simplement, à cet égard, un texte de Google où la firme de Mountain View énonce les conditions d'utilisation de ses services. Le paragraphe suivant en donnera une idée.

Google n'accepte aucune responsabilité d'aucune sorte en ce qui concerne la précision, le contenu, l'exhaustivité, la légitimité, la fiabilité, l'opérabilité ou la disponibilité des informations ou des données affichées dans les résultats obtenus par les Services de recherche Google. Google n'accepte

aucune responsabilité en ce qui concerne la suppression, l'impossibilité de stockage, la transmission incorrecte ou la transmission inopportune desdites informations ou des données. Google n'accepte aucune responsabilité en ce qui concerne les dommages susceptibles de résulter du téléchargement ou de l'utilisation des informations ou des données disponibles sur Internet via les Services de recherche Google. (p. 181)

Sur les sujets abordés dans les encadrés E et F, tout semble ainsi se passer comme si l'auteure pensait disposer *a priori* des connaissances utiles, pertinentes, pour porter un jugement – sans autre forme de procès – tant sur les problèmes de la traduction automatique que sur le style et le contenu des déclarations de la firme Google quant à sa responsabilité face à ses usagers. En chacun de ces cas, pourtant, un geste simple aurait consisté à enquêter sur le sujet ainsi problématiquement croisé, et cela en utilisant les ressources du Web, combinées éventuellement avec... les « services de recherche » de Google.

Commençons par la question des *conditions d'utilisation* – que Google aborderait de façon discrétionnaire et tyrannique. La requête "**conditions d'utilisation Google**" adressée au moteur de recherche Google propose une page de résultats dont nous reproduisons les trois premiers seulement.

Conditions d'utilisation Google
Conditions d'utilisation Google. Bienvenue chez **Google!** 1. Vos rapports avec **Google.** 1.1 L'utilisation des produits, logiciels, services et sites Internet ...
www.google.com/accounts/TOS?loc=fr&hl=fr - 47k - [En cache](#) - [Pages similaires](#)

Google Terms of Service - [[Traduire cette page](#)]
(D) **Google** is transitioning to no longer providing the Services to users in the ... 14.3 No conditions, warranties or other terms (including any implied ...
www.google.com/accounts/TOS?loc=FR - 43k - [En cache](#) - [Pages similaires](#)
[[Autres résultats, domaine www.google.com](#)]

[Ecrans - Critiqué, Google modifie ses conditions d'utilisation](#)
à la lecture des **Conditions d'utilisation** (Terms of Service) rédigées par **Google**, pour l'utilisation de son application d'édition en ligne, **Google Documents** ...
www.ecrans.fr/Critique-Google-modifie-ses,2678.html - 26k - [En cache](#) - [Pages similaires](#)

En cliquant sur le premier lien, on arrive aux conditions d'utilisation de Google en français.

Conditions d'utilisation Google
Bienvenue chez Google!
1. Vos rapports avec Google

1.1 L'utilisation des produits, logiciels, services et sites Internet de Google (désignés collectivement les « Services » dans les présentes et excluant tous services fournis par Google en vertu d'un accord écrit distinct) est soumise aux stipulations d'un accord juridique conclu entre Google et vous-même. « Google » désigne Google Inc., dont le principal établissement est sis au 1600 Amphitheatre Parkway, Mountain View, CA 94043, États-Unis. Les présentes explicitent la composition de l'accord et en exposent certaines stipulations.

1.2 Sauf disposition contraire expressément convenue par écrit avec Google, l'accord entre Google et vous-même comportera a minima et dans tous les cas les conditions et modalités exposées dans les présentes, ci-après dénommées les « Conditions universelles ».

Le deuxième lien renvoie au même texte, en français encore ¹⁰² ; la mention « Traduire cette page » permet d'obtenir une *traduction logicielle en français* du texte anglais relatif aux conditions d'utilisation : voici, sans plus de commentaires, le début de cette traduction.

Conditions d'utilisation Google

Bienvenue dans Google!

1. Votre relation avec Google

1.1 Votre utilisation de Google en matière de produits, de logiciels, de services et de sites Web (collectivement appelés les "Services" dans le présent document et à l'exclusion de tout service qui vous est fourni par Google dans le cadre d'un accord écrit séparé) est soumise aux conditions d'un juriste L'accord entre vous et Google. "Google" Google Inc moyens, dont la principale se trouve au 1600 Amphitheatre Parkway, Mountain View, CA 94043, États-Unis. Ce document explique comment l'accord est composé, et énonce certains des termes de cet accord.

1.2 Sauf accord écrit auprès de Google, vous êtes d'accord avec Google tient toujours compte, au minimum, les modalités et conditions énoncées dans le présent document. Telles sont ci-après dénommées les "Conditions universelle".

Le principal mérite du deuxième lien est de nous faire connaître ce qu'est vraisemblablement l'expression en anglais que Google rend en français par les termes « conditions d'utilisation » : *Terms of Service* (expression souvent abrégée en TOS). Le troisième lien, qui ne renvoie pas à un texte de Google mais se fait l'écho d'une critique à son encontre, confirme cette petite conjecture, puisqu'on y lit : « à la lecture des conditions d'utilisation (Terms of Service)... ». Le document correspondant apparaît sur un site Liberation.fr, *Écrans* : nous laisserons le lecteur intéressé s'y reporter ¹⁰³.

On peut en ce point utiliser les outils linguistiques que propose Google avec son service de traduction automatique. En tapant "**terms of service**" et en demandant une traduction de l'anglais au français, on obtient bien « conditions d'utilisation ».

Texte à traduire

Texte original :

Terms of service

[Texte traduit automatiquement :](#)

Conditions d'utilisation

Anglais à : Français



Traduire

Inversement si on tape « conditions d'utilisation » et qu'on demande à ce service de fournir une traduction en anglais, on arrive à l'expression *terms of use*. Nous limiterons notre petite enquête aux deux expressions ainsi rencontrées : "terms of service" et "terms of use". En

saisissant la première expression dans la boîte de recherche de Google, on retrouve (voir la copie d'écran ci-après), les conditions d'utilisation de Google (en deuxième position).

Yahoo! Terms of Service - [[Traduire cette page](#)]
Yahoo! provides its **service** to you subject to the following **Terms of Service** ("TOS"), which may be updated by us from time to time without notice to you. ...
info.yahoo.com/legal/us/yahoo/utos/ - 42k - [En cache](#) - [Pages similaires](#)

Google Terms of Service - [[Traduire cette page](#)]
All of these are referred to below as the "Additional **Terms**". Where Additional **Terms** apply to a **Service**, these will be accessible for you to read either ...
www.google.com/accounts/TOS - 44k - [En cache](#) - [Pages similaires](#)

USATODAY.com - [[Traduire cette page](#)]
Be aware that we may occasionally release information about visitors when release is appropriate to comply with law, to enforce our **Terms of Service** ...
www.usatoday.com/marketing/privacy-policy.htm - 39k - [En cache](#) - [Pages similaires](#)

Terms of Service - USATODAY.com - [[Traduire cette page](#)]
These **Terms of Service** govern your use of the USATODAY.com website (the "Site") only and do not govern your use of other USA TODAY services, ...
www.usatoday.com/marketing/tos.htm - 43k - [En cache](#) - [Pages similaires](#)

Terms of Service - [[Traduire cette page](#)]
Instructions on how to report violations of these **Terms of Service** may be located at the FAQ "How do I contact the Abuse team?" To expedite the processing ...
www.livejournal.com/legal/tos.bml - 41k - [En cache](#) - [Pages similaires](#)

Second Life | Terms of Service - [[Traduire cette page](#)]
This offer is conditioned on your agreement to all of the **terms** and conditions contained in the **Terms of Service**, including your compliance with the ...
secondlife.com/corporate/tos.php - 74k - [En cache](#) - [Pages similaires](#)

De ces conditions d'utilisation, nous ne reproduisons ici, volontairement, que le point 3 (le document complet en comporte 20).

3.1 Toute traduction de la version anglaise des Conditions vous est fournie le cas échéant par Google à des fins de commodité uniquement. Seules les versions des Conditions en langue anglaise régiront vos rapports avec Google.

3.2 En cas de contradiction entre les stipulations de la version anglaise et ceux de la traduction, les stipulations de la version en langue anglaise prévaudront.

Il n'est pas inutile de fournir l'original en anglais (puisque celui-ci fait seul foi).

3.1 Where Google has provided you with a translation of the English language version of the Terms, then you agree that the translation is provided for your convenience only and that the English language versions of the Terms will govern your relationship with Google.

3.2 If there is any contradiction between what the English language version of the Terms says and what a translation says, then the English language version shall take precedence.

Examinons maintenant le premier lien affiché ¹⁰⁴, qui renvoie aux “terms of service” de Yahoo!. Ces conditions d’utilisation s’appliquent à tout ce que propose Yahoo!, ainsi que le précise le point 2, “Description of Service”.

Yahoo! provides users with access to a rich collection of resources, including various communications tools, forums, shopping services, search services, personalized content and branded programming through its network of properties which may be accessed through any various medium or device now known or hereafter developed (the “Service”).

De nombreux passages de ce document montreraient la même attitude, impétueusement impérative, quant aux conditions imposées à l’utilisateur. Ainsi lit-on au point 14, “Modification of Service”, la clause suivante, dont le style brutal ne doit plus étonner.

Yahoo! reserves the right at any time and from time to time to modify or discontinue, temporarily or permanently, the Service (or any part thereof) with or without notice. You agree that Yahoo! shall not be liable to you or to any third party for any modification, suspension or discontinuance of the Service.

Plus loin, au point 19 (“Disclaimer of Warranties”), sous l’entête “You expressly understand and agree that”, on peut lire encore cette assertion définitive en lettres capitales : “YOUR USE OF THE SERVICE IS AT YOUR SOLE RISK”.

Le troisième lien concerne la “Privacy Policy” de USATODAY.com ¹⁰⁵ ; la page correspondante comporte un lien qui renvoie vers les “Terms of Service” du même quotidien en ligne. On y retrouve les mêmes précautions, qui pourraient paraître exorbitantes pour qui n’est pas familier de ce genre de document. Ainsi en va-t-il par exemple avec ce paragraphe de la section “Disclaimers”.

We do not warrant or make any representations regarding the use or the results of the use of the site or materials on this site or on third-party sites in terms of their correctness, accuracy, timeliness, reliability or otherwise.

De même, le lien qui renvoie aux conditions d'utilisation du site *Second Life* (qui se présente comme un "3-D virtual world entirely created by its Residents") fait découvrir un paysage désormais familier¹⁰⁶ : le texte correspondant indique au lecteur qu'il y trouvera "the terms on which Linden Research, Inc. ('Linden Lab') offers you access to its services." Ainsi lira-t-on au point 1.6 cette claire mise en garde : "Second Life is subject to scheduled and unscheduled service interruptions. All aspects of the Service are subject to change or elimination at Linden Lab's sole discretion". La suite de ce passage nous fait de même retrouver un type de discours familier.

Linden Lab reserves the right to interrupt the Service with or without prior notice for any reason or no reason. You agree that Linden Lab will not be liable for any interruption of the Service, delay or failure to perform, and you understand that except as otherwise specifically provided in Linden Lab's billing policies posted at <http://secondlife.com/corporate/billing.php>, you shall not be entitled to any refunds of fees for interruption of service or failure to perform. Linden Lab has the right at any time for any reason or no reason to change and/or eliminate any aspect(s) of the Service as it sees fit in its sole discretion.

Les exemples d'une telle rhétorique peuvent être aisément multipliés. Sur le site de MSN (Microsoft Network), on trouve ainsi, tout à fait dans la même veine, l'explicitation suivante¹⁰⁷.

17. NOUS NE FOURNISSONS AUCUNE GARANTIE.

Nous fournissons le service « en l'état », « avec toutes ses imperfections » et « tel que disponible ». Nous ne garantissons pas l'exactitude ni l'opportunité des informations fournies dans le cadre du service. Les parties Microsoft n'accordent aucune garantie ou condition expresse. Vous pouvez bénéficier de droits supplémentaires accordés aux consommateurs en vertu des lois de votre pays, que ce contrat ne peut modifier. Nous excluons toutes les garanties implicites de qualité, d'adéquation à un usage particulier, d'absence de défaut de fabrication et d'absence de contrefaçon.

Semblablement, le site de Microsoft offre une abondante "information on terms of use", où l'on peut lire par exemple, sous le titre "Acceptance of terms", ce propos dont le ton paraîtra comminatoire au non-initié¹⁰⁸.

The services that Microsoft provides to you are subject to the following Terms of Use ("TOU"). Microsoft reserves the right to update the TOU at any time without notice to you. The most current version of the TOU can be reviewed by clicking on the "Terms of Use" hypertext link located at the bottom of our Web pages.

Le même ton et les mêmes contenus prévalent encore s'agissant de la société Systran¹⁰⁹, qui se présente comme « le leader des solutions et produits logiciels de traduction en versions desktop, serveur et services en ligne ». Le texte des mentions légales que son site propose aborde en tout premier lieu la question de l'exactitude de la traduction, à propos de laquelle la condition suivante est stipulée¹¹⁰.

SYSTRAN s'efforce de fournir des traductions les plus exactes possible, toutefois **aucune traduction automatisée n'est parfaite ni prévue pour remplacer des traducteurs humains**. Les utilisateurs doivent être conscients de l'importance de la qualité du texte source pour la traduction.

Dans le même style, sous le titre « Contenu et responsabilité », on lit un peu plus loin : « SYSTRAN ne saurait être tenu responsable des erreurs ou omissions pouvant apparaître sur ce site Web, et se réserve le droit de le modifier sans préavis. » Mais venons-en maintenant à la question de la traduction automatique. Là encore, on peut, à l'aide d'une recherche simple et quasi immédiate, situer le document proposé par BC dans un contexte qui donne une perspective plus juste sur les usages positifs, même s'ils sont limités, des systèmes de traduction automatique. La requête "**uses of automatic translation**" présentée au moteur de recherche Google permet déjà de donner une certaine profondeur de champ à la question examinée. Le premier lien renvoie à l'article de l'encyclopédie Wikipedia intitulé "Machine translation". Cet article comporte notamment une section intitulée "Applications" où, une fois rappelé ce truisme que "no system provides the holy grail of 'fully automatic high quality machine translation' (FAHQMT)", le rédacteur anonyme note ceci.

Despite their inherent limitations, MT programs are used around the world. Probably the largest institutional user is the European Commission, which employs a highly-customised version of the commercial SYSTRAN MT system to automatically translate a large volume of document preliminary drafts for internal use.

Le même rédacteur (ou un autre) écrit encore ce qui suit.

Relying exclusively on machine translation ignores that communication in human language is context-embedded, and that it takes a human to adequately comprehend the context of the original text. Even purely human-generated translations are prone to error. Therefore, to ensure that a machine-generated translation will be of publishable quality and useful to a human, it must be reviewed and edited by a human.

Cela rappelé, l'utilité de la traduction logicielle apparaît multiple. Elle permet notamment, devant une langue qu'on ne connaît pas – au point de ne pouvoir déchiffrer même grossièrement un texte simple et court dans cette langue –, d'en obtenir une traduction approximative dans une langue connue. Voici par exemple le chapeau d'un article paru dans une revue en ligne rédigé en néerlandais ¹¹¹.

De Franse president Jacques Chirac heeft zich achter het Frans-Duitse zoekmachine-project geschaard waar het Franse Thomson en het Duitse Deutsche Telekom mee bezig zijn. De steun van Chirac voor het project is niet verrassend. Eerder beklagde Chirac zich al over de dominantie van Google.

Le lecteur francophone – supposé non néerlandophone – y repèrera évidemment la référence au « président Jaques Chirac » à « Thompson », à « Deutsche Telekom », et croira y voir la mention d'un certain projet (*project*). En revanche il butera peut-être sur “zoekmachine” et autres “Frans-Duitse”. Voici alors la traduction en anglais que fournit le service de traduction en ligne de Google.

Texte traduit automatiquement :

The French president Jacques Chirac has supported the French-German search engine project, the way in which the French Thomson and the German Deutsche Telekom are doing. The support of Chirac for the project is not surprising. Chirac has previously complained about the dominance of Google.

On y voit tout de suite que Jacques Chirac a apporté son soutien au projet de moteur de recherche franco-allemand tel que l'envisagent la firme française Thomson et la firme allemande Deutsche Telekom, etc. Si le service de traduction de Google ne propose pas de traduction directe du néerlandais vers le français, on peut lui demander de traduire en français la traduction en anglais. La version fournie (voir ci-après) paraîtra moins utile, car sa lisibilité semble nettement plus faible que celle de la traduction en anglais. Mais qui ne posséderait que

cette version-là pourrait encore, en utilisant ses propres connaissances, en rétablir conjecturalement le sens d'une manière raisonnable.

Texte traduit automatiquement :

Le président français Jacques Chirac a soutenu le franco-allemand de moteur de recherche du projet, la manière dont le français Thomson et l'allemand Deutsche Telekom font. Le soutien de Chirac pour le projet n'est pas surprenant. Chirac a déjà plaint de l'hégémonie de Google.

Un certain nombre de comparaisons, au demeurant, tendent à montrer que la traduction automatique offerte par Google serait actuellement supérieure aux services offerts par Systran dans leur version gratuite en ligne (même si Google ne propose pas actuellement de traduction du néerlandais vers le français). Voici par exemple ce qu'on obtient par le truchement du service de traduction de *AltaVista*¹¹² en demandant successivement la traduction du néerlandais vers le français, puis du néerlandais vers l'anglais.

Le président Jacques Chirac français s'est rangé derrière le projet de moteur de recherche franco-allemand où le Thomson français et le Deutsche Telekom allemand occupé être. Le soutien de Chirac pour le projet n'est pas étonnamment. Plutôt plaint Chirac lui déjà au sujet du dominantie de Google.

The French president Jacques Chirac has spawned himself behind the Franco-German search engine project where the French Thomson and the German Deutsche Telekom busy to be. The support of Chirac for the project is not surprisingly. Earlier pitied Chirac itself already concerning the dominantie of Google.

En dépit de leurs imperfections parfois redoutables, les systèmes de traduction automatique peuvent, ainsi qu'on l'a suggéré, être utilisés pour obtenir rapidement, même pour des textes d'une certaine longueur, une première version sur laquelle des traducteurs humains pourront ensuite opérer. C'est ainsi que le quotidien de langue castillane *El periódico de Catalunya* a lancé dès octobre 1997 une édition en catalan qui « clone » chaque jour l'édition en castillan. Dans un rapport sur cette innovation, la précision suivante est soulignée¹¹³ : “Por clónico entendemos que ambas ediciones son absolutamente idénticas, que se trata exactamente del mismo periódico y que la versión catalana es íntegramente traducida todos los días a partir de la castellana.” Bien entendu, une fois obtenue la traduction automatique d'un texte, des correcteurs s'affairent. Dans le cas du *Periódico*, ils sont 16 à temps plein et 16 à temps

partiel – ces derniers arrivent au journal à 18 heures seulement –, qui tous ensemble mettent environ 4 heures pour établir la version définitive en catalan d'un numéro de 70 à 80 pages. Par comparaison, on estime que, sans la version préalablement obtenue par traduction automatique, il faudrait 80 correcteurs et un temps double pour parvenir au même résultat.

Un moteur de recherche européen ?

Le livre de BC se termine par un ultime encadré, l'encadré G, intitulé *Airbus numérique !* L'auteure y reproduit simplement le texte d'une présentation faite à l'AFP en janvier 2007 par Jean-Luc Moullet, « chef de file du projet et vice-président solutions logicielles de Thomson »¹¹⁴. Depuis cette époque, plusieurs changements sont intervenus à propos de chacun des deux projets commentés dans le texte cité : Galileo, d'une part (Ricard, 2007), Quaero, d'autre part. Ce dernier projet avait été annoncé en avril 2005 par Jacques Chirac (alors président de la République Française) et Gerhard Schröder (alors Chancelier de la République fédérale d'Allemagne), lors d'un conseil des ministres franco-allemand tenu à Reims. La presse – et plus généralement l'opinion publique informée – en retiendra qu'il s'agissait alors de concevoir et de mettre au point un moteur de recherche « européen » en vue de combattre la domination de Google en ce domaine. Lorsqu'elle écrit son livre, BC consacre à ces deux dossiers un peu plus de quatre pages (pp. 221-225), dont les deux dernières concernent le projet Quaero. L'examen auquel elle procède se conclut sur cette note désenchantée (qui concerne en vérité le moteur de recherche Exalead, auquel Quaero « est lié ») : « Bref, il est clair qu'il ne s'agit pas d'une autre approche de la recherche mondiale, c'est pour l'instant une vision directement entrepreneuriale, qui permettra peut-être de marginaliser un peu PageRank – plutôt de la concurrence traditionnelle qu'un nouveau concept. » Quelques pages plus haut (p. 209), elle avait toutefois exprimée à nouveau son vœu le plus cher, dont la réalisation vaudrait bien quelques accommodements avec l'ordre de l'argent : « Que Quaero (le programme) ou Exalead (le moteur) gagnent autant d'argent que Google s'ils le peuvent, du moment qu'ils donnent accès pour tous à de bonnes et inventives structurations des œuvres et de la connaissance. »

Conclusion

La formule du schéma herbartien présentée dans l'*Introduction générale* peut s'écrire, on le sait, $(S(X; Y; Q) \rightarrow M) \hookrightarrow R^\forall$, où $M = \{ R_1^\diamond, R_2^\diamond, \dots, R_m^\diamond, O_{m+1}, O_{m+2}, \dots, O_n \}$. Dans l'enquête dont nous concluons ici l'examen, la question *génératrice* $Q = Q_0$, jugée « imparable » par BC, est à peu de chose près celle-ci : « Comment vivre vraiment mieux, ou au mieux, avec le Net et avec quelque chose comme Google ? » La question est large, sans doute ; mais c'est bien elle, manifestement, qui mobilise l'enquêtrice et lui fait trouver l'énergie et la persévérance nécessaires à son enquête. Bien entendu, le fait que BC soit pour le moins dubitative, voire suspicieuse à l'endroit des réalités que cette enquête vise (le Net, Google) colore ses pensées et ses gestes ; mais ils ne les déterminent pas complètement. (Si besoin en était, on trouverait dans l'ouvrage déjà mentionné de Marie-Monique Robin sur l'activité de la firme Monsanto, dont cette auteure ne brosse certes pas une hagiographie, un contre-exemple évident à la thèse contraire.) Une certaine animosité inaugurale explique (en partie) le fait qu'il y ait enquête ; elle n'en explique pas finement la structure ou le contenu. De ce contenu, nous voudrions donc tirer, à titre de conjectures vraisemblables, certains enseignements instructifs du point de vue de la didactique de l'enquête codisciplinaire et de ses outils.

Avant cela même, notons que c'est bien essentiellement la même question Q_0 (mais dépouillée de toute connotation acrimonieuse) que nous pouvons mettre au principe de notre propre travail. Pourtant, ce qui nous importe ici, ce sont moins les éléments de réponse que l'auteure a pu tenter d'apporter dans son livre que sa manière d'y parvenir – ses matériaux, ses outils, l'usage qu'elle en fait. La suite de ce travail fondera et présentera nos propres éléments de réponse à la question Q_0 : on verra alors que le procédé de « fabrication » de notre réponse comme le matériau utilisé seront en ce cas fort différents de ceux maniés par BC. Dans les développements qui suivent, nous nous en tiendrons cependant, répétons-le, à ces aspects de l'enquête conduite par BC qui, nous semble-t-il, relève de plein droit de la didactique de l'enquête codisciplinaire en construction aujourd'hui dans le cadre de la théorie anthropologique du didactique.

Dans cette perspective, l'un des premiers points qu'il convient de noter est sans doute le suivant. Ici comme en toute enquête sur une question donnée, l'enquêtrice est conduite à

rencontrer et, parfois, à décider d'étudier (de façon plus ou moins approfondie) des questions « engendrées » par l'enquête sur Q_0 . En l'espèce, on peut brosser le tableau (partiel) suivant du questionnement dégagé par l'enquêtrice, en ne retenant que les plus saillantes des questions que nous avons croisées.

- Q_1 Comment peut-on envisager d'organiser toute l'information du monde ?
- Q_2 Que signifie « organiser », et contre quoi cela se fait-il éventuellement ?
- Q_3 . Que signifie la devise de Google, *Don't be evil* ?
- Q_4 . De quelle « mission » est-il question lorsque Google se réfère à « sa mission » ?
- Q_5 . Qu'est-ce que l'algorithme PageRank et quels principes de classement incorpore-t-il ?
- Q_6 . Quels rapports d'affinité ou d'opposition PageRank entretient-il avec la démocratie ?
- Q_7 . En quoi Google participe-t-il à une possible « *Bigbrotherisation* » de nos sociétés ?
- Q_8 . Que s'est-il passé entre Google et les autorités chinoises ?
- Q_9 . D'où vient le mot « Internet » ? Et la chose elle-même ?
- Q_{10} . Comment Internet est-il géré ? Par quelles institutions ?
- Q_{11} . Qu'est-ce l'*Internet Archive* ?
- Q_{12} . L'Internet et Google annoncent-ils la mort de l'imprimé ?
- Q_{13} . *Quid* d'un moteur de recherche « européen » ?

Dans le même temps, ainsi qu'on l'a assez souligné, la même enquêtrice *rate* un certain nombre de questions qui nous sont apparues cruciales dans notre propre contre-enquête, telle par exemple celles-ci.

- Qu'est-ce qu'un *mission statement* ?
- En quoi consiste exactement le *mission statement* de Google ?

Cela noté, la manière dont l'étude est conduite peut inspirer deux séries de remarques qui nous paraissent essentielles. La première touche aux *outils de l'étude*. Une fois fixé les *objets d'étude* que portent en elles la question génératrice initiale (Q_0) et les questions engendrées subséquentes (Q_1 , Q_2 , etc.), il reste en effet à déterminer les *outils* pertinents à l'étude de ces questions. Il s'agit là une question fondamentale de didactique de l'enquête codisciplinaire, notamment parce que l'enseignement scolaire en fait, en règle générale, l'économie en ne demandant aux élèves que d'*appliquer* un savoir *préalablement enseigné* à tel type de problèmes pour lequel ce savoir fournit toute faite, ou quasiment, une technique de résolution. Or ce mode d'organisation de l'étude – que l'on peut baptiser (en reprenant les mots d'un

certain jargon scolaire) le schéma *Théorie/Applications* – ne convient plus dès qu'on ne peut raisonnablement supposer un univers « organisé », lieu d'une harmonie préétablie (au plan épistémologique, s'entend), où toute question possible trouverait dans les connaissances *existantes* les principes de sa réponse. L'habitus scolaire à cet égard est net : subjectivement, il porte à penser que toute difficulté rencontrée devrait céder en faisant peser sur elle le poids de ce que nous savons *déjà*, à l'exclusion de toute connaissance nouvelle *pour nous* – sans parler de connaissance neuve *absolument*. C'est là une modalité de ce que Yves Chevallard (2003) a nommé plus généralement l'*adultisme*, posture cognitive niant en acte le caractère ouvert du champ des connaissances utiles – existantes ou à produire – et qui implique concrètement l'incapacité pratique à envisager d'apprendre *encore*.

C'est en ce point, soulignons-le aussi, que le schéma herbartien engage le fer avec une certaine éducation scolaire où ce qu'on *sait*, ici et maintenant, serait le critère clé, contre une éducation ouverte – « herbartienne » – où ce qui compte est *ce qu'on peut (encore) apprendre* – ajoutons : qui ait quelque utilité supposée pour le projet que l'on se propose, ici et maintenant. Ce n'est pas, au reste, que l'enquête de BC ne fasse place à des connaissances manifestement neuves pour elle, bien au contraire ! Mais ces connaissances paraissent souvent dépourvues de fonctionnalité par rapport à la question au principe de l'étude : elles nous donnent fréquemment, avouons-le, le sentiment de participer d'une thésaurisation épistémologique qui, au-delà de l'érudition apparente et de la critique sans fard, trouve en elle-même sa propre fin. Or tant l'absence de fonctionnalité des connaissances nouvelles glanées ici et là que le retranchement derrière des savoirs anciennement et hautement maîtrisés est à peu près le contraire de l'ambition éducative « herbartienne » : car c'est pour répondre à la question *Q*, qui est pour moi une *vraie* question, une question *vive*, voire *vitale*, que je me penche sur les œuvres O_{m+1} , O_{m+2} , ..., O_n évoquées ci-dessus, dont plusieurs me seront nouvelles, et que j'aurai donc à « apprendre » *ab ovo*, pour expliquer, pour comprendre, pour *faire* – plutôt que pour surprendre, pour dominer, pour donner congé. Si les parcours d'étude et de recherche (PER) ont été introduits en TAD comme des dispositifs clés d'une formation scolaire à venir, c'est pour leur productivité espérée en termes d'apprentissages « adventices » (mais cruciaux), non moins que pour recentrer de façon délibérée, explicite, assumée l'éducation des jeunes générations sur ces questions qui, à un moment donné, *comptent* pour le présent et le futur de la société où ils doivent entrer. Nous plaçant ici dans une perspective de recherche visant à explorer plus avant la viabilité écologique et économique de ce projet éducatif, nous ne saurions, bien entendu, ignorer qu'il est des assujettissements personnels si puissants que, s'ils n'interdisent pas à la personne de se mettre en route à propos de telle question, s'ils y poussent même en certains cas, ils en fixent

par avance le parcours, les haltes, les péripéties, sans grand égard pour ce que le réel parcouru pourrait nous conduire à apprendre. C'est là une limite certaine au projet d'une éducation « par les PER » ; et il n'est pas sans importance que, à travers l'enquête de BC, nous l'ayons frontalement rencontrée, même si là n'est pas l'objet central de notre travail.

Les remarques précédentes n'impliquent nullement que l'on doive jeter par dessus bord ce que nous avons nommé les savoirs indigènes. Sauf contre-indication évidente, il est bien sûr raisonnable de « faire jouer » les connaissances qui nous sont familières, dont nous maîtrisons le mieux l'emploi, pour tenter de leur faire rendre, sur la question étudiée, ce qu'elles peuvent donner. Sans doute même est-ce par là qu'il convient de commencer, et peut-être de revenir à certains moments de l'enquête. Dans le cas que nous avons examiné, il semble pourtant que cette stratégie soit cruellement décalée par rapport aux problèmes affrontés. Nous donnerons ici un ultime exemple de ce mode d'attaque, qui montre l'auteure, dans les dernières pages de son essai, tentant en quelque sorte de *modéliser* la réalité profuse de Google à l'aide d'une conceptualisation portée par la langue grecque ancienne, qui lui est si familière.

Google est à coup sûr du côté de l'invention prométhéenne : une intelligence rusée, liée à un savoir-faire technique simple et efficace, d'ailleurs protéiforme et prompt à saisir l'occasion. Cette description complexe serait beaucoup plus parlante en grec. Elle se ferait (et ce sont parfois des mots que, chemin faisant, j'ai utilisés dans ma description) en termes de *mêtis* « plan, plan habile, sagesse habile et efficace, ruse », qui caractérise le divin Ulysse, le dessein de Zeus et la mobilité tentaculaire du poulpe ; de *tekhnê* « savoir-faire, métier, technique, art, compétence, expertise, manière de faire, moyen, système, artifice » ; de *mêkhanê* « moyen, trouvaille, invention ingénieuse, machine (de guerre), machinerie (de théâtre), machination, expédient, truc, machin, talent, habileté, art, ressources », de *kairos* « point critique, moment opportun, instant propice, à propos, occasion, avantage, profit », de *kerdos* « gain, profit, avantage, amour du gain, desseins profitables ». Elle relèverait aussi du vocabulaire de la rhétorique, en termes de *prepon* « ce qui se distingue, se montre, s'annonce par son extérieur, ce qui a l'air de, ce qui a rapport avec et convient à, le bienséant, l'adapté à l'auditoire comme au sujet traité » ; et surtout en termes de *doxa*, *dokounta*, *endoxa*, « opinion, réputation, apparence, semblant, croyance », « apparences, jugements qui paraissent convenir, sont crédibles et sur lesquels un consensus peut s'établir », « idées reçues, opinion établie, prémisses solides pour des raisonnements (sic) probables ». (p. 244)

Comme BC le dit elle-même, il s'agit là d'une *description* – qui explique peu car expliquer n'est pas sa visée. La modélisation proposée, au vrai, ne produit pas de connaissances ni de

questions nouvelles. D'une façon plus générale, on l'a constaté, le rendement de ce type de stratégie est des plus faibles : le fait est connu, comme en témoigne par exemple cette notation, due à l'historien Paul Veyne, à propos de Michel Foucault (Veyne, 2008, p. 158).

Dans la dernière interview que la vie lui a permis de donner, il s'est exprimé très clairement : on ne trouve jamais la solution d'un problème actuel dans la réponse d'une autre époque, car celle-ci répond nécessairement à une question qui était différente. Il n'y a pas de problèmes qui traversent les siècles, l'éternel retour est aussi un éternel départ (il aimait ces mots de René Char).

En plusieurs cas, une telle manière d'opérer se révèle même contre-productive, au point de conduire à des paralogismes regrettables : que l'on songe ici à la tentative d'élucider le "*our mission*", qui manque l'essentiel. L'avantage est, sans doute, que l'on opère sur son propre territoire, où l'on a ses habitudes et ses aises. Mais on juge l'arbre à ses fruits ; or, on l'a vu en plusieurs cas, la greffe épistémologique tentée ne prend guère.

Il y a d'ailleurs en tout cela, notons-le aussi, une surprenante anomalie épistémologique. Dès lors en effet que l'étude d'un objet est suffisamment avancée, dès lors qu'on en a acquis quelque *working knowledge*, l'objet étudié devient, en règle générale, un *outil* de sa propre étude. L'historien étudie une question d'histoire en s'outillant de la science historique disponible, même s'il est par là conduit à remettre en question certains de ces outils, et même s'il s'adjoint d'autres outils, empruntés à d'autres disciplines de connaissance – économie, anthropologie ou biologie, par exemple. Le mathématicien, de même, étudie une question de mathématiques essentiellement à l'aide d'outils mathématiques : c'est même là un cas limite où, à l'hétérogenèse qui prévaut en nombre de domaines, semble se substituer une monogenèse fortement affirmée, qui porte à croire que les mathématiques ne devraient rien à ce qui n'est pas strictement elles-mêmes. Dans le cas qui nous occupe, l'étude de l'Internet et de Google devrait ainsi fournir des outils coextensifs à l'objet étudié. S'il en est bien ainsi dans les contre-enquêtes auxquelles nous avons procédé, force est de constater que rien de tel, quasiment, n'émerge dans le compte rendu d'enquête fourni par BC.

L'usage que nous avons fait du couple constitué de l'Internet et du moteur de recherche Google se laisse résumer en peu de mots. Tout d'abord, cet outil nous a permis, très classiquement, de conduire des « recherches » (*searches*), plus ou moins approfondies selon le cas, sur une foule de sujets : sur les « Barbara Cassin » de l'Internet, sur les activités publiques des personnes remerciées par l'enquêtrice (la « tribu »), sur le *globish*, sur Victor Klemperer, sur la notion de *Aufhebung*, sur le Dieu Pan et la panique, sur la devise *In God We Trust*, sur le verbe *to organize* et les langues européennes, sur l'expression "*all the*

information in the world” et son emploi chez Google et ailleurs, sur l’algorithme PageRank, sur les notions d’*internet*, d’*intranet* et d’*extranet*, sur le mot anglais *web*, sur la devise *Don’t be evil*, sur l’utilisation dans la presse de la traduction automatique, etc. Ensuite et surtout (au plan de la méthodologie), le même assortiment d’outils nous a permis un type de geste crucial : au lieu de laisser aux objets apparaissant dans l’enquête leur statut immédiat de pures singularités, qu’on ne pourrait que commenter, fût-ce savamment et indéfiniment, nous avons cherché à replacer ces objets dans des *séries* paradigmatiques, qui les situent au sein d’un ensemble *homogène*, seule opération susceptible d’en faire apparaître le sens. Ainsi a-t-on vu la singularité de première prise du syntagme “*our mission is...*” s’évanouir dès lors que celui-ci se trouve situé dans son véritable univers langagier, celui des « organisations » (à but lucratif ou non) du monde états-unien (et, de façon croissante, des mondes soumis à son influence) ; et de même pour l’expression “*our mission is to organize*” ou pour celle de “*terms of service*”.

Ces contre-enquêtes rendent patents, par contraste, deux faits essentiels. Le premier est en quelque sorte un universel de la vie praxéologique : quoique référée au moins symboliquement à un « collège invisible », quoique développée au sein d’une petite « société de semblables », l’enquête de BC pâtit, nous semble-t-il, de n’avoir pas participé véritablement, presque substantiellement, de ce qu’on nommerait, en d’autres contextes, le travail d’une *communauté scientifique*, dont la culture praxéologique sur les questions qu’on y étudie non seulement oriente les recherches (ce qui peut, certes, gêner la découverte) mais aussi bien interdit certains égarements, certaines bévues, certains fourvoiements qu’un moindre contrôle épistémologique laisse bientôt resurgir. Qu’une telle supervision par une culture appropriée s’affaiblisse et l’on entendra aussitôt ici ou là, comme on dit, « beaucoup de bêtises » ! L’effet est d’autant mieux assuré que la question dont nous poursuivons l’étude dans ce travail ne concerne pas, rappelons-le, tel collectif d’experts particulier, mais bien le collectif – si l’on peut dire – *des citoyens* tout entier : vers quelle culture praxéologique faudrait-il que nous tendions tous ensemble pour vivre mieux avec le Net, et en particulier pour ne pas être des victimes impuissantes – sinon innocentes – des approximations et contrevérités qui, à tout moment, peuvent fuser à son propos ? Cette question conduit, on l’a vu, à un second constat : pour tenter d’être au mieux avec Internet, il faut à l’évidence chercher à être mieux avec la *recherche d’informations* sur Internet – la RII –, en se rendant capable d’en faire un usage pertinent, acéré, contrôlé. Ce constat spécifie la question précédente, sans réduire véritablement le champ de l’enquête : vers quelle culture praxéologique faudrait-il que nous tendions tous ensemble pour penser et pratiquer la RII à

notre avantage – sans être victime des fables et autres inventions qui, à tout moment, peuvent nous tomber dessus ? Telle est la question qui nous occupera désormais.

ANNEXE 2

Les « aides à l'étude » : une somme de compétences et d'expériences

Qui sont les personnes mentionnées par BC dans ses remerciements ? Sans prétendre entrer dans la vérité de relations « didactiques » dont nous ne connaissons au mieux que quelques-uns des effets, nous donnons ici un petit nombre d'indications qui nous ont paru pertinentes pour peindre l'environnement possible de l'enquêtrice et de son enquête.

Christine Buci-Glucksmann est une philosophe, professeur émérite à l'Université de Paris VIII où elle a enseigné dans le cadre de l'UFR « Arts, Philosophie et Esthétique ». Dans la présentation d'une conférence intitulée « Vers une esthétique du virtuel », dont le texte est en ligne sur le site du Centre interdisciplinaire de recherches sur l'esthétique du numérique (CIREN) de l'Université de Paris VIII, on trouve notamment les lignes suivantes, où l'on voit se tresser nouvelles technologies et esthétique extrême-orientale.

Le développement des nouvelles technologies du virtuel a entraîné une transformation historique majeure qui touche toutes les pratiques artistiques : le passage d'une culture des objets et des stabilités à une culture des flux et des instabilités. On s'interrogera sur les modalités et les implications d'une telle mutation, en mettant en valeur une nouvelle théorie de l'image : l'image-flux qui ouvre une approche polysensorielle des dispositifs de l'art.

C'est pourquoi les images-flux sont inséparables de nouvelles formes du temps : temps machinique et fluide, éphémère. On dégagera l'esthétique du virtuel qu'elles mettent en œuvre à partir de ses nouveaux modèles : les topologies, les artefacts, les enveloppes, les surfaces et secondes peaux. De tels modèles propres à l'art comme à l'architecture donnent naissance à une esthétique des transparences et des fluidités, qui renoue avec l'éphémère comme valeur positive de l'art propre au Japon, qu'il s'agisse d'êtres artificiels, de lieux ou de paysages.

Notons ce postulat subreptice : l'Internet serait du côté du labile, du fluant, de l'instable.

Laurent Catach, responsable multimédia des dictionnaires Le Robert, a été le directeur éditorial de la version cédérom du Grand Robert. De plain-pied dans le monde du multimédia,

il souligne dans plusieurs déclarations la richesse hypertextuelle du cédérom dont il a dirigé la réalisation, tout en défendant logiquement les productions traditionnelles sur papier. Dans un entretien donné à un magazine en ligne pour les professionnels de l'informatique, *ITR Magazine.com*¹¹⁵, Laurent Catach rappelle que le Grand Robert comporte quelque 13 500 pages et constitue un ensemble « d'une richesse sans égale » ; il poursuit ce panégyrique dans les termes suivants.

Mais l'intérêt majeur de ce CD-ROM est de pouvoir enfin révéler et déployer l'extraordinaire structure interne de l'ouvrage, qui est à ma connaissance le plus vaste réseau hypertexte jamais conçu au sein d'un même dictionnaire – et peut-être même au sein d'un même ouvrage. En effet, le Grand Robert contient de multiples types de renvois internes, notamment des renvois analogiques (des mots vers d'autres mots) mais aussi des renvois de citations, d'expressions, de tableaux, etc. Tous ces renvois ont été analysés et codés de façon à fonctionner sous forme de liens hypertextes directs dans la nouvelle version électronique, ce qui procure une facilité de navigation totalement inédite.

Tout cela est posé en opposition discrète au monde Internet proprement dit. C'est ainsi que, interrogé pour le quotidien *Le Monde* à propos de Wikipedia (Mathieu, 2005), Laurent Catach déclare que c'est là « un sujet intéressant, comme un laboratoire d'usage, une espèce de grand tableau où tout le monde peut écrire », avant de préciser sa position en ces termes.

Wikipedia est l'inverse d'une demande éditoriale. Celle-ci consiste à sélectionner ; or il n'y a pas, ici, de notion de sélection. Le jour où la moitié des articles de ce site seront signés par des professeurs d'université ou des Prix Nobel, il aura gagné ses lettres de noblesse.

On voit affleurer là une valorisation fort répandue, semble-t-il : celle du monde académique classique, et surtout de ses plus hautes cimes, seules capables d'autoriser ce qui s'écrit ici ou là et de le valider.

Daniel Ferrer est un angliciste dont le doctorat d'État abordait les œuvres de James Joyce, Virginia Woolf et William Faulkner. Directeur de l'Institut de Textes et Manuscrits Modernes (ITEM) de 1994 à 1998, il s'affronte aux difficiles problèmes des « éditions génétiques hypertextuelles », c'est-à-dire à la « représentation hypertextuelle des dossiers génétiques complexes »¹¹⁶. À nouveau, une configuration paradigmatique s'affirme : sur une assise des plus classiques, en elle-même hautement valorisée, vient s'inscrire une couche de

modernité que le mot d'hypertexte résume – même s'il ne s'agit jamais que de structurer des univers quasi fermés, où de surcroît les droits d'entrée sont élevés.

Jean-Pierre Guillemant est directeur des systèmes d'informations de la société Armatis, qui, avec ses quelque 2200 employés, offre des services de télémarketing et de centres d'appels, et dont *Anne-Sophie Legendre* est le contact pour la presse : le point d'appui est ici clairement dans l'univers économique engendré par les technologies modernes de communication.

Valérie Tesnière – dont le nom est incorrectement orthographié dans les « Remerciements » – appartient à un autre univers : elle occupe d'importantes fonctions à la Bibliothèque Nationale de France et, plus largement, dans le monde des bibliothèques (elle a ainsi présidé le jury du concours de recrutement des bibliothécaires ouvert au titre de l'année 2007). Elle est notamment l'auteure (avec Noémie Lesquins) d'une étude parue en mars 2006, intitulée *La bibliothèque numérique européenne. Une stratégie culturelle de la Toile* (Tesnière & Lesquins, 2006). Derechef, on retrouve le paradigme dans lequel des professionnels ancrés dans un domaine traditionnel, de bonne noblesse culturelle (en l'espèce, les auteures mentionnées sont l'une comme l'autre archiviste-paléographe), ont à se confronter aux nouvelles technologies et se trouvent pris entre une éthique du métier, d'une part, et des enjeux idéologiques, économiques et politiques, d'autre part¹¹⁷. Si, bien entendu, elles ne peuvent ignorer le monde Internet, ces auteures l'abordent avec prudence, non sans marquer des réserves à propos de certains thèmes devenus rapidement des classiques du genre, comme dans ce passage à propos de la fiabilité des résultats d'une recherche sur Internet.

Beaucoup déplorent aujourd'hui l'opacité des critères d'affichage de Google (page ranking) et de ses algorithmes. C'est un entrepôt. La transparence des critères de classement des résultats ainsi que la certification de l'origine de la source de numérisation sont des atouts permettant de rallier l'essentiel de la communauté des chercheurs et des médiateurs de l'information, avec valeur d'usage pédagogique pour le grand public.

Caroline Wiegandt a réalisé un parcours qui l'a conduite du professorat d'histoire-géographie à la Bibliothèque Nationale de France, où elle est depuis 2001 directrice des services et des réseaux. Présidente de l'association des professionnels de l'information et de la documentation (ADBS) depuis juin 2003, elle adopte dans ses déclarations une posture ferme mais irénique, soucieuse de concilier ancien et nouveau. Dans un entretien à *Captain-doc*, le « guide de la documentation électronique », elle fixe ainsi pour objectif à l'association qu'elle

préside « d'ouvrir les professionnels de la documentation à la nécessaire modernisation de leur métier, tâche qui n'est pas aisée pour un métier qui se voit bouleversé en profondeur par l'arrivée des nouvelles technologies de l'Internet ». S'agissant des documentalistes, elle indique alors sans complaisance l'effort qu'il leur faudra assumer, dans un partenariat ouvert avec les « informaticiens » (Wiegandt, 2005).

La gestion moderne des documents, l'organisation des contenus et des flux, tout cela demande la compétence de véritables documentalistes, qui savent se servir des outils. Ce qui implique aussi de leur part une formation pour adapter leur savoir-faire aux outils mais aussi faire le marketing de leur fonction.

La problématique de l'articulation pacifique et féconde entre informaticiens et documentalistes mérite d'être soulignée : elle introduit une touche relativement rare dans un univers livré souvent aux charmes parfois violents et toujours ambigus de l'*agôn*.

D'une autre manière, à partir de l'autre rive, *Peter Wittenburg*, dont la formation est en électronique et en informatique, et qui travaille au *Max-Planck-Institute for Psycholinguistics* de Nimègue (*Nijmegen*, Pays-Bas), évolue lui aussi dans un univers où l'ancien et le nouveau entretiennent des rapports organiques, celui des *Language Resource Archives* (LRA).

Language Resource Archives are centres that store digital language-related data of various types, ranging from simple texts to complex structured encyclopedia extended by multimedia objects and layers of relations and to complex structured collections. Language Resource Archives make these data accessible to interested users and that also offer solutions for long-term preservation of the data.

La fonction de préservation de données linguistiques assumée par les LRA concerne notamment les langues « en danger », à propos desquels un réseau international a travaillé dans le cadre d'un programme intitulé DOBES (*Dokumentation Bedrohter Sprachen*)¹¹⁸ : exemple de ce fait que les nouvelles technologies permettent quelquefois, d'une certaine manière, de « sauver » des mondes qui se meurent.

Xavier Perrot appartient à une autre génération. Enseignant, dans le cadre de l'université de Paris I, à l'IUP « Métiers des Art et de la Culture », mais aussi professeur à l'École du Louvre, il s'est spécialisé en muséologie et nouveaux médias (sa thèse portait sur la « production des hypermédias et des interactifs multimédias pour les musées »). Après avoir travaillé d'abord comme développeur et producteur multimédia, il a été associé au

Studio for Creative Inquiry de la Carnegie Mellon University à Pittsburgh (États-Unis). Expert en produits culturels multimédias, il conseille sur ce sujet diverses organisations, en Europe et aux États-Unis. Et, comme nous le savons, il a su être un conseiller « Informatique et Internet » crucial pour BC.

Parmi les membres de la petite tribu connivente (ou, du moins, aidante) auxquels nous nous limitons ici, on doit compter au nombre de ceux qui ont un commerce des plus sérieux avec l'informatique celui que BC désigne – sans grande originalité – comme un « inventeur » : *Roland Moreno*. L'inventeur de la carte à puce (1975), donc, semble en vérité se situer à quelque distance du reste de la tribu : peut-être est-il d'ailleurs une tribu à lui tout seul, comme son site pourrait le suggérer¹¹⁹. Dans son livre, p. 226, BC le mentionne d'une manière sympathiquement ambiguë. « Un délire potache, note-t-elle, fait que souvent les plus grands inventeurs en la matière (Roland Moreno) font circuler des blagues ou des jeux inappréciablement bêtes, si bêtes qu'ils en deviennent LewisCarrolliens. » Il est certain que R. Moreno s'intéresse aux mots et aux jeux avec les mots : l'une de ses inventions, le « Radoteur », est, selon l'article de Wikipédia qui lui est consacré, un logiciel « inspiré d'une idée de Claude Shannon » et « basé sur un algorithme de génération automatique de “mots” nouveaux à partir d'une ou plusieurs listes de mots issus du dictionnaire – une tentative pour automatiser la création de néologismes ou de noms de marques ».

Nous sommes ici au carrefour de plusieurs mondes : celui des techniques et des sciences, celui du langage et des langues, celui des arts, ceux aussi du droit et de l'industrie. Moreno est sans doute d'abord un praticien de génie plutôt qu'un véritable philosophe¹²⁰. Pour ce qui est des sciences, les « Remerciements » incluent le nom de *Françoise Balibar*, physicienne et historienne de la physique spécialiste d'Einstein, ainsi que celui du philosophe sénégalais Souleymane Bachir Diagne, dont la thèse, dirigée par Jean-Toussaint Desanti, était consacrée à l'algèbre de Boole¹²¹. Un membre éminent de la tribu que l'on doit mentionner à cet égard est *Maurice Matieu*, né en 1934. Matieu est d'abord un mathématicien de formation ; mais très tôt il se donne tout entier à la peinture, réalisant une première exposition dès 1965 à la galerie Aimé Maeght. Il est, si l'on peut dire, au cœur de la tribu ; entre lui et BC la relation est forte : BC écrit pour lui des textes, lui se fait pour elle illustrateur, ensemble ils conçoivent et construisent une œuvre¹²². Dans une riche série « d'ombres » figure en plusieurs versions l'ombre de Barbara Cassin, et celle aussi de plusieurs autres philosophes – Alain Badiou, Jacques Rancière, Philippe Sergeant, Juliette Simont, Pierre Verstraeten¹²³. Les échanges, les mises en relation foisonnent.

Du côté des arts et de la philosophie des arts, nous avons déjà croisé Christine Buci-Glucksman ; ajoutons ici *Jacqueline Lichtenstein*, qui enseigne l'esthétique dans le

département de philosophie de Paris X-Nanterre, et aussi *Éric Alliez*, premier nommé dans la liste alphabétique des remerciements. Philosophe (il enseigne à l'Université du Middlesex à Londres), il dirige la réédition des œuvres de Gabriel Tarde aux éditions « Les empêcheurs de penser en rond ». Spécialiste d'esthétique, il a édité (avec Elisabeth Samsonow) un ouvrage intitulé *Telenoia. Kritik der virtuellen Bilder* (Turia+Kant, Vienne, 1999). Marquons ici combien sont multiples les liens entre membres de la tribu : Alliez et BC sont ainsi tous deux membres du comité de rédaction de la revue *Multitudes*, dont Elisabeth Samsonow est par ailleurs membre du comité éditorial.

Le monde du droit, on l'a dit, n'est pas absent : dans la liste que nous suivons ici, il est représenté notamment par *Olivier Cayla*, agrégé de droit public, qui, dans le cadre du Centre des normes juridiques de l'EHSS, enseigne la théorie des normes, et cela « en croisant les grands textes de la philosophie politique moderne et les décisions les plus récentes de la jurisprudence constitutionnelle et administrative », précise le texte auquel nous nous référerons ici ¹²⁴. Il s'efforce, lit-on encore, « de mettre en lumière les conditions formelles d'édiction de la norme juridictionnelle et d'élucider les rapports nécessaires entre la pragmatique du langage et la fonction de souveraineté ». Surtout – du point de vue qui est le nôtre ici –, Olivier Cayla a co-édité avec BC et Philippe-Joseph Salazar un travail collectif (Cassin, Cayla & Salazar, 2004) intitulé « Vérité, réconciliation, réparation », numéro 43 de la revue *Le genre humain*, à propos des travaux de la commission « Vérité et réconciliation » créée en Afrique du Sud après la fin de l'apartheid.

Les questions de langage et de traduction sont, pour BC, la grande affaire. Plusieurs des noms figurant dans la liste des remerciements le rappellent discrètement. *Jean Dixsaut* est angliciste, et traducteur ; *Xavier North*, agrégé de lettres et délégué général à la langue française et aux langues de France du Ministère de la Culture, a participé avec BC à la journée internationale de la langue maternelle qui s'est tenue le 21 février 2007 à l'Unesco à Paris, autour du *Dictionnaire des Philosophies* dirigé par BC, ouvrage auquel a participé, parmi nombre de contributeurs, *Irène Rosier-Catach*, directrice de recherche au CNRS, spécialiste des théories du langage au Moyen Âge et coresponsable avec Alain de Libera de la section logique-langage du *Vocabulaire*. BC remercie encore *Heinz Wismann*, auteur (avec Pierre Judet de la Combe) d'un rapport alors récent intitulé *L'avenir des langues. Repenser les humanités* (Judet de la Combe & Wismann, 2004), mais avec qui BC a été très tôt en contact au cours même de sa formation d'hellénisante ¹²⁵.

Quelques autres noms seraient à mentionner. Celui de *Xavier Fels*, qui assume depuis plusieurs années d'importantes responsabilités chez PSA ¹²⁶, renvoie au monde du droit, de la haute administration, de la diplomatie et des affaires (Fels est membre du conseil exécutif du

MEDEF). En un tout autre point de « l'univers BC » sans doute, on trouve encore *Daniel Saadoun*, professeur de philosophie, qui siège, au Centre National de Livre (CNL) à côté de BC dans la commission « Philosophie ». Et puis il y a, « bien sûr », *Aalam Wassef*, qui dirige la collection *Banc Public* chez Albin Michel, dans laquelle il a publié le livre de BC après avoir, semble-t-il, encouragé BC à l'écrire. Autant que nous puissions le savoir, Aalam Wassef relève d'une configuration plusieurs fois rencontrée : appartenant au monde de l'édition, frotté au multimédia, il entretient avec le monde Informatique et Internet des rapports à la fois organiques et distanciés.

¹ Les situations auxquelles nous nous référons ici se distinguent ainsi des situations des professionnels de l'Internet ou des usagers professionnellement intéressés à la vie de l'Internet, qui procèdent à des recherches d'information pour participer à cette vie même et en influencer le cours.

² Ce passage se trouve à la page 249 de l'ouvrage. D'autres passages semblent témoigner d'une hostilité de première prise, que l'étude conduite par BC n'aurait pas sensiblement réduite : ainsi l'auteure parlera-t-elle encore, page 250, de « la solitude subjective et quasi onaniste du googleur de fond ».

³ Voir le site de cette association à l'adresse <http://www.deafct.org/board.htm>.

⁴ Voir http://www.khyf.org/our_sponsors.htm.

⁵ Voir <http://books.google.com/>.

⁶ Nous avons nous-même eu l'expérience d'un tel « collège invisible » dans notre apprentissage du référencement : la tribu en question, en ce cas, communique sur des forums de discussion et peut même se réunir physiquement, dans des rassemblements plus ou moins larges, qui se produisent quelques fois dans l'année.

⁷ On trouvera ce passage reproduit à l'adresse suivante : http://agora.qc.ca/francophonie.nsf/Documents/Internet-Les_carnets_dun_francophone_par_Jean-Marie_Borzeix.

⁸ On trouvera davantage d'indications – pour l'essentiel recueillies sur l'Internet, notamment grâce à... Google – dans l'annexe 2 à cette première partie de notre travail.

⁹ Plusieurs ouvrages de BC (dont sa thèse d'État) portent sur les sophistes, notamment *L'effet sophistique* (Gallimard, Paris, 1995). Voir Denis Slatka, « Éloge de la sophistique », *Le Monde*, 7 avril 1995.

¹⁰ On trouvera le texte de l'entretien contenu dans la plaquette de présentation du *Vocabulaire européen des philosophies* à l'adresse suivante : http://robert.bvdep.com/public/vep/Plaquette_VEP.pdf.

¹¹ Sur l'auteur du blog cité (Stéphane Bortzmeyer), voir <http://fr.wikipedia.org/wiki/Utilisateur:Bortzmeyer>.

¹² Cassin, 2007, pp. 24-25. Nous avons laissé la légère anomalie orthographique qui affecte dans ce passage le slogan *Don't be evil*.

¹³ Les différentes phases du travail rhétorique étaient classiquement au nombre de cinq : l'invention (*inventio*), la disposition (ou structuration : *dispositio*), l'élocution (qui préside à l'arrangement des mots dans le discours : *elocutio*), la mémorisation (*memoria*), la récitation proprement dite (*actio*).

¹⁴ Cassin, 2007, p. 83, note 2.

¹⁵ L'expression "mission accomplished" existe en anglais comme en français.

¹⁶ On notera en passant un fait que nous n'irons pas davantage interroger ici : dans la rhétorique des organisations « à l'anglo-saxonne », *mission* fait fréquemment couple avec *vision*.

¹⁷ Dans ce qui suit nous utilisons le dictionnaire en ligne *OneLook Dictionary*, qui permet de consulter un grand nombre d'ouvrages.

¹⁸ Voir <http://dictionary.cambridge.org/results.asp?dict=B&searchword=mission>.

¹⁹ Voir par exemple http://www.epo.org/about-us/office/mission_fr.html.

²⁰ Le "mission statement" prêté ici à Microsoft n'est pas celui que nous avons rencontré un peu plus haut ("Our mission is to enable people..."). On peut évidemment se demander s'il n'y a pas là un parti pris à l'encontre de la firme créée par Bill Gates ; question que nous n'examinerons pas davantage ici.

²¹ Voir http://www.franklincovey.com/fc/library_and_resources/mission_statement_builder.

²² Voir Cassin, 2007, pp. 82-88.

²³ Les éléments rassemblés dans ce qui suit ont été acquis le 6 septembre 2007.

²⁴ Lorsqu'on mentionne comme ici l'opérateur de conjonction, le moteur de recherche Google affiche sentencieusement ceci : « L'opérateur "AND" n'est pas nécessaire : par défaut, Google tient compte de tous les termes de recherche spécifié. » Observons tout de même que si l'explicitation de l'opérateur AND n'est pas nécessaire *pour Google*, elle peut être utile *pour l'utilisateur* lorsqu'il rédige sa requête ou la relit pour la contrôler (par exemple).

²⁵ Voir : <http://sbinformation.about.com/cs/businessplans/a/mission.htm>.

²⁶ Voir <http://simple-entrepreneur.blogspot.com/2007/09/writing-meaningful-mission-statement.html>.

²⁷ Voir <http://Taylorandfrancis.metapress.com/index/G65324Q3P7722264.pdf>.

²⁸ Voir <http://www.sound-business-practices.com/rationale/mission-statements.php>. Le mot *combo* utilisé dans ce texte est une abréviation de *combination* : appliqué d'abord à un petit groupe de jazz (ou d'autres musiques), il tend à désigner familièrement, en anglais américain, toute « combinaison » – par exemple "*a burger, fries, and shake combo*".

²⁹ Voir <http://www.lavoisier.fr/notice/frCFO32RSUOOVSAR.html>.

³⁰ Voir Lashinsky, 2007, qui nous apprend en outre que le "mission statement" a été récemment modifié. (À propos du vocable *gizmo* employé par l'auteur, on peut se reporter à l'article de même nom de l'encyclopédie Wikipedia.)

³¹ Voir Cassin, 2005, p. 88 et suiv.

³² *Humain, trop humain II, Le voyageur et son ombre*, § 55. Voir Nietzsche, 2000, p. 355.

³³ Le nom de Jean-Noël Jeanneney n'a que trois occurrences dans le livre de BC. On aura noté qu'il n'apparaît pas, en outre, dans la liste des personnes remerciées.

³⁴ Dans cette citation et dans les suivantes, c'est nous qui soulignons.

³⁵ Sa mission est formulée ainsi : "The mission of GlobalWarmingSolution.Org is to bring about a timely and effective solution to global warming. We advocate that global greenhouse gas emissions be reduced 80% below 1990 levels by 2025."

³⁶ Voir <http://www.globalwarmingsolution.org/about%20us/info.html>.

³⁷ Voir <http://www.secularstudents.org/mission>.

³⁸ Voir <http://www.palmbeachpoetryfestival.org/about>.

³⁹ Les effectifs indiqués dans le tableau ont été recueillis dans la matinée du 1^{er} octobre 2007 : on les considère ici comme des indicateurs grossiers mais réalistes de la fréquence d'emploi, dans le contexte étudié, des 47 verbes retenus.

⁴⁰ On ne s'étonnera pas que des persécutés aient adopté sur nombre de points la langue de leurs bourreaux. Avec la lucidité dont il est coutumier, Klemperer note (1947/1996, p. 120) que « la LTI était une langue carcérale (celle des surveillants et celle des détenus) », et qu'une telle langue « comporte inéluctablement (en manière de légitime défense) des mots secrets, des ambiguïtés fallacieuses, des falsifications, etc. », dont les persécutés usent, dans toute la mesure du possible, pour lutter contre la persécution dont ils sont les victimes.

⁴¹ Le *Vocabulaire* dirigé par BC ne contient pas d'entrée *Organique*, ou *Organiser*, ou *Organisation*, etc. Or ces mots, comme les analyses de Klemperer le suggèrent, constituent un point de difficulté du vocabulaire européen. Cela n'a pas échappé à Raymond Williams qui, dans son célèbre *Keywords. A vocabulary of culture and society* (1983), consacre à *Organic*, "an especially difficult word", trois pages denses.

⁴² Voir <http://dictionnaire.reverso.net/anglais-francais/organize>.

⁴³ Voir <http://www.chambersharrap.co.uk/chambers/features/chref/chref.py/main?title=21st&query=organize>.

⁴⁴ Cela ne signifie nullement, bien entendu, que les documents qu'on ne trouve pas sur le Web puissent être ignorés. À propos de l'emploi du verbe *organiser* discuté ci-dessus, nous avons ainsi trouvé cet extrait d'un manuel des classes primaires des premières décennies du XX^e siècle (cité par Lambrichs, 1993, p. 83) : « Pasteur avait organisé dans le vieux collège Rollin une chambre pour la mère et l'enfant. » (L'enfant est le petit Joseph Meister, le premier à avoir été soigné de la rage selon le protocole pasteurien.)

⁴⁵ Cette lettre figure dans le document déjà cité. C'est nous qui soulignons.

⁴⁶ Le *Corporate Overview* de Google (<http://www.google.com/corporate/>) s'ouvre par ces mots : "Google's mission is to organize the world's information and make it universally accessible and useful." Dans la version française correspondante, on lit ceci : « Google a pour mission d'organiser à l'échelle mondiale les informations dans le but de les rendre accessibles et utiles à tous. »

⁴⁷ L’auteur de l’article le présente très exactement comme “chief engineer of Google’s system for scanning books in the library collections”.

⁴⁸ À l’adresse http://fr.wikipedia.org/wiki/Liste_des_moteurs_de_recherche. On peut se demander s’il ne s’agirait pas là d’une description influencée par l’ouvrage de BC. L’article en question a depuis été supprimé (http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Pages_%C3%A0_supprimer/Liste_des_moteurs_de_recherche). Toutefois, au 31 août 2008, on pouvait le retrouver à l’adresse suivante : http://dictionnaire-referencement.wranks.com/index.php?title=Liste_des_moteurs_de_recherche.

⁴⁹ En fait, l’expression recherchée n’y figure pas, mais elle est mentionnée par Google comme apparaissant dans des pages ayant un lien vers cette page.

⁵⁰ Selon l’auteur la mission de Google serait “to collect ‘*all the world’s information*’ and make it accessible to everyone”, ce qui, encore une fois, revient à admettre la quasi-équivalence entre “*all the world’s information*” et “*the world’s information*”.

⁵¹ C’est Larry Page qui parle.

⁵² Voir <http://www.google.com/press/pressrel/pressrelease1.html>.

⁵³ Voir <http://brainofphil.blogspot.com/2006/04/google-search-omniscience.html>.

⁵⁴ Barbara Cassin utilise l’édition du livre de John Batelle parue en septembre 2005. Nous nous référons ici à l’édition parue en 2006.

⁵⁵ Barbara Cassin utilise l’édition du livre de Vise et Malseed parue en novembre 2005. Nous nous référons ici à l’édition parue en septembre 2006.

⁵⁶ On peut imaginer aussi – mais c’est là bien sûr pure conjecture – que la formule “*all the information in the world*” a pu être préférée sciemment, pour des raisons linguistiques, la formule utilisant le génitif saxon – “(*all the world’s information*)” – étant sans doute jugée plus opaque au lecteur francophone que la formule utilisée par BC. Notons cependant que BC n’est pas innocente de petites altérations vénielles : sur les 21 mentions du nom de Sergey Brin que contient son livre, BC écrit ainsi 13 fois – fautivement – « Bryn » ; de même écrit-elle « Turing » le patronyme d’Alan Mathison Turing (1917-1954).

⁵⁷ Concluant le chapitre 17, intitulé « Système et organisation », et se référant alors à la description qu’il vient de donner d’un tic de langage datant d’une époque qu’on voudrait croire révolue, Klemperer se moque gentiment de lui-même en ces termes (p. 145) : « Cela fait déjà un bon moment que j’écris : c’était... c’était. Mais qui a dit hier encore : “Il faut que je ‘m’organise’ un peu de tabac” ? Je crains que ce ne soit moi-même. »

⁵⁸ D’une façon plus générale, à propos d’une étude souhaitée suffisamment complète de la *LTI*, Klemperer écrit ceci (1947/1996, p. 38) : « ... un énorme travail devra être fourni dans des domaines extrêmement variés : germanistes et romanistes, anglicistes et slavistes, historiens et économistes, juristes et théologiens, techniciens et biologistes devront consacrer des essais et des thèses à de très nombreux problèmes particuliers avant qu’un

esprit ample et courageux puisse oser décrire la *Lingua Tertii Imperii* dans sa globalité la plus pauvre et la plus riche. »

⁵⁹ Plutôt que de dire que l'algorithme de PageRank *permet* de classer les réponses, sans doute vaudrait-il mieux dire qu'il *contribue* à les classer.

⁶⁰ Cassin, 2007, pp. 62-63. L'acronyme IR dans "IR score" signifie "information retrieval" (ce que BC omet de signaler à son lecteur).

⁶¹ Voir "*Paging Usher*", in Battelle, 2006, p. 237 et suiv.

⁶² Cassin, 2007, p. 140. (C'est BC qui souligne.) Le texte cité a depuis été légèrement modifié par Google : voir http://www.google.fr/why_use.html.

⁶³ BC mentionne (p. 148) les désagréments survenus à BMW Allemagne pour avoir tenté de modifier le classement au profit du site de cette firme.

⁶⁴ De la *requête* de l'utilisateur et non *de l'utilisateur* lui-même – ce qui serait le cas si le moteur de recherche gardait la mémoire des requêtes précédentes de l'utilisateur en question et en tenait compte d'une certaine manière dans les réponses qu'il apporte à ses requêtes.

⁶⁵ Cassin, 2007, pp. 94-102.

⁶⁶ Le vocable « PageRank » a 50 occurrences dans son livre (à quoi s'ajoute une occurrence de « Page Rank »). Il est vrai que « Google », quant à lui, y a plus de sept cents occurrences !

⁶⁷ Cassin, 2007, p. 60. Nous allons revenir sur ce fait.

⁶⁸ Voir <http://dbpubs.stanford.edu:8091/diglib/pub/slides/berkeleydlijan98/berkeleygoogle2/sld014.htm>.

⁶⁹ Voir <http://www.google.com/corporate/history.html>.

⁷⁰ Voir <http://www.journaldunet.com/diaporama/0603google-story/4.shtml>.

⁷¹ Sur les mathématiques de l'algorithme PageRank, voir Chevallard, 2008.

⁷² Voir <http://infolab.stanford.edu/~backrub/google.html>. BC fait référence à cet article dans une note à la page 42 de son livre ; elle traduit et commente son titre à la page 54. « Le titre est assez dada : *The Anatomy of a Large-Scale Hypertextual Web Search Engine* [Anatomie d'un moteur de recherche hypertexte à grande échelle pour la toile] ».

⁷³ On peut imaginer une page Web proposant un très grand nombre de liens constituant une *webographie* sur un thème donné. Alors même que le fait d'être citée par elle n'accroîtra que peu le PageRank d'une page, le contenu de cette webographie peut être des plus précieux.

⁷⁴ Il s'agit en fait du *Citation Index*. On a vu que BC était familière de ce genre d'imprécision. À nouveau, une courte recherche sur Internet aurait permis de choisir la bonne expression : même si la distinction devient parfois impalpable, une *citation* est une référence, "*a short note recognizing a source of information or of a quoted passage*", tandis qu'une *quotation* est le "*quoted passage*" lui-même (d'après le *OneLook Dictionary*). Sur le

classement de Shanghai des universités (*Academic Ranking of World Universities*), on pourra se reporter à l'adresse http://fr.wikipedia.org/wiki/Classement_acad%C3%A9mique_des_universit%C3%A9s_mondiales.

⁷⁵ Cassin, 2007, pp. 102-103. Il s'agit là à nouveau d'un passage du texte *Pourquoi Google ?* que l'on trouve à l'adresse http://www.google.fr/intl/fr/why_use.html.

⁷⁶ Cet adjectif n'est jamais utilisé par BC. Mais il nous semble qu'il exprime bien l'essence de l'objet d'une réprobation que cette auteure manifeste en mille façons.

⁷⁷ Cassin, 2007, pp. 102-103. L'affirmation qu'un système serait opaque parce qu'il est gouverné par une formule *mathématique* paraît très discutable : si le classement des résultats réalisé par Google relevait *seulement* de la formule mathématique reproduite plus haut, son *principe* serait au contraire de part en part transparent. D'un certain point de vue, les mathématiques sont *le contraire* de ce que Brin et Page eux-mêmes, dans leur *Anatomy of a Large-Scale Hypertextual Web Search Engine*, désignent comme étant du *black art*.

⁷⁸ Rendue célèbre par George Orwell dans son *Animal Farm* ("All animals are equal, but some animals are more equal than others"), cette formule est devenue courante en français (et en anglais).

⁷⁹ Cassin, 2007, p. 125. C'est nous qui soulignons.

⁸⁰ Batelle 2006, p. 258. Vannevar Bush (1890-1974) publie en juillet 1945, dans *The Atlantic Monthly*, un article qui fera date, "As We May Think". Il y décrit un dispositif qu'il nomme *memex* (contraction de "memory extender"), qui semble avoir inspiré ou, du moins, préfiguré l'idée d'hypertexte (voir l'article "Memex" de l'encyclopédie Wikipedia).

⁸¹ Le lancement de Gmail est commenté par Vise et Malseed, qui y consacrent le chapitre 14, intitulé *April fools*, de leur livre (pp. 152-163).

⁸² Cassin, 2007, p. 24, note 4.

⁸³ BC utilise l'édition 1995 du livre de C. Huitema (1996).

⁸⁴ Tout en précisant que le mot est utilisé de façon ironique, l'ouvrage *The Internet for Dummies* cité ci-après propose cette définition bienveillante des *dummies*, que l'on retiendra seule ici : "People who don't know everything but are smart enough to seek help". Dans la langue anglaise courante, le mot est sensiblement plus venimeux, comme le suggère par exemple le *Compact Oxford English Dictionary*, qui précise (pour ce qui est du sens du mot visé ici) : "informal, chiefly N. Amer. a stupid person."

⁸⁵ Cité par Michael Hauben (s. d.). C'est nous qui soulignons.

⁸⁶ Voir <http://www.youtube.com/watch?v=LKTH6f1JfX8>.

⁸⁷ Selon l'article "Google China" de Wikipedia, la formule se traduit en anglais comme ceci : "In accordance with local laws, regulations and policies, part of the search result is not shown."

⁸⁸ D'après Kevin J. Delaney (2006), qui présente MacLaughlin comme un "Google Senior Policy Counsel".

⁸⁹ Sur l'erreur 404, voir l'article "HTTP 404" de Wikipedia.

⁹⁰ BC semble s'inscrire ici dans une tradition d'érudition polémique qu'illustra notamment Pierre Larousse, l'auteur du *Grand dictionnaire universel XIX^e siècle*, dont la notice consacrée à Bonaparte présente dans les termes suivants le futur Napoléon I^{er} : « général de la République française, né à Ajaccio (île de la Corse), le 15 août 1769, mort au château de Saint-Cloud près de Paris, le 18 brumaire, an VIII de la République française, une et indivisible (9 novembre 1799). »

⁹¹ BC emprunte pour cela des informations à un article du quotidien *Le Monde* (Piquard, 2006).

⁹² L'idée, semble-t-il, était d'obtenir la création d'une extension .xxx. Cette proposition fut à l'époque rejetée par l'ICANN.

⁹³ Voir Foucart, 2006. (Cet article a été reproduit dans *Le Monde* 2 daté du 17 novembre 2007, pp. 74-75.)

⁹⁴ Voir <http://www.archive.org/index.php>.

⁹⁵ La référence est ici au film éponyme (1990) dont une « accroche » est la phrase suivante : “In this world, they can steal your mind, erase your memory, and give you another identity. But the most dangerous thing that can happen to you is...TOTAL RECALL.”

⁹⁶ D'après l'article “Internet Archive” de l'encyclopédie Wikipedia, cette organisation à but non lucratif se présente en ces termes : “The Archive’s mission is to help preserve those artifacts and create an Internet library for researchers, historians, and scholars”.

⁹⁷ Le terme de *mêtis* est, précise BC, le vocable grec « qui caractérise la ruse d’Ulysse et l’intelligence prométhéenne des sophistes ». Sur la notion de *mêtis*, voir Vernant et Detienne, 1974. On peut se demander si la notion de *mêtis* n’est pas moins pertinente ici, pour comprendre le capitalisme américain, que la notion de *business model* : Alexa permet et d'alimenter l'Internet Archive et de gagner l'argent sans lequel ce projet philanthropique ne pourrait exister dans le monde capitaliste.

⁹⁸ BC reproduit la petite inexactitude qui apparaît dans la source qu'elle utilise, à savoir l'interview des Google guys paru dans le numéro de *Playboy* de septembre 2004, où “will” est remplacé par “could”. John Ashcroft a été le 79^e *United States Attorney General* (à la tête du *Department of Justice*) dans le gouvernement de George W. Bush à partir de décembre 2000 et jusqu'au 10 novembre 2004 ; il est connu pour ses positions conservatrices, voire traditionalistes, sur nombre de sujets de société notamment.

⁹⁹ BC a signé dans le *Vocabulaire européen des philosophies* (qu'elle a dirigé) l'essentiel de la notice sur la notion de *doxa*, dont l'encadré D reprend nombre de formulations.

¹⁰⁰ La requête “Google.org” fait apparaître comme premier résultat un lien *Google.org* (dont l'adresse est <http://www.google.org/>). En cliquant sur ce lien, on obtient un texte en anglais dont le début a été reproduit ci-après ; en cliquant sur la mention « Traduire cette page » figurant à côté du lien, on obtient alors le texte en « français » ci-dessous (réalisé le 28 novembre 2007). De tels jeux sont classiques : voir ainsi Eco, 2006, chapitre 1.

¹⁰¹ Lorsque BC fait parler le traducteur automatique de Google, cette firme utilise un système déjà performant produit par la société Systran ; depuis lors, Google a mis en service un traducteur de son propre cru (Google, 2008).

¹⁰² Avec une adresse très légèrement différente : <http://www.google.com/accounts/TOS?loc=FR> au lieu de <http://www.google.com/accounts/TOS?loc=fr&hl=fr>.

¹⁰³ Il contient un article d'Astrid Girardeau daté du 27 novembre 2007, intitulé « Critiqué, Google modifie ses conditions d'utilisation ».

¹⁰⁴ Voir <http://info.yahoo.com/legal/us/yahoo/utos/>.

¹⁰⁵ Voir <http://www.usatoday.com/marketing/tos.htm>.

¹⁰⁶ Voir <http://secondlife.com/corporate/tos.php>.

¹⁰⁷ Voir <http://help.live.com/help.aspx?project=tou&mkt=fr-fr>.

¹⁰⁸ Voir <http://www.microsoft.com/info/copyright.msp#E1>.

¹⁰⁹ Voir <http://www.systran.fr/>.

¹¹⁰ Voir <http://www.systran.fr/traduction/systran/entreprise/politiques/mentions-legales>.

¹¹¹ Voir <http://www.emerce.nl/nieuws.jsp?id=812356>.

¹¹² AltaVista (<http://fr.altavista.com/>) est un moteur de recherche qui offre également un service de traduction. Voir <http://babelfish.altavista.com/> et <http://translation.langenberg.com/> où la traduction du néerlandais vers le français est disponible.

¹¹³ Voir Montoliu, 1998. De la phrase en espagnol qui suit, Google propose cette traduction : “For clone understand that the two issues are absolutely identical, it is exactly the same newspaper and that the Catalan version is fully translated every day from the Spanish.”

¹¹⁴ Voir Cassin, 2007, pp. 252-254. Dans les précisions qu'elle donne, BC commet deux petites erreurs. Elle indique que J.-L. Moullet « a présenté Galileo et Quaero à l'AFP en janvier dernier ». Ici, « janvier dernier » signifie « janvier 2006 » (et non janvier 2007, année de la parution de son livre). Par ailleurs, le texte qu'elle reproduit est une déclaration faite à l'AFP en décembre 2005 pour annoncer une présentation programmée, elle, en janvier 2006 (dans le cadre l'Agence de l'Innovation Industrielle, alors chargée du projet Quaero). Le texte qu'elle reproduit est cité dès les derniers jours du mois de décembre 2005 (voir ainsi <http://www.silicon.fr/fr/silicon/news/2005/12/27/quaero-futur-google-europeen>) et apparaît en ligne dans la même période (voir par exemple <http://www.clubic.com/actualite-30463-hi-tech-l-offensive-europeene-galileo-et-quaero.html>).

¹¹⁵ ITR : Informatique, Télécoms, Réseaux.

¹¹⁶ L'ITEM est une Unité Mixte de Recherche CNRS / École Normale Supérieure qui se consacre à l'étude des manuscrits d'écrivains pour élucider les processus de la genèse. Cet institut est « né de la rencontre d'un besoin ponctuel – le recensement et le classement des manuscrits de Heine, acquis par la Bibliothèque Nationale – et

d'une conjoncture théorique – la nécessité de penser la diachronie de l'écriture que le structuralisme, alors triomphant, avait négligé » (<http://www.item.ens.fr/index.php?id=106>).

¹¹⁷ La dimension politique du projet de BNUE est clairement manifestée dans l'entame de l'article de V. Tesnière et N. Lesquins : « Face aux défis lancés par les moteurs de recherche tels que Google, relayés par la presse sensibilisée à la suite du débat public ouvert par Jean-Noël Jeanneney [*Quand Google défie l'Europe : plaidoyer pour un sursaut*, Paris, Mille et une nuits, 2005], le président Jacques Chirac lance l'idée d'une bibliothèque numérique européenne. Le ministre de la Culture demande dès le 30 août 2005 d'établir un état des lieux précis et les contours d'une contribution française à un projet de bibliothèque numérique européenne (BNUE). Cinq groupes de travail réunissent des acteurs du privé (représentants du monde de l'édition, des secteurs industriels de la numérisation et des moteurs de recherche) et du secteur public (Culture, Éducation nationale, Recherche, Industrie, Finances). Renaud Donnedieu de Vabres. »

¹¹⁸ Voir le site <http://www.mpi.nl/DOBES/>.

¹¹⁹ Voir <http://www.rolandmoreno.com>.

¹²⁰ Même s'il a publié en 1995 un ouvrage qui a connu un réel succès de curiosité : *Théorie du bordel ambiant* (Éditions de l'Archipel, Paris).

¹²¹ Diagne est l'auteur d'un ouvrage paru chez Belin en 1989 sous le titre *Boole. L'oiseau de nuit en plein jour*. Il a également traduit en français les *Laws of thought* (Vrin, 1992).

¹²² BC et Maurice Mathieu ont conçu ensemble l'ouvrage intitulé *Voir Hélène en toute femme. D'Homère à Lacan* (Cassin & Mathieu, 2000). Voir http://www.plumart.com/vf3301/html/body_1133matieu.html et aussi http://www.simhaprod.com/vid/p_cassin.html.

¹²³ Voir <http://perso.orange.fr/maurice.matieu/Animalsans.html>.

¹²⁴ Voir <http://lodel.ehess.fr/cenj/document.php?id=60>.

¹²⁵ Voir Cassin avec Briffard, 2006.

¹²⁶ Plus exactement PSA Peugeot Citroën, résultat de l'absorption (en 1975) de Citroën SA par Peugeot SA (PSA).

DEUXIÈME PARTIE

QUE PEUT-ON SAVOIR ?
EXPLORATIONS PRAXÉOLOGIQUES

Introduction

Vrac, le mot me plaît. Quelle sorte de mot, au juste ?

« Locution adverbiale » me souffle un Petit Robert.

Où ? À la page 1928 de l'édition de 1967.

Il se trouve là en belle compagnie : exactement entre « voyou » et « vrai ».

François Nourissier, 2008

Rappelons la question formulée à l'issue de la première partie de ce travail : *vers quelle culture praxéologique faudrait-il que nous tendions tous ensemble pour penser et pratiquer la RII à notre avantage – sans être victime des fables et autres inventions qui, à tout moment, peuvent nous tomber dessus ?* Dans toute la suite, nous nous attacherons plus précisément à une question que l'étude de la question précédente pousse en avant, si elle ne s'y ramène pas entièrement : *que pouvons-nous savoir* qui nous permette de penser et de pratiquer la RII « à notre avantage » ?

Toutefois, cette question n'est pas précisément définie : à cause de la polysémie du verbe *pouvoir*, se dissimule derrière le « *que pouvons-nous savoir* » une pluralité d'interrogations. Une clarification, une « désambiguation » (pour parler comme l'encyclopédie Wikipédia) s'imposent, sous peine de confondre des objets proches, certes, mais distincts. Si l'on s'interroge à propos d'une personne ou d'une institution en se demandant « Que peut-elle savoir en fait de RII ? », on peut vouloir dire « Qu'est-elle *capable* de savoir – et donc d'apprendre – en fait de RII ? », en imaginant notamment que certaines connaissances en la matière – par exemple touchant les mathématiques de l'algorithme PageRank – pourraient bien se situer (actuellement) *au-delà* de sa capacité (actuelle) d'apprendre (en un temps limité), et donc de savoir. C'est là une problématique commune, mais que, en vérité, *nous écarterons de ce travail*, en supposant que personnes et institutions font preuve d'une intelligence « intrinsèque » *toujours suffisante pour apprendre* (en un temps suffisant), quel que soit le contenu à apprendre.

On se rapprochera de ce qui est au cœur de notre recherche avec un deuxième sens de *pouvoir*, qui donne à la question « Que peut-elle savoir en fait de RII ? » la valeur de « Que se

peut-il qu'elle sache en fait de RII ? ». Mais attention : ce qui nous importe alors, ce n'est pas davantage de connaître ce que telle personne ou telle institution déterminée *sait*, de fait, en matière de RII. Ce que nous visons, c'est la connaissance de ce qu'il est *possible* – au sens de *probable* – qu'elle sache en la matière ; non pas, donc, ce qu'elle sait, mais ce qu'elle *pourrait bien savoir*. Considérons ici l'ensemble Π des praxéologies π relatives à la RII ¹. La question que nous venons d'évoquer équivaut alors à se demander quelles sont les praxéologies $\pi \in \Pi$ qu'il est très probable, ou au contraire peu probable, voire très improbable, etc., qu'une personne ou une institution, appartenant à telle catégorie C de personnes ou d'institutions, *connaisse*.

Il nous importe surtout de souligner ici que la réponse à la question précédente (pour une catégorie donnée C de personnes ou d'institutions) apparaît liée à un autre sens encore du verbe *pouvoir* : celui d'*autorisation accordée*, de permission délivrée à une personne ou à une institution. « Que peut-elle savoir en fait de RII ? » signifie en ce sens « Qu'est-elle ou qu'a-t-elle été autorisée à savoir en fait de RII ? », par exemple par une institution « scolaire » qui aurait tenté ainsi d'abolir un décret peut-être jamais formellement énoncé mais partout tacitement admis et reconnu (pour paraphraser Rousseau parlant du contrat social) selon lequel cette personne ou cette institution « n'aurait pas à connaître » tel ou tel savoir. Une telle autorisation transforme alors la personne ou l'institution en *étudiant* potentiel légitime, mais ne dit rien encore sur le succès de l'étude à laquelle celle-ci se livrera peut-être, c'est-à-dire sur sa qualité d'*apprenant* (si l'on entend ce mot *stricto sensu*), soit encore sur la réalité des apprentissages ainsi autorisés et sur la nature du savoir effectivement acquis *in fine* – et cela par exemple parce que d'autres autorisations, venant d'autres points de l'univers des institutions de la société, lui auront été refusées (ce qui est à relier sans doute aussi à la volatilité de nombre d'apprentissages scolaires).

La famille des conditions et contraintes d'autorisation ou de proscription est nombreuse ; mais elle comporte aujourd'hui une contrainte extrêmement prégnante et relativement uniforme (elle pèse *a priori* semblablement sur beaucoup de personnes et d'institutions), qui est susceptible de restreindre fortement l'accès au système praxéologique Π en même temps que d'en infléchir la dynamique de développement en en favorisant la stagnation ou même la dégénérescence : il s'agit de la *péjoration* répétée dont est victime aujourd'hui encore le « monde Internet », fait massif dont la première partie de ce travail fournit un exemple éclatant. Un épisode récent en rappellera ici l'essentiel. Dans une chronique hebdomadaire intitulée « Moderneries », un collaborateur du quotidien *Libération*, Pierre Marcelle, publie le samedi 31 mai 2008 un billet au titre à demi sibyllin, écrit dans un français « modernisé » : *Google limites* (Marcelle, 2008a). Voici l'affaire. Quatre

personnes, dont l'auteur du billet, ont quelque temps auparavant tenté d'identifier une peinture reproduite sur l'affiche d'une exposition tenue à Paris en 1992-1994. Ces personnes croient y voir « une enfant aux cheveux rouges » et un « corbeau géant » ; l'une d'elles suggère qu'il pourrait s'agir d'une œuvre du peintre allemand Emil Nolde (1867-1956). Acceptant cette conjecture, le groupe entreprend de rechercher sur Internet l'identité exacte du tableau. Il s'avouera vaincu, après des essais multiples sur lesquels nous passerons ici. « Au terme épuisé de cette tentative d'identification », comme l'écrit emphatiquement l'auteur du billet, ce dernier croit pouvoir tirer des conclusions qui, peut-on penser, sont en grande partie pour lui de pures confirmations de ses préjugés : 1) « le moteur de recherche, qu'il s'appelle Google, Voila ou Yahoo, n'a pas réponse à tout » ; 2) « sa capacité à trouver n'est pas proportionnelle au nombre de critères qu'on lui suggère » ; 3) bien qu'ayant « craché plusieurs centaines de milliers d'oracles » (*sic*), la « machine » semble bien ne rien savoir du tableau cherché ; 4) la recherche n'aboutira *que* parce qu'un « comparse » voyageur, heureux possesseur d'un catalogue du « Brücke Museum de Copenhague », leur permet d'identifier – à l'ancienne, donc – le tableau de Nolde. La première des conclusions précédentes est, certes, un truisme ; mais cela n'empêche pourtant pas l'œuvre évoquée d'être d'une identification *facile* sur Internet : à la date indiquée, il suffisait de présenter à Google, *par exemple*, la requête **Nolde images** ; passé le tout premier résultat affiché (renvoyant à la moisson opérée par Google Images), on tombait sur un lien conduisant à une biographie du peintre avec, en bas de page, des liens vers un petit nombre de ses peintures, dont l'une intitulée... *Child and Large Bird*². La deuxième des conclusions ci-dessus – où il est question de proportionnalité – est parodique et bouffonne ; la troisième, on l'a vu, est donc platement fausse. La quatrième est la clé de tout : elle suggère que ce que Google fournit *en vrac* a généralement trait à *n'importe quoi*, à l'exception de ce que l'on recherche.

Dans son billet du samedi précédent intitulé *Google attitude*, le même auteur écrivait sans détour : « La seule nécessité ou de simples curiosités sollicitent cent fois par jour la pêche au clic dans la mare aux milliasses de possibles, en même temps qu'elles la rendent inéluctablement absurde. » On saisit là de façon presque caricaturale la corrélation (qui témoigne peut-être d'un véritable lien causal) entre péjoration de Google (et du monde Internet) et indigence praxéologique en matière de RII. On comprendra alors que, dans le premier chapitre de cette deuxième partie, nous ayons tenté de lever, *dans le cadre de notre travail*, la ruineuse hypothèque pesant sur le monde Internet, et cela en essayant de montrer que ce qui est présenté comme une évidence par certains aujourd'hui peut être fondamentalement discuté, et en évoquant la possibilité que la vérité de demain ne soit peut-être plus celle qui va de soi pour nombre d'esprits « critiques » d'aujourd'hui. Si, par

métaphore, on identifie l'univers culturel dans lequel s'enferment ceux-là à un univers euclidien, nous appelons ainsi le lecteur à une expérience mentale consistant à s'imaginer plongé dans un monde *non euclidien*, où l'ordre classique s'efface au profit d'ordres inédits, auxquels on est sûr de ne rien comprendre en les nommant, sans autre forme de procès, « désordres ».

Cet obstacle surmonté – admettons-le au moins –, il s'agit alors de se tourner vers l'ensemble Π pour découvrir quels éléments praxéologiques s'y rencontrent, selon quelle cartographie, organisée selon quelles frontières « naturelles » ou « artificielles ». Ainsi que nous le verrons plus tard, ce n'est qu'alors que le système des autorisations et des interdits en matière de RII pourra être pris en compte réellement. L'équipement praxéologique *probablement* accessible à telle catégorie de personnes et d'institutions dépend de ce qu'est l'équipement praxéologique *légitimement* accessible à cette catégorie, notion qui ne prend son sens que si l'on précise ce qu'est l'équipement praxéologique qui lui est en quelque sorte « *matériellement* » accessible – qui est, comme on dit en anglais, *out there*. Mais qu'entend-on par là ? Comment examiner l'ensemble Π des praxéologies π relatives – directement ou indirectement – à la RII ?

L'ensemble Π peut être défini en extension comme constitué, à un moment donné, par la réunion des équipements praxéologiques en matière de RII de toutes les personnes et de toutes les institutions d'un ensemble donné de sociétés. Bien entendu, une telle définition n'est guère opérationnelle : s'il est en effet possible de montrer que telle praxéologie π appartient à Π (parce qu'on peut exhiber telle personne ou telle institution dont l'équipement praxéologique la contient), il semble impossible en revanche d'énumérer exhaustivement l'ensemble Π . En fait, nous réduirons un peu le problème en redéfinissant Π pour en faire un ensemble plus « tangible ». Pour cela, il nous faut introduire alors l'ensemble Σ des *exposés* relatifs à la RII, et même des exposés *écrits* (sous forme imprimée ou électronique). Dans toute la suite de ce travail, nous nommerons *exposé* (sur la RII) toute *totalité organisée, en général ouverte, de déclarations écrites* (à propos de la RII) ³. L'ensemble Σ ainsi caractérisé, on doit le souligner, n'est pas défini par une institution établie qui imposerait au chercheur sa propre nomenclature comme allant de soi (ainsi qu'il en va par exemple lorsque, cette institution étant, disons, le collège, le chercheur y considère l'ensemble des manuels de technologie ou de mathématiques de 4^e disponibles en une période donnée, et elle seulement). En conséquence, et hormis si l'on se réfère à un sous-ensemble spécifiquement défini de Σ , les exposés composant Σ apparaîtront en général *non homogènes* du point de vue des critères culturels usuels : ainsi un exposé sur la RII, au sens où nous l'entendons ici, pourra-t-il être aussi bien un texte de quelques pages qu'un traité de plusieurs centaines de pages, par

exemple ⁴. Selon la logique de la recherche scientifique, le critère définitoire retenu est ici *interne à la recherche* et, en l'espèce, est de nature *didactique* : appartient à notre ensemble Σ de référence *tout texte prétendant donner à connaître quelque chose de la RII*. Un exposé, donc, pourra consister aussi bien en tout ou partie d'un *dossier* (publié dans une revue grand public), d'un *mémoire* (de master, de thèse, etc.), d'un *article* (de magazine, de revue scientifique, etc.), d'un *traité*, d'un *recueil*, etc. L'ensemble Σ des exposés sur la RII ainsi défini inclura en conséquence aussi bien, en totalité ou en partie, des traités savants ou des ouvrages demi-savants ⁵ – qui, en règle générale, seront pour nous d'abord des *outils d'étude*, tout en étant aussi des *objets* d'étude éventuels – que, à l'autre extrémité, de simples notules figurant dans des documents plus ou moins étendus – qui seront d'abord des *objets d'étude*, sans que soit exclue leur emploi contrôlé comme *outils* d'étude.

Au lieu de l'ensemble Π initialement évoqué, nous nous référerons donc désormais à l'ensemble, que nous notons encore Π , des éléments praxéologiques relatifs à la RII présentés ou évoqués dans les exposés composant Σ . Nous faisons l'hypothèse que la réduction ainsi opérée, qui assimile à ses seules « mises en texte » (de toute nature) une réalité praxéologique vivant dans *l'activité* d'institutions et de personnes, se fait sans perte essentielle, parce que, à l'âge du traitement de texte et du Web, avec ses forums de discussion et ses blogs, même les praxéologies apparemment les plus intimes (du point de vue d'une personne ou d'une institution) tendent à se mettre par écrit – il y a ainsi un déplacement sensible de la culture de l'oral, porteur de la parole vive, vers celle de l'écrit, qui fixe le propos, le pérennise et le propage. C'est pour la même raison que, ainsi qu'on l'aura noté, alors qu'un exposé peut certes être délivré oralement, sans même le support d'un texte écrit préalable, on n'a retenu ici, par méthode, que les exposés *écrits* (en y incluant, le cas échéant, les textes mettant en forme écrite des exposés oraux). Cela étant précisé, il reste à explorer, avec l'ensemble Σ des exposés, l'ensemble Π des praxéologies relatives à la RII qui s'y trouvent « exposées » : c'est ce que nous commencerons à faire dans les trois derniers chapitres de cette deuxième partie.

CHAPITRE 6

Ordre et désordre dans la connaissance

La question de la recherche d'information sur Internet dans le cadre d'une *enquête codisciplinaire* à propos d'une question Q (enquête concrétisée en un *parcours d'étude et de recherche*), c'est-à-dire la question des *techniques* de recherche d'informations, des *technologies* associées et des éléments *théoriques* qui ne peuvent manquer de faire sentir leur influence dans un sens ou dans un autre, cette question est prise, dans l'espace de débat qui a émergé depuis deux décennies, dans un jeu de tensions dont nous avons rencontré la plus importante tout au long de la première partie de ce travail : tension entre, d'une part, les adversaires du « vrac » et, d'autre part, ceux à qui nous n'avons jusqu'ici pas donné la parole, pour qui le vrac, s'il n'est pas un objet de désir, n'est pas pour autant un objet de trouble, et moins encore d'aversion. Cette première tension est solidaire d'une autre, que nous distinguerons par méthode, mais avec laquelle elle semble constituer un couple mal dissociable : la tension entre partisans des « experts » (et autres « autorités »), d'une part, et ceux qui rechignent – apparemment – devant les expertocraties traditionnelles.

Des ordres tout faits à interroger

Nous prendrons comme un *fait* la situation que les adversaires du vrac et les partisans de l'expertise établie déplorent : il y a le *Web*, où figurent des ressources de toutes sortes, dont certaines sont accessibles par les moteurs de recherche existants à un moment donné, et dont d'autres restent inaccessibles au moins durant une certaine période de développement historique des moteurs de recherche⁶ ; et il y a la *recherche* de certaines de ces ressources dans le cadre d'une *enquête* visant à apporter une réponse R à une question Q . Tel est le schéma auquel nous nous arrêterons ici, en faisant nôtres les principes suivants d'une problématique théorique de la recherche d'information sur Internet. Le premier principe stipule que, lorsqu'on entreprend d'étudier une question Q , les *ordres* mis sur certains ensembles de ressources – les “well controlled collections” évoquées par Larry Page et Sergey Brin⁷ –, et cela généralement *d'un point de vue étranger* à l'enquête que l'on mène, doivent

être regardés par méthode comme des *désordres*, dont il s'agit pourtant de tirer profit. En d'autres termes, ce principe énonce qu'il *n'existe pas d'ordre privilégié à vocation universelle*, et qu'il n'existe pas même d'ordre constitué *par ailleurs, antérieurement* à l'enquête (conduite à propos de la question *Q*), et que cette enquête devrait privilégier. À cet égard, un principe *de doute* doit prévaloir, qui conduit à regarder tout ordre même apparemment approprié comme *un désordre à interroger*. De ce point de vue, ainsi, le système de textes composant une encyclopédie organisée de façon classique – antérieurement au surgissement du « monde Internet » – sera *par principe* regardé comme un désordre au moins aussi contingent que tel ensemble de ressources réunies par un moteur de recherche en réponse à une certaine requête.

Nous reprenons ainsi à notre compte en tant que principe ce qui est par ailleurs la *réalité* du comportement d'un moteur de recherche vis-à-vis d'un système « organisé » de ressources, tel qu'un *annuaire* peut par exemple en proposer : le moteur de recherche, en effet, ne reconnaît pas l'*annuaire comme tel*, dans la mesure notamment où il n'en « connaît » que les pages qui lui sont accessibles. L'idée derrière ce principe, du point de vue de la recherche d'informations, est, bien entendu, que si un ordre apparaît pertinent dans l'enquête sur la question *Q*, il devra *in fine* être *(re)constitué par l'enquêteur*, par exemple à partir de la moisson de résultats fournis par tel ou tel moteur de recherche ainsi que des liens hypertextes (*hyperlinks*) que les pages proposées et consultées offriront à l'enquêteur. Le même principe exactement s'applique à l'utilisation des résultats fournis par un moteur de recherche : qu'on en approuve ou non les critères, tout moteur de recherche fournit des résultats *ordonnés*, et met ainsi *de facto* un ordre sur un ensemble de ressources accessibles à l'utilisateur ; mais nous considérerons aussi – par principe – que cet ordre doit être regardé comme un désordre à partir duquel un « système » pourra éventuellement être créé ou recréé par l'enquêteur.

Un symptôme : Shmoogle

La déconstruction des ordres documentaires peut être aujourd'hui aisément illustrée à l'aide du logiciel *Shmoogle* dû à Tsila Hassine.

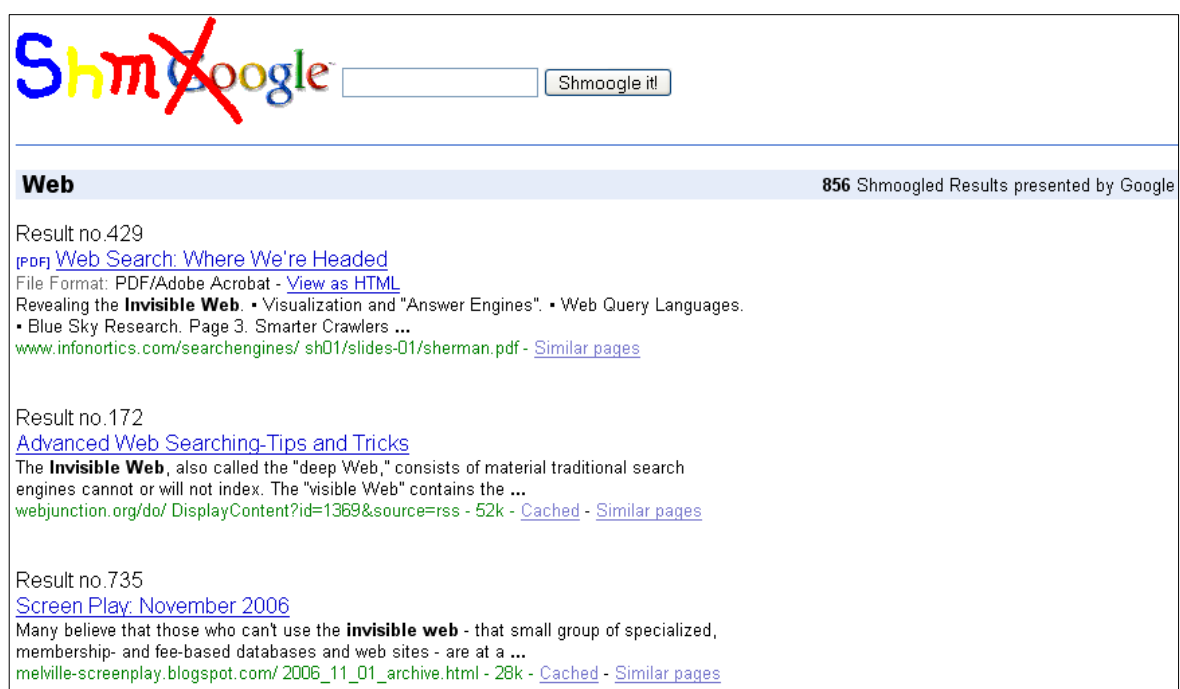
À propos de la créatrice de Shmoogle, on peut lire sur le site du Piet Zwart Institute de Rotterdam cette notule biographique⁸ : “Tsila Hassine grew up in Tel Aviv, Israel, where she completed B.Sc's in Mathematics and Computer Science. She spent 2003 at the New Media department of the HGK Zürich [Zürcher Hochschule der Künste, École supérieure des arts de Zürich], and in 2004 she joined the Piet Zwart Institute in

Rotterdam, where she pursued an MA in Media Design, until graduating in June 2006. Tsila is currently a researcher at the Jan Van Eyck Academie in Maastricht."

Shmoogole offre à la demande un *échantillon au hasard* des résultats que Google propose en réponse à une requête donnée ⁹. Prenons ainsi pour requête "**invisible web**". On a reproduit ci-après le haut de la première page des résultats affichés par Google, suivi du début d'un échantillon de 856 résultats extraits par Shmoogole.



The screenshot shows the Google search interface. At the top is the Google logo. To its right are navigation links: Web, Images, Actualités, Maps, Nouveau!, Groupes, and plus ». Below these is a search bar containing the text "invisible web" and a "Rechercher" button. To the right of the search bar are links for "Recherche avancée" and "Préférences". Below the search bar, it says "Rechercher dans : Web (selected), Pages francophones, Pages : France". The results section is titled "Web" and shows "Résultats 1 - 10 sur un total d'environ 403 000". The first result is titled "Invisible web invisible" and describes various resources for discovering the invisible web. The second result is titled "Invisible web, web caché : bibliographie, articles, études" and provides a bibliography of articles and studies. The third result is titled "Explorer le web invisible" and provides information about the invisible web.



The screenshot shows the Shmoogole search interface. At the top is the Shmoogole logo, which is a stylized "Shm" followed by a red "X" over the word "Google". To the right of the logo is a search bar and a "Shmoogole it!" button. Below the search bar, it says "Web" and "856 Shmoogled Results presented by Google". The results section shows three results. The first result is titled "Web Search: Where We're Headed" and is a PDF file. The second result is titled "Advanced Web Searching-Tips and Tricks" and describes the invisible web. The third result is titled "Screen Play: November 2006" and discusses the invisible web and its membership.

Voici maintenant, pour la même requête, un second échantillon extrait de la même population de réponses de Google (on suppose ici que cette population n'a pas varié entre-temps).

Shmoogled

Web 856 Shmoogled Results presented by Google

Result no.637
[Classes at the Edge](#)
Improve searching skills with an in-depth look at search engines, directories, and the **invisible web** including advanced searching options and tips. ...
thelibrary.springfield.missouri.org/programs/edge/classes.cfm - 37k - [Cached](#) - [Similar pages](#)

Result no.172
[Advanced Web Searching-Tips and Tricks](#)
The **invisible Web**, also called the "deep Web," consists of material traditional search engines cannot or will not index. The "visible Web" contains the ...
webjunction.org/do/DisplayContent?id=1369&source=rss - 52k - [Cached](#) - [Similar pages](#)

Result no.284
[PDF The Deep Web: Resource Discovery in the Library of Texas](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
While the terms '**invisible Web**' and 'deep Web' are frequently used interchangeably, The **invisible-Web** Directory: A Browsable Deep Web Directory ...
www.unt.edu/wmoen/publications/TLJarticle_deepweb_PrePrintFeb2004.pdf - [Similar pages](#)

On peut par ce moyen explorer aléatoirement les profondeurs du Web *visible*, tel du moins que le donne à connaître le moteur de recherche Google. Il ne s'agit là, au vrai, que d'une autre manière de « faire parler » le Web à l'aide d'un moteur de recherche ; et l'on pourrait même imaginer – naïvement ! – que Google rajoute à ses options de « recherche avancée » la fonctionnalité offerte par Shmoogled.

En réalité, Shmoogled procède bien d'une intention critique à l'endroit de Google, comme le texte suivant, qui figure sur le site du Piet Zwart Institute, l'indique non sans maladresses, mais sans ambages (Hassine, s. d.).

Shmoogled: No More Google Goggles!

What is Shmoogled? Shmoogled is simply a Google randomizer. It takes Google query results and presents them in random order. One click – and the user is liberated from the domination of the search engine and corporate Internet, back to the real (virtual) world.

.....

The gap between retrievability and accessibility is widening in direct proportion to the Web's exponential growth. Every HTML page is retrievable – the user types the URL, and it's there. But is every page equally accessible? What are the chances a user will accidentally surf her way to a random page?

There are other issues at hand, concerning the democratic character of the Web. The web, as mentioned above, was built by the masses. It is the knowledge of the masses that has brought the Internet and its content to its current size and state, and it isn't served well by the search engines. Google's PageRank link-based algorithm allows for the emergence of a Pareto traffic distribution, meaning that 20% of the sites get 80% of the traffic. Moreover, viewing the net through Google goggles distorts the proper character of the Web. Since Google presents its results according to their "popularity", the first impression we receive of the "state of the net" is indelibly shaped by the first results page. This page doesn't necessarily represent the Web's (majority) answer to our query.

Shmoogle is the remedy. Since Shmoogle presents the results in a completely random order, the impression a user receives thereby is a better reflection of the Web's "opinion" concerning that query.

Personally, I don't believe Shmoogle should replace Google, but rather complement it. I believe it does. It should exist alongside Google as an alternative to the Google-dictated information hierarchy, and also bring back a bit of the serendipity we could all sometimes benefit from.

Remember: There is no single best answer, and no power-free hierarchy.

Il ne semble pas que, à l'heure où nous écrivons ces lignes, la firme Google ait réagi à l'apparition de Shmoogle, même si la chose est attendue par certains observateurs (Durand, 2005). L'idée à l'origine de Shmoogle semble en vérité profondément antinomique du projet googléen de fournir la « meilleure information possible » répondant à une question donnée, projet que symbolise de façon non anecdotique le lien "I'm feeling lucky" (en français : « J'ai de la chance ») proposé à l'utilisateur par Google et censé idéalement l'amener directement sur le « résultat parfait », qui comblerait l'attente de celui-ci.

Le bouton "I'm feeling lucky" ferait perdre à la firme Google quelque 110 millions de dollars par an, car son usage implique qu'aucune page de résultats n'est affichée, ce qui se traduit par une « perte de clics » sur les liens sponsorisés (sponsored links), lesquels fournissent à Google l'essentiel de son revenu. Il semble que Google préfère cependant maintenir ce bouton, et cela au nom de son idéal du "best result [first]" (Abondance, 2007).

Dans une conférence donnée à Paris le 19 juin 2007 dans le cadre d'un "Google Press Day", Marissa Meyer, vice-présidente de Google en charge des "search products and user experience", précisait ce point de vue dans les termes suivants ¹⁰.

In the early days of search engines, users got used to hunting and packing through results, clicking the next button, going through the first few pages to try finding results. We realized that our job was trying to find the best result first.

Pourtant, sans négliger aucun des apports des moteurs de recherche (dont les performances sont en constante évolution), nous adopterons ici le point de vue déjà explicité : face à la volonté de bonne adéquation *a priori* affichée par le fournisseur de résultats, l'utilisateur fera valoir, dans la dynamique même de son travail d'enquête, son ambition de rechercher une pertinence d'abord *conjecturée*, ensuite *constatée*, enfin *vérifiée*. Ainsi, aux savoirs et savoir-faire des fournisseurs de résultats quels qu'ils soient, nous opposerons les savoirs et les savoir-faire – et plus généralement les praxéologies – *potentiellement mobilisables par l'utilisateur* enquêtant à propos d'une certaine question Q .

L'expertise supposée ne suffit pas

Ainsi que nous l'avons rappelé, la vision exposée jusqu'ici, dans laquelle on entend tirer profit des ordres documentaires préétablis sans valider pour autant les points de vue ayant présidé à leur établissement, semble marquée par une autre difficulté au regard du point de vue classique en matière documentaire : elle paraît en effet ignorer les « experts », les « savants », les « autorités », et ne se fier qu'au seul génie de l'enquêteur, à son *ingenium*, à sa capacité d'invention. En réalité il n'en est rien. L'expert d'une question donnée est celui qui, à propos des réponses faites ici ou là (R^\diamond), et de la réponse qu'il aura à cœur de faire connaître après l'avoir produite (R^\heartsuit) aura pu et aura su mettre en œuvre une *dialectique des médias et des milieux* rigoureuse autant que vigoureuse. Mais plusieurs facteurs viennent compliquer une telle notion. Tout d'abord, il n'est pas toujours facile, pour l'enquêteur profane en un certain domaine, d'identifier celles et ceux qui y développent véritablement une activité d'expert. Ensuite, telle personne qu'on peut regarder comme ayant habituellement un fonctionnement praxéologique d'expert pourra oublier les exigences de son « expertise » à propos de telle question dont le profane sera pourtant enclin à penser qu'elle en relève typiquement : il suffira de légères variations pour que l'assignation d'expertise perde tout fondement. Enfin, l'enquêteur – ou le collectif d'enquête – qui, sur telle question Q^* engendrée par la question Q , n'est pas, en règle générale, un expert (même s'il est peut-être en train de le devenir), peut fort bien se méprendre sur tel élément de réponse que le meilleur expert du moment lui aura fourni, en une communication où, même quand l'émission est bien contrôlée, la réception reste encore incertaine. De là que l'enquêteur *doive travailler sur le*

message reçu, quand bien même la *source* de ce message serait la plus parfaite et la plus sûre possible.

Le point de vue de méthode ainsi esquissé éclairera les développements à venir. Notons en passant que ce point de vue paraît consonant avec le principe qui, semble-t-il, sous-tend l'une des rares « défenses et illustrations », due au philosophe Michel Serres, de l'encyclopédie en ligne Wikipédia. Dans un entretien sur une radio nationale le 25 février 2007, celui-ci, qui préside aux destinées du « Corpus des Œuvres de Philosophie en Langue française » (publié aux éditions Fayard depuis 1984), s'exprimait dans les termes suivants ¹¹.

[Wikipédia est] une entreprise qui m'enchanté parce que, pour une fois, c'est une entreprise qui n'est pas gouvernée par des experts. J'ai une grande grande confiance dans les experts, bien entendu. À qui voulez-vous que je fasse confiance sinon à des experts ? Mais cette confiance envers les experts est limitée puisque les experts, qu'ils soient mathématiciens, astronomes ou médecins ne sont que des hommes. Par conséquent, ils peuvent se tromper et il y a là dans cette entreprise de liberté, de communauté, de vérification mutuelle, quelque chose qui, dans la gratuité, la liberté, m'enchanté complètement et me donne une sorte de confiance dans ce que peut être un groupement humain.

Il s'agit là d'un point de vue encore minoritaire, qui a pour nous le mérite de faire contraste avec des points de vue aujourd'hui (apparemment) « évidents », et dont pour cela on oublie aisément qu'ils participent *objectivement* des conditions et contraintes déterminant l'espace praxéologique au sein duquel il convient de situer et d'analyser ce qui relève (ou pourrait relever) d'une « didactique du Web ».

CHAPITRE 7

Deux analyses praxéologiques

Le livre qui a fourni sa matière à la première partie de notre travail n'avait nullement pour objet « d'enseigner l'Internet » ou « le Web ». Par contraste, nous examinerons dans ce chapitre deux exposés ayant l'un et l'autre une *intention didactique* à l'endroit de leurs lecteurs potentiels à propos de praxéologies – qui y sont soit présentées, soit évoquées – appelées par la recherche d'information sur Internet (RII). Ces analyses praxéologiques ébaucheront le repérage du curriculum (ou des curriculums) que l'on peut regarder comme actuellement en voie d'émergence lente – sinon de stabilisation – en matière de RII. La création d'une tradition curriculaire se fait en effet *dans la durée*, sous la pression non seulement de l'évolution sociale extérieure, mais aussi en fonction de l'obsolescence interne des choix curriculaires antérieurs, laquelle fait que, tout à coup, on s'aperçoit que tel objet d'enseignement a perdu de son crédit comme objet à *enseigner*. Ce que nous saisirons à examiner les textes choisis n'est donc *a priori* formé que de fragments de curriculums *possibles*, partiellement réalisés dans ces textes.

Le choix de ces deux textes n'a en vérité d'autre fondement que leur commune facture rudimentaire, relative à un contenu dont pourtant bien des points, nous semble-t-il, n'en ont pas moins le statut de petites énigmes pour le tout-venant des usagers. Le premier est en français. Pour le contraste, le second a été choisi en anglais, alors même que, comme nous le découvrirons plus loin, cette langue constitue, pour la diffusion praxéologique en matière de RII, une frontière dont le franchissement, pour l'utilisateur français, ne va pas de soi.

La RII, un thème presque absent

Le premier exposé, sur lequel nous nous arrêterons maintenant, est un petit ouvrage récemment publié à l'intention du grand public, ou plus précisément de cette partie du grand public supposée avoir acquis depuis peu « un ordinateur tout neuf, “avec Windows Vista” ». Intitulé *Windows Vista et Internet facile*, paru aux éditions First (Paris, 2007), qui publie par ailleurs la collection « Pour les Nuls », cet abrégé est signé de Henri Lilen, auteur qui a écrit

un nombre considérable d'ouvrages pour l'utilisateur non spécialisé, voire pour le néophyte, dont un nombre respectable aux éditions First, qui lui doivent notamment les titres ci-après ¹².

Facile sécurité de l'ordinateur (2007) / *Mon aide mémoire pour dépanner mon ordinateur* (2007) / *Spécial débutant - Je me mets à Excel* (2007) / *Windows Vista facile* (2007) / *L'ordinateur à la maison facile* (2007) / *Internet* (2006) / *Office 2007 par étapes* (2006) / *Spécial débutant - Je résous les problèmes de Windows XP* (2006) / *Open office.org 2.0 pour les Nuls* (2006) / *Guide des meilleures astuces Windows XP 2006* (2006) / *400 questions-réponses pour améliorer, protéger, dépanner votre PC* (2005) / *Openoffice.org* (2005) / *Guide des meilleures astuces Windows XP* (2005) / *Poche Micro Internet* (2006) / *PC* (2005) / *Guide des meilleures astuces Internet sur PC* (2004) / *Protéger son PC en 10 étapes* (2004) / *Je me mets à Excel* (2003)

L'exercice auquel cet auteur s'est livré dans l'ouvrage que nous allons parcourir rapidement n'est donc pas isolé. L'expérience de l'auteur avec les publics visés par ses multiples écrits peut être tenue pour significative d'une certaine problématique de diffusion de praxéologies dans le domaine « Informatique & Internet ». Voici la table des matières abrégée de l'ouvrage.

Table des matières

Chapitre 1. Ce que vous devez savoir avant de commencer / Chapitre 2. Découvrez Windows Vista / Chapitre 3. Pratiquez Windows Vista / Chapitre 4. Familiarisez-vous avec le clavier, la souris ou le dispositif de pointage / Chapitre 5. Découvrez les programmes et les documents / Chapitre 6. Vos programmes et vos documents au quotidien / Chapitre 7. Place aux loisirs ! / Chapitre 8. Partez à la découverte d'Internet / Chapitre 9. Surfez sur le Web ! / Chapitre 10. Le courrier électronique au quotidien / Chapitre 11. Devenez expert en messagerie / Veillez à votre sécurité / 20 sites qui méritent votre attention / Index

La table des matières développée du chapitre 1 se présente comme suit.

Chapitre 1. Ce que vous devez savoir avant de commencer

L'ordinateur et son fidèle serviteur, Windows Vista / Tout ce que vous pouvez faire en surfant sur Internet / Êtes-vous bien équipé ? / L'ordinateur et l'Internet, combien ça coûte ? / Et maintenant, démarrez !

Qu'apprend-on dans la deuxième section de ce premier chapitre à propos de ce qu'on peut faire « en surfant sur Internet » ? La section est découpée en neuf sous-sections dont voici les titres : *Achetez en ligne / Téléchargez de la musique ou des films / Visionnez des vidéos / Participez à des forums / Consultez des sites d'informations / Trouvez du travail / Restez en contact avec les amis et la famille / Voyez vos interlocuteurs en direct / Achetez, mais vendez aussi !*

Comme on n'aura pas manqué de le noter, la question de la recherche d'information à l'aide de moteurs de recherche *n'apparaît pas* dans cet inventaire. La sous-section intitulée « Consultez les sites d'informations », ainsi, comporte simplement cette indication (p. 13) : « Les sites Web d'informations se sont multipliés. Tous les grands journaux ont le leur. Vous pouvez les consulter à tout moment, et le plus souvent, c'est gratuit. » (Ce texte est accompagné d'une copie d'écran d'une page du site LeMonde.fr.) Toutefois, le chapitre 8, on l'aura observé, s'intitule « Partez à la découverte d'Internet », tandis que le chapitre 9 a pour titre « Surfez sur le Web ». Qu'en est-il des sujets qui y sont abordés ? Voici la table des matières développée de ces deux chapitres.

Chapitre 8. Partez à la découverte d'Internet

Ouvrez le navigateur Internet Explorer / Affichez votre première page Web / Cliquez sur un lien pour surfer / Visitez un site en tapant son adresse / Revenez sur une page précédente / Naviguez en cliquant sur des onglets / Retrouvez la page d'accueil d'un site

Chapitre 9. Surfez sur le Web !

Trouvez l'adresse du site Web qui vous intéresse / Recherchez un site avec Google / Récupérez du texte provenant d'une page Web / Récupérez une image ou une photo sur une page Web / Adoptez les Favoris pour naviguer plus rapidement / Prévisualisez puis imprimez une page Web / Visitez des blogs

Le sujet qui nous occupe ici – la recherche d'information sur Internet – pourrait être abordé, comme on le voit, dans la deuxième section du chapitre 9 : « Rechercher un site avec Google ». Pour plus de sûreté, nous parcourrons cependant l'ensemble des chapitres 8 et 9 – en commençant par le chapitre 8.

Le Web et ses liens

Au début du chapitre 8 – nous sommes à la page 112 d'un ouvrage qui en comporte 174 –, on suppose que le lecteur a installé sa *connexion à Internet*, en passant par un *fournisseur*

d'accès à Internet, un FAI. « Connexion à Internet », « FAI », tels sont les objets nouvellement rencontrés à ce stade. La notion de FAI fera d'ailleurs l'objet, à la page suivante, de cette définition : le FAI, c'est le « fournisseur auprès duquel vous vous abonnez et qui vous donne accès à Internet ». La notion de l'Internet elle-même semble apparaître ici pour la première fois ; elle fait donc également l'objet d'une (très sobre) définition : « Internet est un réseau mondial d'ordinateurs interconnectés. Ils vous fournissent de nombreux services. En anglais, *net* signifie réseau. » On aura noté l'apparition, au passage, d'une autre notion, celle de *service* fourni par l'Internet. Pour « surfer » sur l'Internet, le lecteur va alors être invité à « ouvrir un programme spécial, un navigateur ».

Le nom anglais surf désigne – peut-être par onomatopée – le ressac, les déferlantes, les brisants sur le rivage, waves breaking on the shore. D'après le Online Etymology Dictionary (Harper, 2001), le verbe to surf, au sens de to ride the waves of the sea with a surfboard, apparaît en 1917. Ce même dictionnaire date de 1993 le « sens Internet » du mot (“Surf the internet or the world wide web”), à savoir to look around casually and randomly, without seeking anything in particular.

Le navigateur est désigné ici comme un *programme*, mot générique qui n'apparaît pas dans l'index de l'ouvrage – non plus que le mot *navigateur*, du reste.

Dans le langage informatique, le mot de programme est historiquement premier : en anglais, program (au sens de programme informatique) est attesté dès 1945 et programmer (« programmeur ») dès 1947. En français, le mot logiciel est plus récent : il répond d'abord au software de l'anglais « informatique », qui apparaît vers 1960 et fait contraste avec hardware (apparu vers 1947). De ce sens premier, il évolue ensuite – on parlera alors d'un logiciel, et plus seulement du logiciel comme opposé au « matériel » – pour désigner « tout ensemble particulier d'informations qui préside à un comportement donné de l'ordinateur, c'est-à-dire tout programme informatique, par exemple tout traitement de texte, tout tableur, etc. » (Ganascia, 1998, p. 166). Le même auteur ajoute : « En ce sens second, un logiciel est une partie détachable et commercialisable du logiciel pris au sens premier, cette partie étant douée de fonctions bien identifiables. » « Logiciel » correspond alors à l'anglais application, mot pour lequel le OneLook Dictionary fournit cette définition : “a program that gives a computer instructions that provide the user with tools to accomplish a task (Example: ‘He has tried several different word processing applications’).”

On observera également que, contrairement à *net*, le nom anglais correspondant, soit *browser*, n'est pas introduit par l'auteur.

En France, la Commission générale de terminologie et de néologie préfère « logiciel de navigation », mais donne « navigateur » comme synonyme. En anglais, le terme générique est browser, mot pour lequel le OneLook Dictionary propose ces deux acceptions : “a program used to view HTML documents”, “a viewer who looks around casually without seeking anything in particular”. L'idée essentielle est ici celle de cheminement non finalisé dans un domaine d'activité quelconque. Le dictionnaire déjà mentionné donne ainsi de browse les acceptions suivantes : “reading superficially or at random / eat lightly, try different dishes / look around casually and randomly, without seeking anything in particular / shop around; not necessarily buying / feed as in a meadow or pasture”. Dans cette perspective, David Weinberger écrit dans un livre récent (Weinberger, 2007, p. 9) : “Before the Web, the word browsing was usually a polite way of telling a salesperson to buzz off. ‘I’m just browsing,’ you reply with a little smile. With that word, a customer declares a lack of commitment. With that smile, she asserts that she’s within her rights: ‘Just try and stop me, salesboy!’ Browsing is more than window-shopping, fantasizing about what it would be like to own something or resenting those who do. You browse when you intentionally ignore the organizational structure the store has carefully imposed on its stock.” En anglais, les vocables to browse et browser laissent entendre ainsi une certaine absence de finalisation (qui s'entend aussi dans to surf, on l'a vu), et, ainsi que le note Weinberger, une certaine inattention aux ordres établis, ce qu'on ne retrouve guère en français dans naviguer et navigateur.

Dans le cas examiné, le navigateur est (Microsoft) Internet Explorer, puisqu'il s'agit d'initier le lecteur à l'emploi de Windows Vista. L'auteur précise alors, copie d'écran à l'appui, la manœuvre à opérer pour *lancer* le navigateur. C'est celui-ci, nous dit-il, qui va permettre « de découvrir les joies du surf » sur le Web. Une définition du Web est alors proposée : « Le Web, ou le World Wide Web (la toile mondiale) est le service le plus spectaculaire d'Internet. Il donne accès à des milliards de pages d'information. » Le Web n'est donc que *l'un* des services disponibles sur l'Internet ; mais c'est aussi l'un des plus remarquables : selon l'auteur, c'est « le plus spectaculaire ».

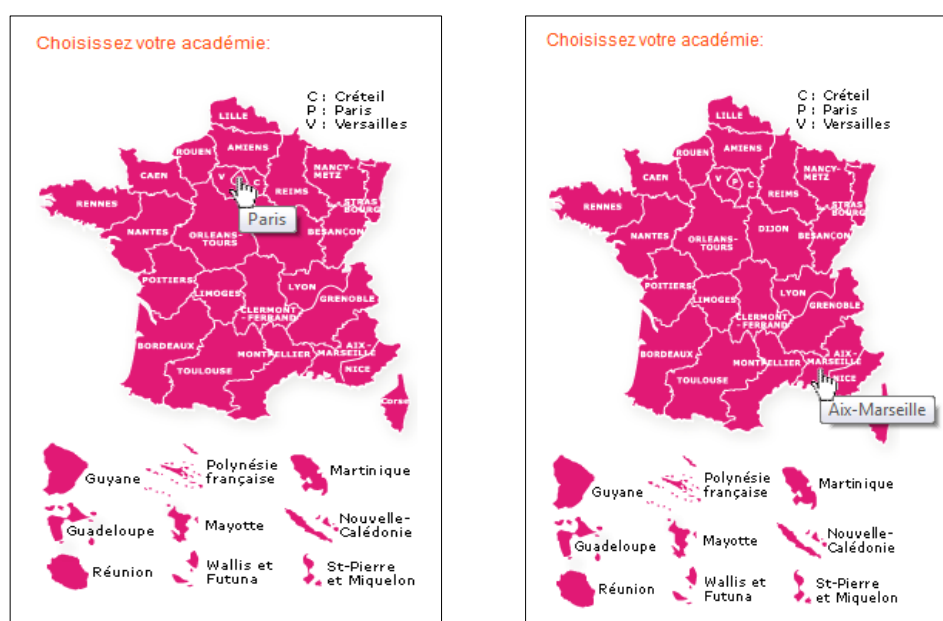
Le folklore habituel des écrits sur l'Internet aurait voulu ici une glose du type « le Web, service conçu en 1989 par l'anglais Tim Berners-Lee au CERN (en Suisse), etc. ». Le recours à l'image utilisée, banale mais qui peut être discutée (en quoi le Web est-il plus spectaculaire que le courrier électronique par exemple ?), est sans doute lié à la volonté de l'auteur de restreindre les informations et les notions mises en jeu dans l'ouvrage.

Lorsque le logiciel Internet Explorer est lancé, indique l'auteur, il apparaît une page dite « page d'accueil »¹³. Cette dernière notion fera l'objet d'une courte définition (p. 115) : « première page d'un site Web procurant généralement un bref sommaire de son contenu et des liens permettant de visiter d'autres pages ». La page d'accueil à laquelle on arrive en l'espèce est supposée être la page d'accueil du site MSN¹⁴. Elle s'ouvre dans une fenêtre dont l'auteur entreprend alors de détailler les différents composants – *barre d'état* (« tout en bas »), *barre d'outils*, *barre de menu*, *barre de titre* (laquelle, en principe, « rappelle le nom du site auquel vous êtes connecté »). Trois autres composants sont mentionnés qui ont un lien plus spécifique avec la navigation sur le Web. Tout d'abord, la *barre d'adresses* « sert à taper l'adresse du site où vous voulez vous rendre ». Ensuite, la *zone de recherche* permet soit de rechercher un mot dans la « page courante » (avec Internet Explorer, car il en va autrement par exemple avec Mozilla Firefox), soit de « rechercher un site *via* un moteur de recherche »¹⁵. Enfin, la *barre d'onglets* conserve, pour chacun des sites visités, un onglet sur lequel il suffit de cliquer pour revenir sur ce site. Le dispositif des onglets semble être tenu par l'auteur pour important, à en juger par la présence d'un commentaire spécifique (p. 115) : « On veut toujours afficher le maximum de choses à l'écran, mais l'encombrement est vite atteint. Une astuce permettant de tourner cette difficulté consiste à utiliser des *onglets*, comme dans le classeur physique dans un bureau. » Pourtant, rien n'est dit en ce point de la manière de faire qu'un lien sur lequel on clique s'ouvre dans un onglet – ce qui n'est pas automatique, mais résulte d'une action délibérée de la part de l'utilisateur. L'auteur y reviendra longuement quelques pages plus loin, où il propose tout une sous-section intitulée « Naviguez en cliquant sur des onglets » dans laquelle figurent les indications techniques suivantes (p. 122).

► Pour ouvrir un nouvel onglet lorsque vous suivez un lien sur une page Web, appuyez sur **Ctrl** lorsque vous cliquez sur le lien, ou encore faites un clic droit sur le lien et cliquez sur **Ouvrir dans un nouvel onglet**.

► Si vous possédez une souris à roulette, vous pouvez cliquer sur un lien avec la roulette pour ouvrir un nouvel onglet.

Jusqu'ici, l'auteur n'a fait qu'évoquer la possibilité d'accéder à des pages Web à l'aide d'un moteur de recherche. Hormis pour la page d'accueil de MSN, que le lancement d'Internet Explorer est censée faire apparaître à l'écran, le lecteur ignore encore comment s'y prendre pour accéder à des pages Web (ou à des sites). La sous-section intitulée « Cliquez sur un lien pour surfer » fait rencontrer une première technique, qui « consiste à cliquer sur des liens qui s'affichent fréquemment dans les pages Web » et qu'on appelle aussi des « hyperliens ». La notion de lien apparaît cardinale et donne lieu d'abord à cette brève définition ¹⁶ : « Texte ou image sur lequel il faut cliquer pour se déplacer vers une page du site ou vers un autre site Web. » L'auteur précise les formes concrètes derrière lesquelles un lien peut se dissimuler : *texte*, *image* (qu'on appelle alors, note-t-il, image *cliquable* ou *point chaud*), enfin *carte image* – comme il en va avec la carte ci-après correspondant aux académies en lesquelles est divisé le territoire français ¹⁷.



Un lien peut donc se dissimuler sous des apparences diverses sur une page affichée. L'auteur explicite alors une question cruciale et la réponse utile (p. 117-118) ¹⁸.

Mais comment savoir si un objet est un lien ? C'est très simple :

- Par défaut, un lien textuel est souligné et s'affiche en bleu.
- Si vous placez le pointeur de la souris sur un lien, la barre d'état affiche l'adresse où il mène. Si l'adresse n'apparaît pas dans cette barre d'état, c'est tout simplement qu'il n'y a pas de lien.
- Positionné sur un lien, le pointeur de la souris se transforme en une main, index pointé vers le haut.

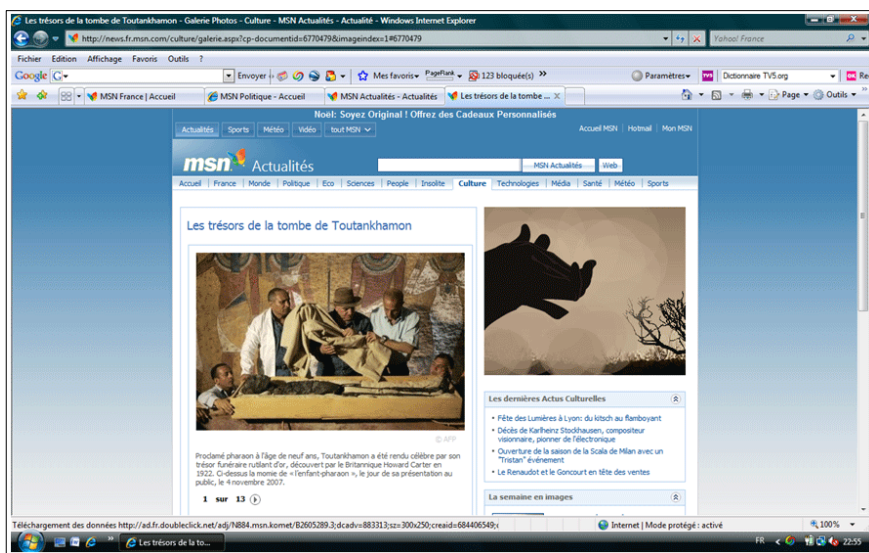


Si aucune de ces manifestations n'a lieu quand vous pointez un objet à l'écran, c'est qu'il n'y a pas de lien, n'insistez pas !

L'importance du rôle attribué aux liens conduit l'auteur à préciser explicitement ce que l'utilisateur pourra voir – et pourra faire – chaque fois qu'il aura identifié un lien sur la page qu'il examine (p. 118) : « Dès que vous cliquez sur un lien, il est dit **suivi**. Internet Explorer modifie sa couleur qui passe souvent du bleu au pourpre. Il reste toujours actif et vous pouvez cliquer dessus. »

Adresses Web et onglets

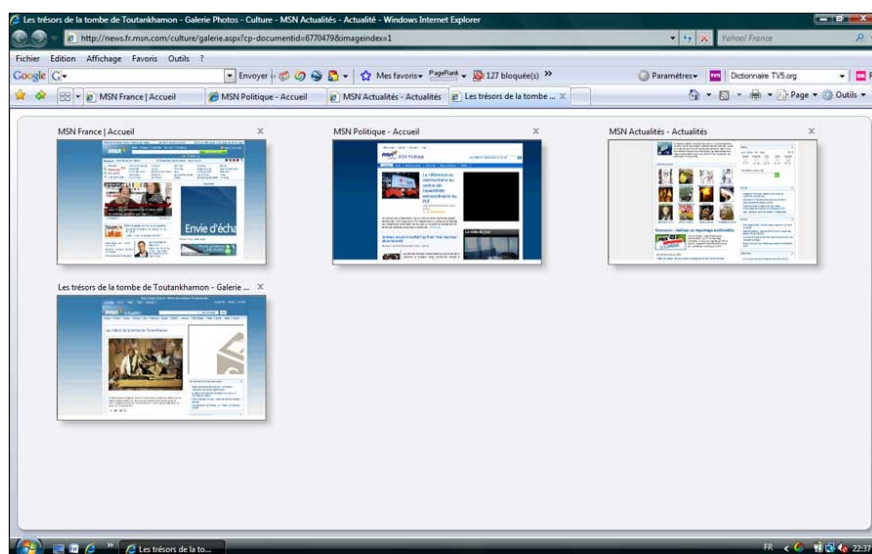
S'il conjugue l'étude du texte de l'auteur avec des essais de mise en pratique, le lecteur de l'ouvrage que nous suivons peut, à ce stade, « surfer sur le Web », mais en étant à la merci des vagues de liens qui se présentent à lui : c'est ainsi que la copie d'écran ci-après montre la page sur laquelle on arrive en suivant un certain cheminement (parmi d'autres possibles) réalisé en trois clics successifs.



L'auteur présente alors une autre manière d'arriver sur une page Web : en *tapant son adresse* dans la barre d'adresses. Il signale d'abord que cette adresse est « parfois » appelée une URL, notation qu'il commente en ces termes : « Ce sigle provient de l'expression **Uniform Resource Locator**, localisateur uniforme de ressources. C'est une adresse sur le Web. » Le lecteur se voit préciser alors comment « taper une adresse » : il lui faut cliquer à l'intérieur de

la barre d'adresse et taper la nouvelle adresse sans s'inquiéter de celle qui y figurait déjà, à ceci près que, si des adresses ont déjà figuré dans la barre d'adresses au cours de la session de travail, elles apparaîtront automatiquement dans une liste déroulante où l'adresse que l'on a entrepris de taper pourrait se trouver : en ce cas, l'utilisateur pourra se contenter de cliquer sur son occurrence dans cette liste ; c'est là ce qu'on appelle, note l'auteur, une *saisie semi-automatique*. Bien entendu si l'adresse à laquelle on s'intéresse n'apparaît pas dans la liste déroulante, il convient de la taper intégralement dans la barre d'adresses. À cet égard, les recommandations suivantes sont faites (p. 120) : « Tapez une adresse Web sans introduire d'espace et toujours en minuscules, en commençant par la formule sacro-sainte `http://` et ce, jusqu'à la première barre oblique. Ensuite, des compléments peuvent demander des majuscules. » L'adresse ayant été tapée, il faut cliquer sur « la petite flèche située sur la droite de la zone d'adresses », à moins qu'on ne choisisse d'appuyer sur la touche **Entrée**. La page demandée s'affiche – en principe. L'auteur conclut : « Vous pouvez ainsi très facilement voyager dans le monde entier. Vous y prendrez probablement vite plaisir. »

La suite du chapitre comporte des indications pratiques non négligeables. Ainsi la section intitulée « Revenez sur une page précédente » commente-t-elle l'utilisation des deux flèches, de sens opposé, qui figurent sous la barre de titre, en haut et à gauche. La sous-section qui suit, « Naviguez en cliquant sur des onglets », a déjà été mentionnée ; elle fournit les indications que nous avons précisées plus haut, et quelques autres encore, notamment à propos des « onglets rapides » : en appuyant sur **Ctrl+Q** on les fait apparaître (voir la copie d'écran ci-dessous) ou disparaître.



Enfin l'ultime section, « Retrouvez la page d'accueil d'un site », attire l'attention sur une vertueuse coutume du Web : la présence habituelle, sur chacune des pages d'un site, d'un lien

placé généralement sur une icône – souvent le logo du site – qui permet de revenir sur la page d'accueil.

À la recherche des moteurs de recherche

Dans ce qui précède, deux traits saillants sont à noter. Tout d'abord, à ce stade, on a encore fort peu avancé vers la recherche d'information sur le Web. Pour aller vers une ressource possible, en effet, il faut jusqu'ici soit connaître l'adresse (URL) où elle se trouve, soit y parvenir à partir d'une page Web à laquelle on aura eu accès, et cela par un cheminement qui exploite autant que faire se peut le système de liens que l'on trouvera devant soi. Ensuite, et peut-être surtout en cette étape liminaire, on ne dispose encore que d'un exposé dont, au plan *technologique* (ce mot étant pris ici au sens de la TAD), la « profondeur de champ » *est très réduite*. Plus clairement qu'ailleurs peut-être, l'auteur a choisi de demeurer autant que possible dans le registre de la « phénoménologie », c'est-à-dire des objets qui apparaissent à l'usager « pur », en se gardant apparemment d'entrer plus avant dans la « nouménologie » du domaine présenté.

Le noumène, selon le mot créé par Kant (1724-1804) à partir du grec nooumena (employé par Platon dans le Timée à propos des Idées, et apparenté à noos, esprit, intelligence), c'est « ce qui est pensé », l'intelligible qui n'est pas l'objet d'une connaissance empirique. L'opposition noumène/phénomène sera notamment reprise par Gaston Bachelard (1884-1962), dès 1932 dans son article Noumène et microphysique (voir Bachelard, 1970). Nous en faisons ici un usage résolument relatif : ce qui est phénomène et ce qui est noumène pour une personne dépend de la position qu'occupe cette personne au sein de l'institution où elle est observée. Ainsi la distinction de la phénoménologie et de la nouménologie n'a-t-elle pas le même contenu pour qui se limite à visiter des sites Web et pour qui est familier de la création de sites Web, par exemple.

Ainsi a-t-on vu le sigle URL présenté comme *parfois* utilisé, et dont l'expression développée est donnée par l'auteur en anglais, avec une traduction française. Mais l'auteur en restera là sur cette entité pourtant emblématique de l'histoire et de la théorie du Web. Dans la même veine, l'auteur mentionne « la forme sacro-sainte `http://` » ainsi que certaines règles touchant à l'usage des majuscules et des minuscules dans une URL ; mais le lecteur n'aura droit à ce sujet, pour tout viatique, qu'à ces notations minimalistes.

Notons ici que l'article de l'encyclopédie Wikipedia en anglais consacré à la notion d'URL s'ouvrait, début octobre 2007, par ces mots : “Uniform Resource Locator (URL) formerly known as Universal Resource Locator, is a technical, Web-related term used in two distinct meanings”. Il se poursuit ainsi : “In popular usage and many technical documents, it is a synonym for Uniform Resource Identifier (URI). Strictly, the idea of a uniform syntax for global identifiers of network-retrievable documents was the core idea of the World Wide Web. In the early times, these identifiers were variously called ‘document names’, ‘Web addresses’ and ‘Uniform Resource Locators’. These names were misleading, however, because not all identifiers were locators, and even for those that were, this was not their defining characteristic...”

La table des matières du chapitre 9, « Surfez sur le Web ! », a été reproduite plus haut. Chacun des chapitres de l'ouvrage comporte, sur la première page, une rubrique intitulée « Découvrez comment... ». On y trouve, sommairement présentés, les principaux *types de tâches* abordés dans le chapitre. Cette liste de types de tâches est en général proche du libellé de la table des matières figurant en tête de l'ouvrage, sans pour autant en être une reproduction *ne varietur*. Le tableau ci-après montre ce qu'il en est pour les chapitres 8 et 9.

	Table des matières	Découvrez comment...
Chapitre 8	1. Ouvrez le navigateur Internet Explorer 2. Affichez votre première page Web 3. Cliquez sur un lien pour surfer 4. Visitez un site en tapant son adresse 5. Revenez sur une page précédente 6. Naviguez en cliquant sur des onglets 7. Retrouvez la page d'accueil d'un site	a) Ouvrir le navigateur Internet Explorer b) Afficher votre première page Web c) Naviguer en tapant l'adresse d'un site Web d) Naviguer en cliquant sur un lien e) Utilisez les boutons d'Internet Explorer f) Surfer en cliquant sur des onglets g) Réafficher la page d'accueil
Chapitre 9	1. Trouvez l'adresse du site Web qui vous intéresse 2. Recherchez un site avec Google 3. Récupérez du texte provenant d'une page Web 4. Récupérez une image ou une photo sur	a) Rechercher des sites spécifiques b) Récupérer du texte dans une page Web c) Enregistrer une image ou une photo dans

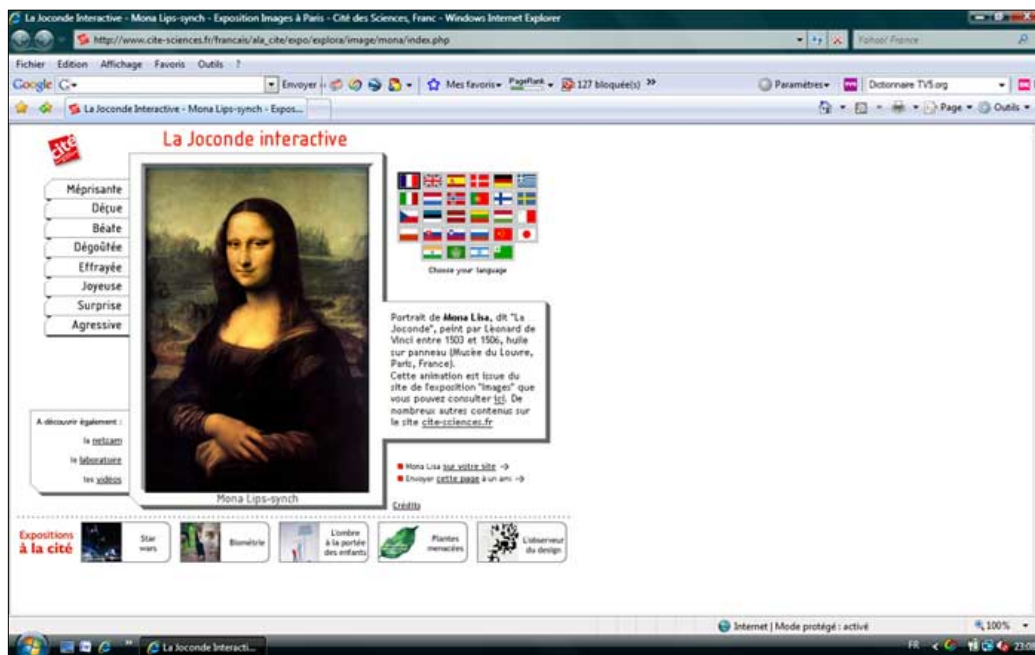
	une page Web	l'ordinateur
	5. Adoptez les Favoris pour naviguer plus rapidement	d) Enregistrez un site dans les Favoris e) Afficher la liste des Favoris f) Ouvrir un site enregistré comme Favori
	6. Prévisualisez puis imprimez une page Web	g) Visualiser avant d'imprimer h) Visiter des blogs
	7. Visitez des blogs	

On observera que, s'agissant du chapitre 8, les différences sont minimes, à une interversion près (il faudrait échanger les points c et d de la rubrique « Découvrez comment... » pour retrouver l'ordre du chapitre). En revanche, pour ce qui est du chapitre 9, les divergences sont plus sensibles. En particulier, la sous-section « Recherchez un site avec Google », présente dans la table des matières, *n'apparaît plus explicitement* dans la liste des types de tâches, en même temps que celle-ci donne une plus grande place à la gestion des Favoris. On peut donc s'attendre à un exposé qui – non sans sobriété, certes – tende à privilégier le stable et l'organisé par rapport à la dynamique buissonnière des recherches sur le Web.

La première sous-section, on l'a vu, s'intitule « Trouvez l'adresse du site Web qui vous intéresse ». Le Web comporte, écrit l'auteur, « des dizaines de millions de sites ». Comment trouver ce qui pourrait nous « intéresser » ? Ou, encore, comment trouver un site que l'on connaît mais dont on n'a pas conservé l'adresse ? On retrouve alors un leitmotiv « pédagogique » de l'auteur, qui s'adresse ici à son lecteur en ces termes¹⁹. « Qu'allez vous devenir ? *C'est simple* encore une fois. » En l'espèce, la solution consiste à demander l'adresse cherchée « à un service de bienfaisance gratuit, un **moteur de recherche** ». L'expression de moteur de recherche fera un peu plus loin l'objet de la définition suivante : « Site Internet, libre et gratuit, servant à trouver les adresses des sites qui vous intéressent. » En l'espèce, le moteur de recherche accessible par défaut est celui de Microsoft, *Live Search*²⁰. Pour l'utiliser, il suffit de taper des mots clés dans la zone de recherche aménagée dans la barre portant la fenêtre des adresses²¹. On rencontre alors l'unique exemple de recherche effective qui sera proposé par l'auteur : ayant choisi pour requête l'expression « La Joconde », il illustre cet essai de recherche en reproduisant la page à laquelle conduit l'un des liens affichés par *Live Search* (voir la copie d'écran, ci-après).

La sous-section suivante, « Recherchez un site avec Google », semble être un additif à ce que l'on vient d'examiner. C'est que, précise l'auteur, « outre celui que propose Internet Explorer par défaut », il existe « de très nombreux moteurs de recherche » ; et, parmi eux,

Google, « le plus célèbre des moteurs de recherche ». L’auteur indique comment y avoir accès (il suffit de taper l’adresse <http://www.google.fr>) et prend pour exemple de requête le mot *chocolat* : un clic sur le bouton « Recherche Google » et « ce service de recherche vous indique quels sont les sites qui parlent du chocolat, en commençant par ceux qu’il juge les plus pertinents. »



L’auteur conclut : « Chaque site est un lien : pour lui rendre visite cliquez dessus. » On retrouve là, *in fine*, l’un de ces minuscules accommodements déjà rencontrés – car un lien n’est pas un site (il pointe vers un site, ou plutôt vers une *page* d’un site). Cela noté, la suite du chapitre – que nous abandonnerons ici – n’apportera rien de plus à la question qui nous occupe principalement.

“Basic Computer Skills”

Nous nous arrêterons plus brièvement sur un deuxième ouvrage paru, lui, en 2003, dans la célèbre collection des “*Teach yourself*” Books²². Signé de Moira Stephen, il est intitulé *Basic Computer Skills* et comporte huit chapitres dont voici les titres.

1. information and communication technology / 2. windows essentials / 3. common skills / 4. word processing / 5. spreadsheets / 6. databases / 7. presentations / 8. internet and e-mail. Nous examinerons essentiellement le huitième et dernier chapitre, dont voici la table des matières détaillée : 8.1 First steps with the Internet / 8.2 Browser settings / 8.3 Web navigation / 8.4 Searching

the Web / 8.5 Favorites / 8.6 Printing / 8.7 E-mail / 8.8 Send a message / 8.9 Inbox / 8.10 Attachments / 8.11 Address Book / 8.12 Message management.

Comme dans l'ouvrage précédemment examiné, la page de titre d'un chapitre comporte une liste de types de tâches abordés dans le chapitre : nous la reproduisons ci-après pour ce qui est du chapitre 8.

In this chapter you will learn

- how to search the Web / • how to customize Internet Explorer / • how to manage favorites /
- about search engines / • how to use Hotmail / • how to send attachments / • about managing your messages and contacts

L'auteure commence par indiquer comment ouvrir « le navigateur » (*browser, browser application*), qui est ici, à nouveau, Internet Explorer, et précise la notion de *page d'accueil (Home Page)*, en différant jusqu'à la page suivante la question du choix par l'utilisateur de sa propre page d'accueil. Elle introduit alors la notion de page Web (*web page*), et définit un site Web (*web site*) comme “a collection of linked pages holding the information for an individual or an organization”. Chaque page Web à visiter possède une adresse, qui est une URL (*Uniform Resource Locator*). L'auteure précise la signification de la syntaxe d'une URL, en allant à peine plus loin, sur ce sujet, que l'auteur précédemment suivi. Le préfixe `http://` identifie, écrit-elle ainsi, “the address as a hypertext (web page) URL”. La notation `www`, signale-t-elle encore, est usuelle dans les adresses Web, mais... elle n'y est pas toujours présente, comme le montre l'adresse <http://news.bbc.co.uk> par exemple. Quant à leur longueur, les adresses peuvent être “very short” pour certaines, “more complex” pour d'autres. Le conseil est donné de noter soigneusement les adresses « intéressantes » rencontrées ici ou là (dans la presse, à la télévision, sur des produits commerciaux, etc.) : car, dès lors qu'il connaîtra l'URL d'une page Web qu'il souhaite visiter, l'utilisateur n'aura plus qu'un geste à faire – “key it into the **Address** field and press [**Enter**]”.

Le lecteur apprend ensuite à sauvegarder une page Web. Puis vient l'examen plus détaillé de la fenêtre du navigateur et des outils qu'on peut y faire afficher et de la manière de personnaliser (*customize*) cette fenêtre, notamment en y faisant apparaître ou disparaître les diverses barres d'outils disponibles. La personnalisation des barres d'outils fait l'objet d'une attention particulière, mais sans explicitation des fonctionnalités des différents boutons disponibles²³. Intitulée “Browser settings” (« Paramétrer le navigateur »), cette deuxième section se termine par un développement sur la manière d'éviter l'affichage des images

(*pictures, graphics*) afin d'accélérer le chargement des pages Web – problème aujourd'hui dépassé (provisoirement ?) du fait de la généralisation des connections à haut débit. L'exposé se poursuit alors par une troisième section au titre prometteur : "Web navigation". À nouveau il est rappelé à l'utilisateur que, s'il connaît l'URL de la page qu'il veut visiter, il lui suffit de taper cette URL dans la zone d'adresses (*address field*) et d'appuyer ensuite sur la touche « Entrée ». Selon un plan déjà rencontré, l'auteure mentionne alors les hyperliens que l'on peut trouver sur une page Web et précise à ce propos ce que l'on avait déjà appris dans l'ouvrage précédent : un hyperlien peut être porté par du texte (ce texte étant "usually colored blue with a blue underline") ou par une image – dans ce cas, si le pointeur de la souris est positionné sur l'image, on voit apparaître "a pointing hand" ²⁴.

On aura noté la grande ressemblance des exposés. L'auteure, ainsi, indique ensuite que, pour aller vers l'endroit où pointe l'hyperlien identifié, il suffit de cliquer sur le texte ou l'image qui le porte, avant de conclure : "as you jump from one location to another following the hyperlinks that interest you, you are *surfing* the net." Elle ajoute encore à cela la mention des flèches permettant de passer à la page précédente (*Back*) ou à la page suivante (*Forward*), lorsqu'une telle page existe, ainsi que l'usage de la rubrique *Historique* (*History*). Comment donc expliquer que deux ouvrages simples choisis par nous d'une façon essentiellement erratique se ressemblent si fort ? On avancera ici la conjecture qu'il y aurait, dans cette institution internationale constituée autour de la formation des commençants au domaine « Informatique & Internet », un *texte du savoir en émergence*, sans doute susceptible de variantes plus ou moins importantes ²⁵. Dans le cas examiné, la ressemblance est confirmée par la suite de l'exposé examiné. La section qui suit, intitulée "Searching the Web", traite du cas où on ne connaît pas l'URL de la page ou du site que l'on recherche. On utilise alors des mots clés (*keywords*) que l'on propose au dispositif de recherche d'Internet Explorer, que l'on lance en cliquant sur l'icône appropriée (*Search*), puis en choisissant l'option "Find a Web page" dans le menu qui s'affiche alors ²⁶. La recherche confiée ainsi au dispositif de recherche d'Internet Explorer s'achève par l'affichage de pages de résultats.

La syntaxe des requêtes

Au-delà du système de recherche intégré au navigateur utilisé, l'exposé s'élargit aux autres moteurs de recherche (*search engines*), dont plusieurs sont mentionnés (Ask Jeeves, Alta Vista, Excite). Le moteur choisi comme exemple est cependant celui de Yahoo!. C'est en ce point que l'exposé s'enrichit légèrement par rapport au petit livre de H. Lilen. Tout d'abord,

l'auteur introduit la distinction entre la recherche par annuaire (*directory*), et la recherche à l'aide des moteurs de recherche proprement dit ²⁷.

Most search engines have a directory to allow you to locate the information you require by working through the various topics. However, if you know what you're looking for it's usually quicker to *search*. In Yahoo! you can specify the keywords and phrases in the **Search the Web:** field.

À propos de cette recherche, l'exposé aborde en passant la *syntaxe des requêtes*. Le premier exemple de requête est ainsi **+Edinburgh +Restaurant -Chinese**. L'auteur explicite la signification des signes + et - en ces termes : "a + (plus) means that the page must contain the word, a - (minus) means that the page must not contain the word". Cette introduction, réalisée à propos de l'emploi du dispositif de recherche d'Internet Explorer, est poursuivie à propos des moteurs de recherche en général et de celui de Yahoo! en particulier. Dans ce dernier cas, la requête prise pour exemple est la suivante : **"short break" +Speyside -camping +skiing**. En l'espèce, le moteur de recherche consulté fournit à l'auteure 18 résultats seulement. Là se termine l'initiation à la recherche d'information sur le Web que l'ouvrage en question propose à son lecteur.

CHAPITRE 8

Un peu au-delà des rudiments

Nous examinerons ici un troisième ouvrage. Intitulé *Guide Bilingue de la Communication Électronique* ou *Electronic Communications Bilingual Guide*, il se situe en quelque sorte, au plan linguistique, sur la frontière entre le français et l'anglais. Publié en 2003 dans la collection « Langues pour tous » chez Pocket, il est dû à deux auteurs spécialisés, James Benenson, ingénieur d'études et enseignant d'anglais à Télécom-Paris, et Brigitte Juanals, docteur en information-communication et maître de conférences à l'Université Charles-de-Gaulle Lille 3. Le titre complet de l'ouvrage est, en vérité, « Guide bilingue de la communication électronique *en 40 leçons* ». Pour donner une idée de l'ampleur du champ couvert par l'ouvrage, qui ne comporte pourtant que 320 pages, nous avons reproduit ci-après la table des matières.

SOMMAIRE • CONTENTS

/// Présentation • *Presentation* / Abréviations, symboles, orthographe • **Abbreviations, symbols, spelling** /// **1. INTRODUCTION • INTRODUCTION** / 1. Le courrier électronique ; éléments de base • **E-mail; basics** / 2. Le courrier électronique ; fonctions avancées • **E-mail; advanced features** / 3. La netiquette, choses à faire et à ne pas faire • **Netiquette, Do's and don'ts** / 4. De la lettre traditionnelle au bavardage • **From traditional letter-writing to on-line chatting** /// **2. SITUATIONS • SITUATIONS** // VIE SOCIALE ET RENCONTRES • **SOCIAL LIFE & ENCOUNTERS** / 5. Appels téléphoniques en ligne • **Telephoning via the Internet** / 6. Participer à une visioconférence (1) • **Participating in a videoconference (1)** / 7. Participer à une visioconférence (2) • **Participating in a videoconference (2)** / 8. Magazines électroniques • **E-zines** / 9. Bulletins d'information • **Newsletters** / 10. Carnets Web • **Weblogs** / 11. Lexique et ressources des carnets Web • **Weblog lexicon and resources** / 12. Forums et listes de diffusion • **Newsgroups, web forums and mailing lists** / 13. Ecriture collective, écriture interactive, écriture collaborative • **Collective writing, interactive writing, collaborative writing** / 14. Communautés d'intérêt en ligne (1) • **Online communities of interest (1)** / 15. Communautés d'intérêt en ligne

(2) • **Online communities of interest (2)** / 16. Séances de bavardage, messagerie instantanée et SMS • **Chat sessions, instant messaging and SMS'ing** / 17. Ecrire une requête avec un moteur de recherche (1) • **Writing a query with a search engine (1)** / 18. Ecrire une requête avec un moteur de recherche (2) • **Writing a query with a search engine (2)** / 19. Jouer en ligne • **Playing online** / 20. Canulars et légendes • **Hoaxes and urban legends** / 21. Arnaques et escroqueries • **Scams and rip-offs** // ETUDES UNIVERSITAIRES ET FORMATION CONTINUE • **ACADEMIC STUDIES AND CONTINUING EDUCATION** / 22. Présentation des formations et diplômes • **Presentation of studies and degrees** / 23. L'apprentissage de l'anglais par l'Internet (1) • **Learning English through the Internet (1)** / 24. L'apprentissage de l'anglais par l'Internet (2) • **Learning English through the Internet (2)** / 25. Participer à un congrès commercial • **Attending a trade convention** / 26. Participer à un congrès scientifique • **Attending a scientific congress** / 27. Écrire une publication scientifique • **Writing a scientific paper** // SITUATIONS COMMERCIALES ET PROFESSIONNELLES • **E-COMMERCE AND PROFESSIONAL SITUATIONS** / 28. Recherche d'emploi • **Looking for a job** / 28. Achats en ligne (chercher des produits et des prix) • **Online buying (searching for goods and prices)** / 30. Achats en ligne (confirmation, modification, plainte) • **Online buying (confirmation, modification, complaint)** / 31. Enchères et achats groupés • **Online auctions and group-buying** / 32. Demandes d'informations & réservations • **Inquiries and bookings** / 33. Réunions • **Meetings** / 34. Offres commerciales • **Business offers** / 35. Correspondance commerciale (propositions) • **Business correspondence (proposals)** / 36. Langues et cultures en ligne • **Languages and cultures online** /// **TOOLS • OUTILS** / 37. Acronymes • **Three-letter acronyms** / 38. Abréviations courantes de bavardage • **Common chat abbreviations** / 39. Binettes • **Smileys** / 40. Extensions de fichiers • **File extensions** /// Index

Nous concentrerons notre enquête sur les leçons 17 et 18, en lesquelles est scindé le traitement du thème « Écrire une requête avec un moteur de recherche » (*Writing a query with a search engine*).

Préparer une requête

Le thème en question occupe 14 pages pleines, sans copie d'écran ou autres images. À l'exception des titres – et d'un vocabulaire bilingue français-anglais qui occupe les deux dernières pages –, le texte est entièrement en français. On peut s'attendre à ce qu'il condense une version d'un texte du savoir plus développée que celles déjà rencontrées. Le découpage adopté par les auteurs fait apparaître quatre volets, désignés par les lettres A, B, C, D :

A. *Preparing a query*, « Préparer une recherche » (pp. 132-133) ; B. *Writing a query*, « Écrire une requête » (pp. 133-136) ; C. *What sort of tool would you choose?*, « Quel type d'outil choisir ? » (pp. 137-143) ; D. *Vocabulary*, « Vocabulaire » (pp. 144-145). Le classement des volets A, B et C en fonction de la surface occupée est le suivant : C > B > A. Ce simple constat quantitatif suggère une première remarque : dans la recherche d'informations sur Internet, le choix du bon outil, ou plutôt du bon outillage, semble s'imposer comme une question importante, du moins dans une certaine version du texte du savoir en émergence²⁸.

La distinction entre les volets A et B (Préparer une requête / Écrire une requête) est vertueuse mais délicate. L'idée sous-jacente est que, pour rendre « plus rapide et plus efficace » les recherches que l'on entreprend, il convient de les « préparer », d'y réfléchir à l'avance, notamment pour sélectionner les mots clés constituant les requêtes « atomiques » possibles²⁹. Une liste de tels mots clés doit être constituée avec l'idée de cerner assez nettement la question sur laquelle porte la recherche. Elle doit en outre être enrichie de mots anglais et de synonymes, tout en évitant « les termes polysémiques », ou encore les sigles et autres abréviations, écartés comme générateurs de « bruit ».

Selon le Dictionnaire bilingue Internet & multimédia (que nous examinerons un peu plus loin : il est également signé des auteurs dont nous commentons le texte ici), la notion de bruit (noise) « fait référence aux informations parasites dans un fichier d'images ou de son, ou à la présence de résultats non pertinents dans les résultats d'une recherche informatisée ». Le « silence » (silence) « fait référence à l'absence de résultats pertinents dans les résultats d'une recherche informatisée ». Selon les mêmes auteurs, on emploie aussi, à ce propos, l'expression an empty search, « recherche vide ou infructueuse ».

On prendra garde encore de respecter l'orthographe, car les erreurs orthographiques faussent les résultats, et on préférera choisir des *noms* plutôt que des verbes ou des adjectifs. Notons ici que ces éléments, constitutifs d'une technique livrée « en kit », sont inégalement justifiés : la technologie de la technique n'est pas uniformément développée. Ainsi est-il conseillé de choisir *aussi* des mots clés en anglais, « car, sur la toile, plus de 80 % des pages sont en anglais et seulement un peu plus de 2 % sont en français ». De même, il faut ajouter des synonymes, et cela « pour augmenter la quantité de réponses pertinentes potentielles ». Semblablement, il faudra porter attention à *l'ordre des mots* de la requête car, « dans beaucoup de moteurs de recherche », « la valeur de pertinence est maximale pour le premier terme, puis diminue ». En revanche, l'exposé n'éclaire pas le lecteur sur les raisons pour

lesquelles il faudrait préférer les noms aux verbes et aux adjectifs, ou pourquoi les sigles doivent être regardés comme générateurs de bruit. D'une façon générale, ce qui est avancé comme motivant tel ou tel geste technique paraît insuffisamment argumenté, comme s'il existait quelque part des savoirs non explicités par les auteurs qui fonderaient pourtant leurs assertions.

*Par contraste, il serait facile par exemple d'attirer l'attention du lecteur sur ce simple fait, aisément contrôlable : si une requête est composée de deux mots M_1 et M_2 , la requête qu'on peut écrire $+M_1 + M_2$ ne conduira pas en général aux mêmes résultats, rangés dans le même ordre, que la requête $+M_2 + M_1$. Le lecteur pourra en faire l'épreuve en tapant par exemple dans la boîte de recherche de Google les quatre requêtes suivantes : **Barbara Cassin**, **Cassin Barbara**, **"Barbara Cassin"**, **"Cassin Barbara"** : le 15 décembre 2007, vers 18 heures, le nombre des résultats affiché par Google était respectivement de 154 000, 153 000, 34 000, 846. Ce résultat, plus faible sans doute que celui avancé par les auteurs, est cependant de nature à susciter des gestes techniques utiles sans pour cela s'appuyer sur un supposé savoir dont la source n'est en tout cas pas révélée dans le texte que nous examinons.*

Écrire une requête : les opérateurs booléens

Le volet A comporte d'autres indications encore, qui se rapportent à des fonctionnalités plus avancées des moteurs de recherche. Ainsi la recherche demandée peut-elle éventuellement se centrer sur des images, des vidéos, des groupes de discussion, etc. Elle peut aussi se faire *par champs*, ou recourir à des requêtes « moléculaires ». C'est à de tels développements que va être consacré le volet B du dossier, « Écrire une requête » (*Writing a query*). Cette partie s'ouvre par une motivation générale qui mérite d'être soulignée : selon les auteurs, en effet, « lorsque vous écrivez une requête, vous avez intérêt à être le plus précis possible pour obtenir des résultats en quantité limitée et de ce fait exploitables ». On a là une assertion technologique qui va déterminer certains gestes techniques que les auteurs vont ensuite expliciter. Avant d'y venir, soulignons un fait qui n'est pas rare dans le travail de constitution d'un discours technologique dans le domaine qui nous occupe : ici, le précepte avancé est présenté comme allant de soi, alors que, précisément, la construction d'une technique supposerait tout un travail technologique, voire technologico-théorique, qui « ouvre les perspectives » : n'avoir par exemple que 18 résultats est déjà trop pour beaucoup d'utilisateurs (qui en laissent un certain nombre inexploités), alors qu'on peut se montrer actif devant des

milliers de résultats (par exemple en accomplissant des gestes d'échantillonnage tels que ceux permis par le logiciel Shmoogler). On notera ici – nous y reviendrons – qu'il est relativement rare dans les textes sur la recherche d'informations sur le Web, que les auteurs fassent une place spéciale, et significative, à l'apprentissage du travail *de lecture des résultats affichés* par un moteur de recherche pour en apprécier – avant même de cliquer sur quelque lien que ce soit – la pertinence *a priori*, estimée sur la base des multiples indices que peut receler le résultat affiché à l'écran.

Pour illustrer cette remarque, on peut par exemple envisager le scénario de travail suivant : en une première étape, mis devant plusieurs pages de résultats, l'utilisateur sélectionne a priori un certain nombre de liens ; sa session de travail devant s'arrêter rapidement, il copie les liens en question et les colle dans un fichier de texte afin de les examiner à loisir lors d'une prochaine session de travail. La deuxième étape est évidemment classique en la matière : elle consiste en l'examen de pages Web supposées « intéressantes », examen qui, bien entendu, soulève un ensemble de problèmes. La première étape, en revanche, ne l'est guère : elle mobilise des praxéologies spécifiquement développées, qui appellent donc une élaboration systématique de la part de l'utilisateur.

L'introduction du volet B s'appuie sur le précepte technologique commenté plus haut pour justifier notamment d'y présenter les *combinaisons de termes* possibles pour la constitution de requêtes « moléculaires » à partir de requêtes « atomiques ». Selon l'usage, le nom de George Boole est mentionné, avant qu'on ne passe en revue les opérateurs OR, AND, AND NOT, NEAR, à quoi s'ajoutent encore, outre bien sûr l'usage des parenthèses, l'opérateur de troncature (symbolisé par le signe *), et l'opérateur d'adjacence (ADJ). De ces opérateurs « booléens », divers exemples sont proposés, que nous reproduisons d'abord sans autre commentaire.

("mc luhan" NEAR marshall) AND (media*) / magicien AND oz AND film / cake* OR desert* / gateau* OR dessert* / car* OR kitten* / chat* OR chaton* / "jean renoir" AND NOT url:com / voiture AND (bmw OR fond) / fruit AND (banane OR pomme) / franc* / "Charlie Chaplin" / Charlie ADJ Chaplin / "catastrophe naturelle" / (jam OR jelly) NEAR (raspberry) / (confiture OR gelee) NEAR (framboise)

Voici maintenant les définitions et les commentaires que donnent les auteurs pour chacun des opérateurs mentionnés qu'ils présentent à leurs lecteurs (pp. 134-135).

AND (ET)

Indique que la recherche se fera obligatoirement sur les 2 mots saisis. Si un document contient uniquement un de ces 2 mots, il ne sera pas pris en compte.

OR (OU)

Indique que la recherche prendra en compte soit l'un des mots saisis, soit l'autre, soit les deux. Cet opérateur est souvent celui pris en compte par défaut, si l'on ne saisit rien entre les mots tapés, sur un grand nombre de moteurs de recherche.

AND NOT (ET NON, SAUF)

Permet d'exclure toutes les pages qui contiendraient un mot non désiré. Le moteur cherchera les pages contenant le 1er mot mais ne prendra pas en compte celles contenant aussi le 2e mot exclu.

parenthèses

Permet de coupler plusieurs opérateurs booléens en utilisant des parenthèses. La recherche portera sur les pages contenant à la fois les deux premiers mots (et non pas l'un ou l'autre), soit sur les pages contenant le dernier mot, soit les deux ensembles.

troncature

Appelés *jokers* ou *wildcards*, et le plus souvent symbolisés par le signe *. Ils permettent d'effectuer des troncatures de mots, c'est-à-dire de ne saisir qu'une partie du terme désiré. La troncature ne peut souvent se faire qu'en fin de mot, mais certains moteurs acceptent une troncature à l'intérieur du mot.

adjacence

Possibilité de rechercher 2 mots adjacents. Cette adjacence de mots est souvent imposée par des guillemets ou par l'opérateur ADJ. Le moteur recherche les pages contenant les mots tels qu'ils sont saisis, et ne prendra pas en compte celles contenant les 2 mots placés dans un autre ordre.

proximité : NEAR (PROCHE DE)

Possibilité de trouver des mots qui sont proches les uns des autres dans l'espace d'une page, sans pour autant être adjacents. Le moteur recherchera les pages qui contiennent les 2 termes même s'ils sont distants d'un nombre maximum de mots spécifiés et sans contrainte d'ordre. Selon les moteurs, cette fenêtre de recherche peut aller de 10 à 50 mots, voire plus.

On aura noté l'usage, dans ce qui précède, de la notion de *joker* ou *wildcard*, mots que ces auteurs, dans leur *Dictionnaire bilingue* déjà cité, rendent par « caractère générique, caractère de remplacement, caractère joker, joker » ou encore « caractère de substitution ».

Ils en proposent exactement la définition suivante : « Caractère de substitution susceptible de représenter un ou plusieurs caractères, soit dans une recherche d'informations dans une base de données ou sur le Web, soit dans une recherche de nom de fichier dans un système d'exploitation. » À la notion de wildcard character sont associées notamment les notions de wildcard search, « recherche avec un caractère de remplacement », et de truncation, troncature (de mots). Notons que le mot wildcard ainsi usité en informatique n'y est en rien une création ex nihilo, comme le montre la notice que consacre à ce mot The American Heritage Dictionary of the English Language (Picket et al., 2000) : “1. Games A playing card whose value can vary as determined by its holder. 2. Sports An athlete or team selected to compete in a tournament or playoff from among those who did not qualify by meeting the regular requirements. 3. Computer Science A symbol that stands for one or more unspecified characters, used especially in searching text and in selecting multiple files or directories. 4. Slang An unpredictable or unforeseeable factor: A surprise witness proved to be the wild card at the trial.”

La recherche par champ

La section suivante est intitulée « Affiner une requête avec la recherche par champs » (expression que les auteurs écrivent aussi « Recherche par champ » dans le *Vocabulaire* qui clôt le dossier, où ils donnent l'expression anglaise correspondante : “*field search*”). Comme à propos des opérateurs booléens, les auteurs signalent que les indications concernant les différents champs auxquels restreindre une recherche « varient en fonction des outils de recherche ». Nous avons reproduit ci-après l'ensemble des indications proposées.

title: title:education

Recherche sur les titres des sites uniquement.

host: host:afp

La fonction **host:** permet de limiter une recherche sur le nom du serveur.

domain: domain:fr, domain:edu, domain:net.

La fonction **domain:** permet de mener une recherche sur les noms de domaines.

+ pour une définition des noms de domaine, voir pp. 19 et 20.

link: link:www.computers.net

Avec la fonction **link:** vous menez une recherche sur l'adresse des liens contenus dans une page, c'est-à-dire que vous trouverez les pages qui contiennent le lien indiqué.

url: url:mit

La fonction **url:** permet de faire une recherche sur les adresses URL complètes qui contiennent le mot ou l'expression indiqués. pour une définition d'URL, voir glossaire alphabétique en fin d'ouvrage.

anchor: anchor:liberation

La fonction **anchor:** permet une recherche du mot ou de l'expression faisant partie d'un lien hypertexte, pour connaître les pages qui pointent sur un site donné. La recherche s'effectue dans le texte des liens.

Les filtres de mots (contient/ne contient pas), les filtres par langues, les filtres par date, les filtres par types de fichiers, sont aussi des moyens très utiles pour restreindre une requête.

text: text:kubrick

Recherche des pages comportant le texte saisi à n'importe quel endroit, à l'exception des images, liens ou URL.

image: image:"citizen kane", image:sunset*

Recherche sur les noms d'images.

Notons que ces indications sont reprises presque mot pour mot dans le *Dictionnaire bilingue...* (Benenson & Juanals, 2004, p. 159), où cet exemple est donné : « *Field search is used to focus on particular parts of Web pages.* La recherche par champ est utilisée pour cibler une zone spécifique de pages Web. » En vérité, les explicitations proposées ne permettent guère au lecteur qui n'aurait aucune autre source d'information sur le sujet d'en tirer profit en pratique. Le texte examiné apparaît ici comme un *leurre didactique*, qui apporte des informations techniques dépourvues d'opérationnalité immédiate, ne serait-ce que parce que leur réalisation concrète dépend du moteur de recherche sur lequel on s'efforcerait de les mettre en œuvre. De ce point de vue, le *topos* du lecteur, étudiant potentiel, apparaît comme un trompe-l'œil : la perspective qui semble s'ouvrir à lui n'est qu'une illusion suggestive.

Outils de la recherche : les annuaires

Le troisième volet du dossier, on l'a noté, est le plus fourni : intitulé « Quel type d'outil choisir ? » (*What sort of tool would you choose?*), il est développé sur plus de six pages.

L'introduction au thème traité distingue d'abord *répertoires*, *moteurs* et *méta-moteurs*, avant de proposer un tableau plus diversifié de la panoplie des outils de recherche.

Les outils de recherche sont présents en nombre croissant sur l'Internet et sont de plusieurs types : moteurs ou annuaires généralistes ou spécialisés (par thèmes, secteurs d'activité, régions ou pays...), versions internationales ou localisées... La mention « portail » indique que les outils proposent aussi toute une série de services, comme des actualités, la météo, des annuaires, un courrier électronique gratuit ou des pages d'accueil personnalisées. Un nombre croissant d'outils proposent aujourd'hui un accès conjoint de type répertoire et de type moteur que vous pouvez désormais combiner ou utiliser séparément. (p. 137)

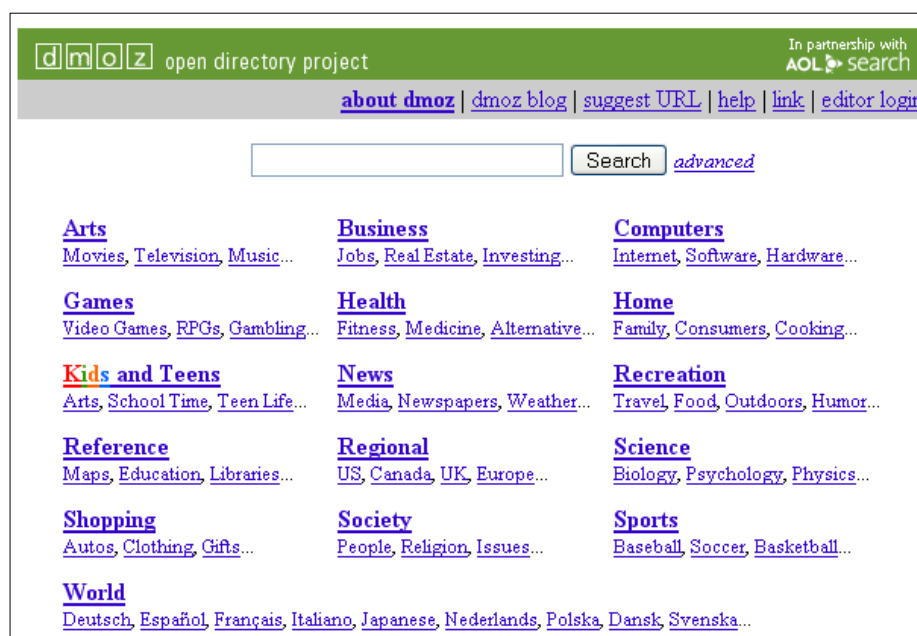
Comment s'orienter en un tel foisonnement ? Les auteurs distinguent à cet égard deux grands types de recherche : celles portant « sur un sujet plutôt large d'intérêt général », d'une part, et celles portant « sur un sujet pointu », d'autre part. Les recherches du premier type utiliseront de préférence des *répertoires*, « comme les annuaires thématiques généraux, des catalogues thématiques, des pages de départ spécialisées », et cela afin d'identifier des sites susceptibles d'apporter « une information plutôt générale » à propos du thème de recherche. Le texte examiné fait place alors à un développement sur la notion de répertoire qui s'appuie notamment sur la distinction entre répertoire et *index*.

Les répertoires sont constitués d'une base de données de liens organisés de façon arborescente et hiérarchique et renvoyant à d'autres sites : des domaines sont eux-mêmes subdivisés en plusieurs sous-domaines, du plus général au plus précis. La collecte de l'information est manuelle, ce qui implique une validation et une classification (par thème, lieu, discipline...) humaine des ressources. La conséquence est une couverture beaucoup moins large qu'un index, mais l'intervention humaine accroît la pertinence de l'information retenue par rapport à une indexation automatique de moteur. Une recherche par index a souvent été ajoutée à la structure hiérarchique, car les répertoires ont multiplié les catégories (Yahoo! compte par exemple plus de 25 000 catégories), rendant le repérage des ressources souvent difficile. (pp. 137-138).

Après une présentation « phénoménologique » des annuaires – c'est-à-dire une description de ce que l'utilisateur va rencontrer (voir les copies d'écran ci-après) –, on voit les auteurs expliciter les éléments d'une certaine *orthodoxie technologique* à propos de la distinction annuaires/moteurs – les annuaires étant censés accroître « la pertinence de l'information

retenue par rapport à l'indexation automatique d'un moteur ». Ce discours, plutôt favorable aux annuaires, ne paraît pas argumenté de façon à convaincre.

L'avantage reconnu aux annuaires semble revêtir l'allure d'un dogme, c'est-à-dire d'une assertion qui va de soi et qu'on n'interroge donc pas davantage. Les trois copies d'écran ci-après illustrent une recherche par annuaire utilisant l'un de ceux mentionnés par les auteurs, DMOZ – ce nom, contraction de Directory Mozilla, désigne l'annuaire résultant de “The Open Directory Project”, lequel se présente comme “the largest, most comprehensive human-edited directory of the Web”, “constructed and maintained by a vast, global community of volunteer editors” (<http://www.dmoz.org/about.html>). Sur la première page, on a choisi d'abord le lien Science.



Sur la page à laquelle celui-ci conduit (ci-après) on a choisi alors le lien Search Engines, lequel apparaît accompagné de la promesse – plutôt décevante – de renvoyer à 7 résultats. On aboutit alors à la troisième page, où se trouvent lesdites 7 références : celles-ci, qui résultent d'un choix parmi les propositions de collaborateurs bénévoles de l'annuaire, peuvent apparaître un rien arbitraire à l'œil habitué aux nombreuses pages de résultats des moteurs de recherche. À l'inverse, ce « florilège » peut renforcer l'idée qu'un certain discours orthodoxe tend à valoriser, en apparaissant à l'utilisateur comme l'expression supposée d'un choix fortement assumé, impliquant la mise à l'écart délibérée de la majorité des pages du Web ayant à voir avec le thème retenu ici.

<ul style="list-style-type: none"> • Academic Departments@ (3) • Chats and Forums (17) • Directories (26) • Educational Resources (366) • Employment (111) • Events (35) • Institutions (743) 	<ul style="list-style-type: none"> • Instruments and Supplies (2,056) • Methods and Techniques (187) • News and Media (165) • Publications (210) • Reference (438) • Search Engines (7) • Software (182)
<ul style="list-style-type: none"> • History of Science@ (438) • Museums@ (553) • Women@ (186) 	
<ul style="list-style-type: none"> • Agriculture (3,248) • Anomalies and Alternative Science (480) • Astronomy (4,048) • Biology (25,607) • Chemistry (4,446) • Computer Science@ (2,218) • Earth Sciences (5,724) 	<ul style="list-style-type: none"> • Environment (6,285) • Math (11,189) • Physics (4,690) • Science in Society (723) • Social Sciences (22,916) • Technology (12,190)

- [Article Sciences](#) - Scientific article search and order engine, a service provided by Institut de l'Information Scientifique et Technique (INIST), National de la Recherche Scientifique (CNRS), France.
- [CiteSeer](#) - Search engine for computer science and related matters. For all results number of citations are given for each paper in a given context. Most of the papers are cached and available for direct download as postscript or pdf.
- [ISI Highly Cited](#) - A search tool to identify individuals, departments and laboratories that have made fundamental contributions to science and technology in recent decades.
- [iSpecies](#) - A species search engine from Rod Page of Glasgow University. Search on either the common or scientific name of a species or querying other data sources.
- [Online Journals Search Engine](#) - A multi-search tool that permits queries of a number of scientific journals.
- [Scirus](#) - Search engine that fetches only scientific websites.
- [SciSeek](#) - Science search engine and directory. Also provides forums, tools and related news.

La recherche par index évoquée ci-dessus concerne l'indexation des ressources de l'annuaire seulement, et non les ressources du Web ou d'une vaste partie de celui-ci. Dans leur *Dictionnaire*, les auteurs que nous suivons définissent un index comme une « liste organisée d'éléments accompagnée du moyen d'y accéder », avant de donner cet exemple d'emploi du mot : « “A search engine's index is generated automatically”. L'index d'un moteur de recherche est créé de façon automatique. » On ne peut pas ne pas voir dans le discours technologique orthodoxe tel qu'il affleure ici un élément *théorique* essentiel, l'opposition entre humains et machines, concrétisée dans l'opposition du choix par un opérateur humain et de l'indexation par un robot. Cet élément théorique, en surplomb par rapport à l'orthodoxie technologique rencontrée, est évidemment contestable, non seulement parce que les critères appliqués dans l'indexation automatique ont été définis par des humains, mais encore parce que, entre l'index *stricto sensu* et les résultats affichés, il existe tout un traitement également défini par des opérateurs humains.

La suite de l'exposé présente sans commentaires les noms et adresses d'un petit nombre de répertoires (*About, DMOZ, La Toile du Québec, Looksmart, Magellan, NBC, Nomade, Yahoo!*). Au moment où nous écrivons, quelques-uns d'entre eux ont vu leur état changer à tel ou tel point de vue : la page d'accueil de Looksmart semble indiquer que ses fonctions d'annuaire sont en voie d'extinction ; l'adresse donnée pour Magellan n'aboutit à rien (il semble que cet annuaire ait été racheté par Excite) ; NBC apparaît plus comme un portail commercial qu'un annuaire ; Nomade est aujourd'hui propriété d'Alice (Groupe Télécom Italia) ; Yahoo!, enfin, propose toujours un annuaire, mais l'accès à celui-ci est bien dissimulé quand on prend pour point de départ la page d'accueil (on le trouvera, pour ce qui est de la version anglaise par exemple, à l'adresse <http://search.yahoo.com/dir>).

Les moteurs de recherche face aux annuaires

Le volet C, *Quel type d'outil choisir ?*, se poursuit par une section intitulée « Pour rechercher une information sur un sujet pointu ». C'est désigner là, en l'espèce, le recours aux moteurs de recherche. Le discours se fait ici subtilement complexe ; le voici dans son intégralité.

Les moteurs de recherche, généralistes ou spécialisés, permettent de mener une recherche extrêmement fine sur des sites ou des documents spécifiques. Mais attention, l'efficacité des moteurs est liée à la capacité de l'internaute à formuler des requêtes adaptées.

Les moteurs sont des bases de données résultant de l'indexation automatisée préalable de pages Web par des robots logiciels d'exploration. Ces ressources sont consultables via une interface d'interrogation au moyen de mots clés, en vue d'obtenir une liste de pages contenant ces mots clés. Les moteurs sont plus ou moins puissants en fonction du nombre de pages indexées, de leur capacité à relier les mots clés ou à lancer des recherches dans des zones précises des pages pré-indexées. Chaque moteur possède un mode d'indexation particulier (titre de l'URL, en-tête, premières lignes ou intégralité des textes d'un document...) qu'il est utile de connaître pour mieux formuler vos requêtes. (p. 138).

La rhétorique de ce passage paraît marquer une attitude ambivalente à l'endroit des moteurs de recherche. D'un côté, les moteurs de recherche permettraient de mener « une recherche extrêmement fine », à propos de ressources « spécifiques ». D'un autre côté, toutefois, leur bon usage supposerait une compétence difficile à acquérir concernant la bonne formulation des requêtes. Le second alinéa du développement reproduit plus haut, qui apporte une information en principe neutre sur les principes structurels et fonctionnels des moteurs de

recherche, peut aussi être pris comme une mise en garde – « un moteur de recherche, c’est aussi compliqué que ça, c’est aussi peu “naturel” que ça », pourrait-on entendre à travers les propos des auteurs. La fin du paragraphe formule à demi-mot le dilemme face auquel l’ouvrage abandonne son lecteur : outre qu’un moteur de recherche est chose complexe, chacun d’eux a ses particularités, qu’il faut connaître, ce qui ne va pas de soi. Ces considérations sont suivies d’une liste d’exemples de moteurs de recherche, avec les adresses correspondantes.

Certains des moteurs de cette liste ont connu des évolutions importantes : Alta Vista a été racheté d’abord par la firme Overture puis par Yahoo! et fonctionne actuellement avec la technologie de recherche de Yahoo! (Yahoo! Search Technology). Hotbot a été racheté par Lycos et ne représente aujourd’hui plus qu’une interface accédant aux technologies des outils Ask, MSN et Lycos. Le moteur de recherche Lokace ne fonctionne plus depuis 2001 (voir <http://c.asselin.free.fr/french/moteursmorts.htm>). La page d’accueil de Northern Light (<http://www.northernlight.com/>) semble indiquer une évolution des activités principalement vers des outils d’entreprises. Le moteur Webtop est arrêté depuis 2001 (voir <http://actu.abondance.com/actu0146.html#mardi>).

Les trois sections suivantes du volet C reviennent alors de façon plus large sur les questions examinées jusqu’ici : une première section est intitulée « Conseils d’utilisation des répertoires et moteurs » tandis que les deux sections suivantes ont pour titre, respectivement, « Pour établir un panorama le plus exhaustif possible » et « Pour cibler une recherche sur un thème précis ». La première de ces trois sections conseille surtout de multiplier et de combiner les consultations de répertoires et de moteurs. Ce type de geste technique trouve son fondement technologique, selon les auteurs, dans le fait que 1) les moteurs « n’indexent pas forcément les mêmes pages », et 2) les options de recherche avancée « sont plus ou moins étendues et varient suivant les moteurs ». Le point de vue technico-technologique proposé trouve en outre un écho, ajoutent alors les auteurs, dans la tendance « au couplage des répertoires et des moteurs », observation que concrétisent à l’époque (2003) quelques exemples dont certains sont aujourd’hui obsolètes.

C’est ainsi par exemple que le guide thématique de Voilà n’existe plus depuis 2007 (voir <http://c.voila.fr/Errors/terminated.html>) et que Yahoo! utilise depuis 2004 sa propre technologie de recherche et n’a plus recours aux services de recherche de Google. Le lien proposé par les auteurs vers le répertoire d’Altavista conduit, dans sa

version anglaise, vers le répertoire de Yahoo! et, dans sa version française, vers une page inconnue sur le site de Yahoo!.

En conclusion, pour « améliorer de façon conséquente la pertinence des résultats », le conseil est donné de « choisir d'abord une catégorie et de taper ensuite des mots-clés dans le cadre de cette catégorie ». Cette proposition technique semble avoir toutes les vertus d'une *formation de compromis* au sens freudien du terme (*Kompromissbildung*) : à travers elle, la puissance des moteurs de recherche est discrètement assumée, sans pour autant que cela renvoie au néant la vénérable tradition documentaire dont les « répertoires » sont le rejeton.

Méta-moteurs et « moteurs de moteurs »

Dans la section « Pour établir un panorama le plus exhaustif possible », le mot d'ordre de consultation plurielle se concrétise par la référence à l'emploi de méta-moteurs, « moteurs de recherche par mots clés, qui proposent une interface de recherche intermédiaire commune à plusieurs moteurs de recherche ». Le recours à un méta-moteur permet, selon le texte que nous suivons, « un gain de temps très appréciable », du fait de la consultation simultanée de plusieurs bases de données. Le revers de la médaille, c'est qu'il n'est plus possible ici de dépasser le niveau de la requête simple, faute d'une syntaxe commune à tous les moteurs consultés par le méta-moteur. La présentation des avantages et des inconvénients des méta-moteurs est complétée, selon l'usage, par une courte liste de méta-moteurs, introduite par cette précision : « Tous les outils retenus ci-après comportent une interface de saisie unique pour formuler vos requêtes à destination de moteurs différents, lancent une recherche simultanée sur plusieurs moteurs et fournissent une liste de résultats synthétisée unique... »

Parmi les méta-moteurs mentionnés, Copernic (<http://copernic.com>), Metacrawler (<http://www.metacrawler.com>) et Search.com (<http://www.search.com>), sont tous trois opérationnels à ce jour. L'outil inFind (<http://www.infind.com>), en revanche, semble être arrêté, comme l'indiquait Danny Sullivan le 8 mars 2001 sur Searchenginewatch.com et comme le confirme la page d'« erreur 404 » à laquelle conduit son nom de domaine. Profusion (<http://www.profusion.com>) informe sur sa page d'accueil qu'il est arrêté définitivement (“The ProFusion.com search engine is down and will not be in operation any longer.”) On notera que, à la date de publication de l'ouvrage que nous suivons (2003), certaines des indications données par les auteurs étaient déjà périmées et auraient dû être écartées.

On aura remarqué que, dans la présentation des méta-moteurs, le thème des annuaires disparaît. Ce thème va pourtant réapparaître dans la troisième des sections annoncées, « Pour cibler une recherche sur un thème précis ». Dans ce cadre, en effet, les auteurs proposent trois « manœuvres » techniques, dont la première consiste précisément à « mener des recherches sur un ou plusieurs répertoires », en se rapprochant le plus possible, à travers les catégories proposées, du thème qui est la cible de la recherche. Une deuxième stratégie consiste à identifier un moteur de recherche ou un répertoire spécialisé sur le thème en question (ou l'incluant dans son domaine de spécialité). Pour cela, nous disent les auteurs, il convient d'utiliser « un moteur de moteur (*sic*) de recherche ou un répertoire de répertoire (*sic*) de recherche (aussi appelé un méta-répertoire ou page de départ spécialisée) ». Aucune information ne vient en ce point illustrer la référence aux notions de « moteur de moteurs » et de « répertoire de répertoires ».

Bases de données : errances et sérendipité ?

Une autre direction en vue de « cibler une recherche » consiste à rechercher d'abord des *bases de données* (*databases*) appropriées, dont un grand nombre « ne sont pas accessibles par les outils de recherche habituels ». Les auteurs fournissent à ce propos l'adresse d'un certain nombre de sites donnant accès à des bases de données. La première adresse est celle du site de la Bibliothèque Nationale de France (<http://www.bnf.fr>), sur lequel, précisent les auteurs, l'utilisateur pourra trouver les catalogues de la BNF, des listes de catalogues des bibliothèques françaises, une sélection de catalogues étrangers ou encore une sélection de méta-moteurs dans les catalogues de bibliothèques.

Cette rapide description ne rend sans doute pas justice à la profusion de possibilités offertes sur le site de la BNF : c'est ainsi que, au moment où nous écrivons, en empruntant le parcours qui, démarrant à la page d'accueil du site de la BNF, passe successivement par « Les signets de la BNF », « Outils internet », « Métamoteurs », l'utilisateur découvrira l'existence du site Agentland, dont la visite lui fera notamment rencontrer la notion d'agent intelligent. En revanche, nous n'avons pas réussi à identifier clairement la « sélection de métamoteurs dans les catalogues de bibliothèques » : les seuls types de sélections d'outils de recherche qu'on y trouve sont d'une part des listes d'annuaires, de moteurs de recherche et de méta-moteurs, d'autre part des listes d'outils d'interrogation de catalogues, tel le Catalogue Collectif de

France, présenté comme permettant de « localiser plus de 20 millions de documents (imprimés, manuscrits, sonores, multimédias, électroniques, numérisés, des images fixes, des images animées) » (<http://ccfr.bnf.fr/portailccfr/servlet/LoginServlet>). Signalons encore la rubrique « Catalogues collectifs », que l'on trouve à l'adresse suivante : http://signets.bnf.fr/html/categories/c_017collectif_nat.html.

La seconde référence proposée est un peu sibylline : « *Lycos, Searchable databases*, sélection de bases de données... » Une recherche à propos de **Lycos** – dénomination qui désigne aujourd'hui plutôt un portail qu'un moteur de recherche ou un annuaire – et de l'expression "**searchable databases**" conduirait l'utilisateur à une plus ou moins longue errance à travers le Web, situation typique où une information peu précise du double point de vue formel (quelle adresse ?) et réel (que sont ces *Searchable databases* dont on semble nous parler ?) conduit à une culture du vagabondage, dans laquelle les parcours à travers le Web semblent d'abord motivés, apparaissent ensuite démotivés, et sont enfin immotivés. On peut voir en cela une technique de « recherche » dont l'objet n'est pas fixé, et dont la technologie la plus répandue aujourd'hui repose sur la notion « vedette » de *sérendipité*.

Par delà l'origine historique de la notion, il n'est pas difficile de s'assurer que, comme à d'autres propos, sa source de diffusion principale se trouve dans le monde anglophone : la requête "serendipity" adressée à Google le 30 décembre 2007 vers 8 h ramène approximativement 11 000 000 résultats, tandis que "sérendipité" n'en suscite que 6420, et "sérendipidité" (traduction parfois rencontrée, due semble-t-il à Edgar Morin, 1999), 612 seulement. Notons que le premier des quelque 11 millions de résultats affichés par Google correspond à une boutique parisienne portant ce nom et faisant commerce de « mobilier & design enfant » : le mot est ou a été à la mode. The American Heritage Dictionary of the English Language (Pickett et al., 2000) retient les acceptions suivantes du mot Serendipity : "1. The faculty of making fortunate discoveries by accident. 2. The fact or occurrence of such discoveries. 3. An instance of making such a discovery." Il propose en outre cette histoire du mot qui appartient désormais au folklore du sujet (van Andel, 2005) : "We are indebted to the English author Horace Walpole for the word serendipity, which he coined in one of the 3,000 or more letters on which his literary reputation primarily rests. In a letter of January 28, 1754, Walpole says that 'this discovery, indeed, is almost of that kind which I call Serendipity, a very expressive word.' Walpole formed the word on an old name for Sri Lanka, Serendip. He explained that this name was part of the title of 'a silly fairy tale,

called *The Three Princes of Serendip*: as their highnesses traveled, they were always making discoveries, by accidents and sagacity, of things which they were not in quest of...’.”

Le même alinéa se poursuit alors par l’indication suivante : « Ministère de la culture et de la communication [<http://www.culture.fr>], *Toutes les bases*, liste de ressources culturelles. » À nouveau, c’est là une référence devenue aujourd’hui introuvable et qui peut, en conséquence, détourner le lecteur d’un approfondissement éventuellement utile, à moins qu’il ne se trouve projeté dans la culture du vagabondage sitographique déjà évoquée. Cette référence a bien existé : la consultation de l’*Internet Archive* atteste de la présence, en 2002, d’une telle rubrique sur le site www.culture.fr, ainsi qu’en témoigne la copie d’écran ci-après ³⁰.



Il semble que la rubrique ait moins disparu que changé de nom, passant de « Toutes les bases » à « Collections » : on trouve en effet sur le site un « Accès en ligne aux données patrimoniales » à l’aide d’un moteur de recherche intitulé *Collections*, qui comporte une interface de recherche avancée ³¹.

Les références concernant les bases de données se poursuivent par la mention d’un site repéré par l’étiquette *Northern Light, Special Collection*, qui, précisent les auteurs, propose à l’époque l’accès à 12 millions de documents. En vérité, cette référence n’est plus valable depuis janvier 2002, si l’on en croit une information présente sur le site *Moteurs de Recherche* ³². Semblablement la référence qui vient ensuite, « *The Invisible Web (Intelliseek)* [<http://www.invisibleweb.com>] », présentée comme renvoyant à un répertoire thématique « de plus de 10 000 bases de données non indexées par les moteurs de recherche habituels », est à nouveau une adresse périmée : si l’on clique sur le lien indiqué, on est redirigé vers une page d’accueil du site *Profusion.com* où, comme on l’a vu, ce site annonce sa fermeture. De même

encore, l'ultime référence proposée au lecteur, « Webdata.com [<http://www.webdata.com>] », est une page introuvable. Si ce type de situations en forme apparente d'impasse ne conduit pas aux errances déjà mentionnées (c'est-à-dire à tenter de sortir de l'impasse en prenant n'importe quelle route), il peut engendrer une lassitude provoquant l'abandon de l'étude éventuellement engagée jusqu'en ce point, à moins qu'il ne provoque l'évitement du sujet d'étude que les indications involontairement erronées fournies au lecteur étaient censées l'aider à aborder.

Le dossier que nous suivons comporte encore trois sections. La première, intitulée « Enregistrer ou télécharger des données », est consacrée aux techniques de recueil des données sélectionnées sur le Web. Différentes manières de faire sont évoquées, à charge pour le lecteur qui voudrait en faire usage de leur donner un contenu technique concret : l'utilisateur pourra ainsi « enregistrer les adresses des sites retenus dans un carnet de liens sur [son] navigateur », « copier le texte et le coller dans un traitement de texte », « enregistrer les fichiers (textes, images, sons, vidéos...) sur un disque de [son] ordinateur », enfin « imprimer les pages qui [l']intéressent. Il s'agit là surtout d'un petit inventaire qui, pour le lecteur tout à fait débutant, apporte une première information sur des types de tâches qu'il lui restera à explorer plus avant, et qui, pour l'utilisateur plus aguerri, offre un tableau auquel confronter ses propres pratiques (il pourra observer par exemple que le fait « de copier les adresses retenues et de les coller dans un fichier de traitement de texte » est une possibilité que les auteurs n'ont pas retenue). À côté de cela, les auteurs expriment – très normalement – leur souci quant au droit de prélever ainsi des productions sur le Web, exhortant le lecteur en ces termes : « Respectez les droits de copyright associés aux textes et aux documents audio-visuels en citant vos sources et en demandant des autorisations si vous souhaitez diffuser les informations recueillis. »

Manuel d'instruction ou programme d'études ?

Le traitement de ce thème d'étude ainsi que de ceux précédemment évoqués fait apparaître le document que nous examinons moins comme un *manuel d'instruction* – qui offrirait au lecteur un ensemble de prescriptions, de commentaires, d'exemples, voire d'exercices, permettant d'apprendre en quasi-autonomie, sans avoir à réinventer l'essentiel de ce qui peut être appris – que comme un *programme d'études*, où est annoncé un « spectacle » didactique non encore réalisé et qui, en fait, reste à « créer » (pour emprunter au vocabulaire du théâtre). C'est ainsi que la deuxième section, intitulée « Évaluer la fiabilité de l'information trouvée », qui aborde un problème tenu pour crucial par nombre de commentateurs et d'analystes,

propose un thème d'étude où la matière praxéologique à étudier paraît tout à fait manquer. Les auteurs se bornent ici, en effet, à mettre en garde le lecteur contre les incertitudes pouvant affecter une information trouvée sur le Web, en lui prescrivant de « toujours vérifier » une telle information et de « chercher à identifier » sa source. Ce qu'ils proposent alors au lecteur en la matière se réduit à ce constat désolant que des données « essentielles » sur un document trouvé sur Internet (nom de l'éditeur du site, nom et qualité de l'auteur, date de création et de mise à jour du document, etc.) « ne sont connues que si elles sont indiquées par le producteur ou le diffuseur d'informations ». Il y a là, à l'évidence, une difficulté devant laquelle les auteurs ont renoncé, non sans raisons peut-être ; ce renoncement n'en marque pas moins une des limites de leur « plan d'études ».

La troisième section est intitulée plus vaguement « Ressources complémentaires ». Les auteurs ont rassemblé là, en une courte liste commentée, les références d'un petit nombre de sites jugés par eux utiles à leurs lecteurs. Le premier est un répertoire, *Search Engine Colossus* (<http://www.searchenginecolossus.com/>), présenté comme proposant « une classification thématique de plus de 1000 moteurs de recherche », devenue aujourd'hui, toutefois, une classification *par pays* surtout (voir la copie d'écran ci-après).

INTERNATIONAL DIRECTORY OF SEARCH ENGINES

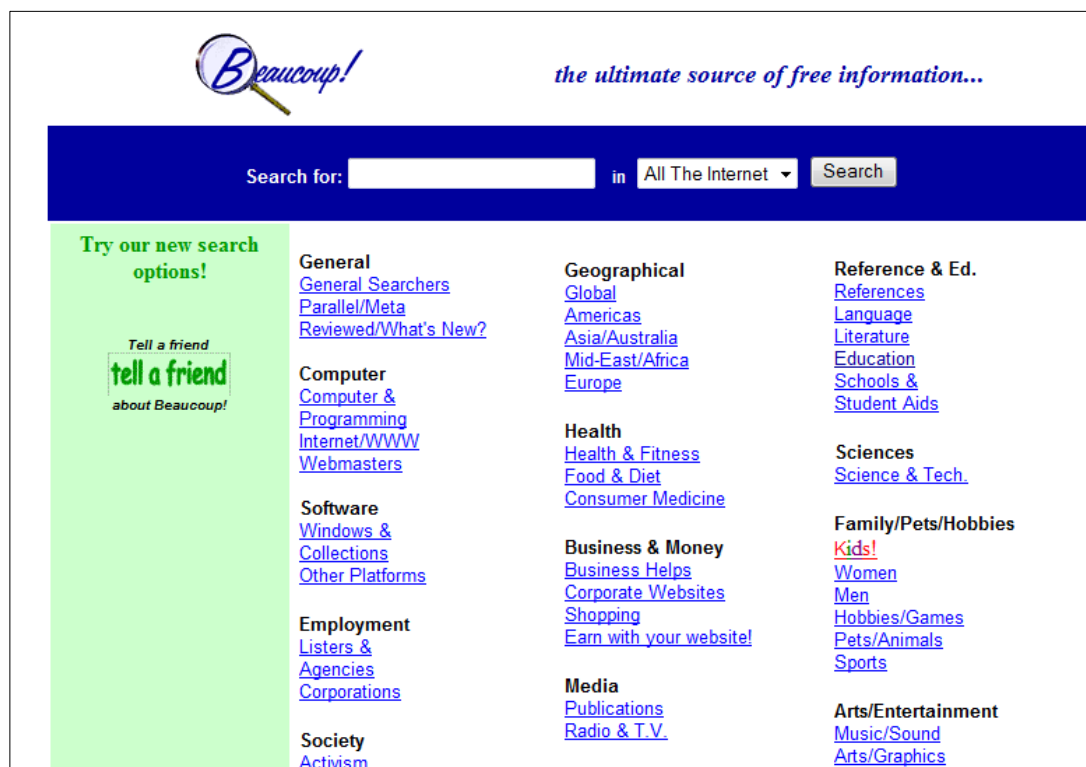
SEARCH ENGINE COLOSSUS

With Search Engine Colossus:
Explore 343 countries and territories!
Search in other languages!
Self submissions to search engines!
Since 1998.

Search Engines	Buscadores	Moteurs de Recherche	Zoekmachines	Suchmaschinen
 Submit Site	A D S	 Comprehensive, quality website directory! Over 10,000 sites listed. Add Your Site Today!	A D S	 Regional Web Directory
Aaland Abkhazia Aceh Adjara Adygea Afghanistan African Union Albania Algeria Altai American Samoa	Dagestan Denmark Djibouti Dominica Dominican Republic Easter Island East Timor East Turkestan Ecuador Egypt	Nagorno Karabakh Nakhichevan Namibia Nauru Nepal Netherlands Netherlands Antilles New Caledonia New Zealand Nicaragua Niger	Global ADS Start your Free Trial Free	


La deuxième référence est censée renvoyer à un moteur de recherche répondant au nom de *Countries and Territories Search Engines* (<http://www.internets.com/country.htm>). L'adresse mentionnée conduit aujourd'hui sur le site d'une société commerciale ; et la requête "Countries and Territories Search Engines" engendre, sur les moteurs de recherche consultés, un nombre très limité de résultats qui renvoient invariablement à la société commerciale susmentionnée ou encore à un blog aujourd'hui abandonné.

Au contraire de cette référence aujourd'hui morte, la troisième référence est encore vivante : c'est celle du répertoire *Beaucoup.com* (<http://www.beaucoup.com>), présenté comme permettant « une recherche par catégories thématiques ou par mots-clés », et dont le lecteur découvrira qu'il s'agit en fait d'un outil de recherche *en anglais* (voir la copie d'écran ci-après).



La quatrième référence, proposée sous le nom *W3 Search Engines*, est présentée comme « le répertoire des outils de recherche du Consortium W3 ». L'adresse indiquée (<http://cui.unige.ch/meta-index.html>) conduit, aujourd'hui, à une page introuvable. La recherche dans l'*Internet Archive*, pour la période où écrivent les auteurs (en 2002-2003), de cette page évanouie (voir http://web.archive.org/web/*/http://cui.unige.ch/meta-index.html), conduit cependant à la page reproduite ci-après. Il semble bien que les auteurs aient commis une confusion en identifiant de façon hâtive l'intitulé *W3 Search Engines* – qu'il faut entendre ici, à l'évidence, comme désignant les moteurs de recherche (*search engines*) disponibles sur le Web: *W3* ou *WWW* ou *World Wide Web* – comme renvoyant à des moteurs de recherche *du Consortium W3* (*W3 Consortium* ou *W3C*), c'est-à-dire ou bien à un *choix* propre au *W3C* de moteurs de recherche existants, ou bien même à des moteurs de recherche *développés* par ce consortium.

À l'entrée "World Wide Web Consortium" de leur Dictionnaire, les auteurs écrivent :
« Organisation officielle basée au MIT. Fondée en 1994 avec l'objectif de prendre la relève de l'IETF, le W3C vise à développer et à normaliser les différentes technologies utilisées sur la Toile mondiale, en particulier les évolutions du langage HTML. »



[What is the CUI ?](#)

W3 Search Engines

This documents collects some of the most useful search engines available on the WWW. Suggestions for additions are welcome!

Some interesting information sources are available only through specialized software. If you are willing to install software on your computer to access them, check our page [tools for information access](#).

Links that appeared to be broken are in a separate [page](#). Please let me know if one of those is still available! **NB:** If you find this page especially useful, why not install a local copy on your machine? It will work from anywhere. [The original source of this page is at [CUI](#)] You may also be interested in trying out Martijn Koster's [CUI](#) (Configurable Unified Search Interface), or visiting [Search Engine Watch](#) site which provides detailed information on how search engines work, along with tips on getting better results.

- [Information Servers](#)
- [Spider-based](#)
- [Natural language queries](#)
- [Other Catalogs](#)
- [Other Meta-Search Engines](#)
- [Software](#)
- [PEOPLE](#)

Cette bévue marque, sans doute involontairement, que, là encore, les auteurs parviennent à une limite extrême du domaine qu'ils ont exploré pour leur lecteur. Cette situation limite est le lieu de perturbations parfois étonnantes, corrélatives sans doute, chez le lecteur, d'un embarras dont nous avons noté certains des effets délétères possibles. Ainsi les auteurs mentionnent-ils encore une « liste d'outils de recherche de forums de discussion et listes de diffusion » que, selon eux, le lecteur devrait pouvoir trouver en une certaine page de leur ouvrage : or la page indiquée ne contient rien de tel ; et on cherchera vainement une telle liste dans l'ouvrage.

L'ultime référence propose des outils permettant une « évaluation quantitative et qualitative des outils de recherche » : il s'agit des deux sites *SearchEngineShowdown* (<http://searchengineshowdown.com>) et *Search Engine Watch* (<http://searchenginewatch.com>), aujourd'hui encore opérationnels mais dont la densité du contenu pourra apparaître décourageante au débutant : en visitant ces sites, en fait, l'utilisateur entrerait dans le domaine des

quasi-spécialistes de la recherche d'information sur Internet (voir les copies d'écran ci-après). Si flou soit-il, l'ensemble des références proposées par les auteurs devrait au moins suffire à suggérer au lecteur qu'il existe un domaine de connaissance appelant une étude particulière, à reprendre régulièrement : celui du monde divers et changeant des outils de recherche. Dans un petit ouvrage intitulé *Le PC pour les nuls* (Gookin, 2007), abordant la question des moteurs de recherche (qui n'est pas, il est vrai, son sujet), l'auteur clôt rapidement son propos par cette mise en garde : « L'utilisation des moteurs de recherche est une véritable science qui dépasse le cadre de notre ouvrage ! » C'est à cette science que le lecteur imaginaire suivi jusqu'ici aura donc reçu une première initiation.

Search Engine Showdown
The Users' Guide to Web Searching

UK Search Engines
If you're not listed you're missing out. Get listed today. £34+VAT.

Search Engine Submission
Solutions for Your Small Business. Business Begins Here.

Ads by Google

[Home](#) [Chart](#) [Reviews](#) [Search](#) [Problems with Page?](#)

Search News

Live Links Lost Again

At least it worked for awhile. The [trick of using the plus](#) for getting Live's link searching commands to work again (link:, linkdomain: and inurl:) that has worked since those commands were [disabled in March](#) has stopped working. It looks like link searching is again unavailable at Live, it continues to be incomplete at Google, so Yahoo! is the best choice for link searching.

While I can understand that problems with "automated usage for data mining" caused Microsoft to [disable the command in March 2007](#), it is really sloppy of Microsoft that all these months later, the [advanced search](#) still has a "links to" section that does not work, and their [online help](#) still has instructions on how to use this non-functioning "links to" section.

[See more »](#)

Dated Dec 19, 2007 in [Live Search](#) - [[#permalink](#)]

LiveSearch Turned Off

Yahoo!-owned [AlltheWeb](#) had their LiveSearch experiment running for over a year, display suggested search terms as you typed. Now, the [LiveSearch address](#) just redirects to Yahoo! I [suspect that this is because of](#) [Yahoo's new search engine](#) features.

Features
Feature Chart
SEs by Feature
News Searches
Directories
Phone Numbers
Others

Reviews
Google
Yahoo! Search
Live.com
Exalead
Ask

News / Blog
Blog
Email Lists
Alerts

Subscribe

Features
Archive Sources
Bookmarks
Book Searching
Custom Engines
Google
Review
Components
Unindexed URLs

Teaching Web Search Skills
Techniques and Strategies of Top Trainers
GREG R. NOTESS

SearchEngineWatch.com

Search ClickZ Network

[ClickZ](#) [Interactive Marketing Events](#) [Search Engine Watch](#) [Search Marketing Events](#)

[SEW Home](#) [Subscribe to Newsletters](#) [Subscribe to RSS Feeds](#) [Advertise](#) [Member Login](#)

PROTECT YOURSELF AGAINST

abcsearch
Building e-Business™
[ADVERTISE NOW!](#)

SIGN-UP TODAY AND GET \$100* OF FREE QUALITY TRAFFIC

You are here: SEW Home

SearchDay **Forums** **Search stats**

[» This Year in Search - Dec 28, 2007](#)
[» 2007: The Year that Search Grew Up - Dec 27, 2007](#)
[» Highlights from the SEW Blog: December 17-21, 2007 - Dec 26, 2007](#)
[» Search Engine Forums Spotlight: December 21, 2007 - Dec 21, 2007](#)

Blog News and views from the SearchEngineWatch blog

[» Norman Mailer On Google - Dec 31, 2007](#)

FREE WHITE PAPER:
RESCUING PROBLEM PPC CAMPAIGNS:
7 METRICS YOU MUST CONSIDER
[FREE DOWNLOAD »](#)

*Optimize the effectiveness

Le dossier préparé par les auteurs du *Guide bilingue de la communication électronique en 40 leçons* à propos de la recherche d'information sur Internet comporte, nous l'avons noté, un quatrième volet, le volet D, qui tient en deux pages (pp. 144-145). Intitulé *Vocabulary* (*Vocabulaire*), il se présente comme un petit dictionnaire qui, curieusement, va du français à l'anglais – à l'instar, il est vrai, de toutes les rubriques « Vocabulaire » de l'ouvrage³³. La raison de ce choix se trouve peut-être dans la volonté d'équilibrer les autres parties bilingues de l'ouvrage, où le texte en anglais est premier (il figure sur un verso, c'est-à-dire sur une page de gauche), tandis que le texte équivalent en français lui fait face sur la page de droite (sur un recto, donc). À titre d'exemple, on a reproduit ci-après le premier paragraphe de la leçon 10, consacrée aux *Weblogs* (les « Carnets Web »), tel qu'il figure sur la page 74 de l'ouvrage, ainsi que le texte en français qui lui répond sur la page 75.

10	WEBLOGS	10	CARNETS WEB
	A. What is a blog?		A. Qu'est-ce qu'un carnet Web ?
	In the last few years a new kind of personal publication has become very popular on the Internet: the weblog, or blog. It's "the talk of the town", a real buzz word, the object of articles in all the mainstream media, particularly after February 2003 when Google (the company behind the Internet's most-used search engine) bought Pyra, the company behind blogger.com (the online software most-used to create weblogs).		Ces dernières années, une nouvelle forme d'auto-édition est devenue très populaire sur l'Internet : il s'agit du carnet web, ou joueb. C'est « le dernier truc en ville », un mot à la mode, qui fait l'objet d'articles dans tous les grands médias, surtout depuis février 2003, lors du rachat par <i>Google</i> (la société qui détient le moteur de recherche le plus utilisé sur l'Internet) de <i>Pyra</i> , la société qui gère blogger.com (le logiciel en ligne le plus utilisé pour créer des carnets web).

Un autre aspect peut apparaître plus curieux encore. L'ouvrage est divisé en trois parties. La première partie, intitulée *Introduction*, comporte les quatre premières leçons. La troisième partie, intitulée *Tools* (Outils), rassemble les quatre dernières leçons. L'essentiel de l'ouvrage, soit les leçons 5 à 36, compose la deuxième partie, intitulée *Situations*. Dans ces 32 leçons, les leçons 17 et 18 apparaissent, au plan linguistique, singulières : contrairement à la quasi-totalité des autres, elles ne comportent pas de développements suivis en anglais avec, en regard, le texte français correspondant. Sur les 30 autres leçons, en effet, une seule échappe à

cette organisation – la dernière, pour des raisons presque évidentes quand on en lit le titre : « L’Internet n’est pas qu’en anglais... » (*The Internet is not only in English...*).

Peut-on interpréter le choix du « tout français » pour les leçons 17 et 18 que nous avons suivies ? Alors que dans les autres leçons l’accent semble mis d’abord sur une familiarisation *linguistique* avec les objets de la leçon, tout se passe ici comme si la nature de l’objet d’étude conduisait à mettre l’accent moins sur les *mots* que sur les *choses*, et, pour cela, à revenir à l’emploi de la langue supposée la plus familière, le français. En ce sens, il s’agirait, plus qu’ailleurs dans l’ouvrage, d’aider le lecteur à maîtriser ces types de tâches consistant à « préparer une requête », à « écrire une requête », enfin à choisir un type d’outil de recherche.

Dans les volets A, B et C du dossier, les mentions en anglais sont rares, voire rarissimes. Pour le volet A, on ne compte que le titre, *Preparing a query*, formulation à laquelle la traduction française en regard est un rien infidèle (« Préparer une recherche »). Le volet B n’est guère moins avare en matière d’emploi de l’anglais : hormis le titre (*Writing a query*, « Écrire une requête »), on n’y trouve guère que des emplois d’expressions obligées, et tout d’abord celles désignant les opérateurs booléens OR, AND, NOT, NEAR, etc., ou encore les expressions (supposées standardisées) permettant de solliciter les recherches par champ (*title, host, domain, link, anchor, text*, etc.). À cela s’ajoutent quelques mots d’anglais utilisés dans des exemples de requêtes (*cake, desert, cat, kitten, jam, jelly, raspberry*, etc.). Ce régime pauvre souffre une exception, le mot *wildcard*, que nous avons commenté plus haut, et un mot compagnon, *joker*, depuis longtemps entré dans la langue française³⁴. Bien que dépassant sensiblement en longueur l’ensemble des volets A et B, le volet D, quant à lui, ne fait apparaître franchement que deux traductions en anglais, celles de « base de données », *database*, et celle de « web invisible », *invisible web*.

Cet inventaire ne doit cependant pas masquer l’existence de fréquentes rencontres d’un *troisième type* avec des fragments de langue anglaise. Le premier type de rencontre est celle avec un texte *en anglais*, comme dans la leçon 10 citée plus haut ; mais on a dit que dans les leçons 17 et 18 un tel type de rencontre n’existe pas (à moins de faire entrer dans cette catégorie l’occurrence isolée du mot *wildcard*). Le deuxième type de rencontre a lieu selon le schéma organisant la rubrique « Vocabulaire », où une formulation en anglais suit la formulation en français qu’elle est censée traduire, comme il en va dans « bases de *données* (**database**) » (p. 141). Le troisième type tient à la rencontre avec des fragments de langue anglaise intégrés dans le nom de certains des outils de la RII. Par delà la floraison de mots anglais qui accompagne les opérateurs booléens ou la recherche par champ, on peut trouver un gisement de tels fragments d’anglais dans les listes de noms et d’adresses de sites Web : le

lecteur attentif aura observé ainsi l'emploi des mots (anglais ou de formation anglaise) *About*, *Looksmart*, *Excite*, *HotBot*, *Lokace*, *Northern Light*, *Webtop*, *inFind*, *Metacrawler*, *Search*, *Searchable databases*, *Special Collection*, *Intelliseek*, *Webdata*, *Search Engine Colossus*, et quelques autres encore, sans oublier un nom connu de tous mais en général ignoré du lecteur français comme mot de la langue anglaise, *Yahoo*.

Les Yahoos apparaissent dans le roman de Jonathan Swift Gulliver's Travels (« Les voyages de Gulliver », 1726), lors du séjour de Gulliver au pays des Houyhnhnms (où il demeurera du 7 septembre 1710 au 5 décembre 1715) : ils y sont dépeints comme des humains dégénérés, de la plus vile espèce. La fortune littéraire de l'œuvre de Swift a fait de yahoo, en anglais, un nom commun désignant qui adopte une conduite brutale, vulgaire, dénuée de manières : "he acted like a yahoo at the restaurant", dira-t-on ainsi. Le nom donné à l'annuaire de recherche Yahoo! – sans doute le plus gros index hiérarchique en son temps – est présenté souvent comme l'acronyme de Yet Another Hierarchical Officious Oracle. Mais les créateurs de Yahoo!, Jerry Yang et David Filo, ont aussi souligné que le choix de ce mot était dû au fait qu'il s'applique à un individu "rude, unsophisticated, uncouth".

Que montre l'examen du vocabulaire choisi par les auteurs ? Lorsqu'on l'entame en partant du début de la liste, un fait saute aux yeux : les premières entrées donnent la traduction en anglais de mots ou d'expressions *qui n'apparaissent pas dans le corps du dossier* (c'est-à-dire dans les volets A, B, et C). Ainsi en va-t-il de « agent intelligent » (*intelligent agent*, *robot*, *bot*), de « anneau, anneau Web » (*Webring*), de « association de termes de recherche » (*combining search terms*), de « balise META » (*META tag*), de « barre » (« de catégories de recherche, de raccourci aux catégories de recherche » : *category bar*). En parcourant l'ensemble du « Vocabulaire », on découvre que d'autres mots encore n'apparaissent *que là*. Ainsi en va-t-il de « boîte » (« de requêtes, de saisie de requêtes » : *search box*), de « recherche d'expression » (*phrase searching*), de « doublon » (*duplicate*), de « filtre familial » (*family filter*), quoique non de « filtre » (qui a quatre occurrences dans le corps du dossier), de « insensible (indifférent) à la casse » (*case insensitive*, *case independant*), de « caractère générique, caractère de remplacement » (*wildcard character*), de « méta-outils » (*meta-tool*), de « mise en majuscules » (*capitalization*), de « naviguer » (*to browse*), de « paramétrage » (« de la langue » : *language setting* ; « individuel » : *individual setting*), de « pages référencées » (*indexed pages*), de « pondération de termes » (*term weighting*), de « recherche booléenne » (*boolean search*, *boolean searching*), de « recherche en plein texte »

(*full text search*), de « recherche multimédia » (*multimedia search*), de « recherche simple » (*generic search, main search, simple search*), même si « requête simple » apparaît une fois dans le corps du dossier, de « recherche en langage naturel » (*natural language search*). Sans tenter d'être exhaustif, ajoutons à cette liste les entrées « technologie de distribution personnalisée » (*push technology*), « technologie de recherche active » (*pull technology*), « terme générique » (*generic term*), « tri par clé » (*key sort*), « version améliorée d'un logiciel, d'un outil de recherche » (*enhanced version of a software, search tool*).

Certains des mots et quelques-unes des expressions que nous venons de recenser sont proches sans doute de mots ou d'expressions apparaissant dans le corps du dossier : un exemple typique en est « pages référencées », qui n'y apparaît pas, alors que « pages indexées » est bien présent (avec quatre occurrences). De même, si « association de termes de recherche » n'apparaît pas dans le corps du dossier, l'idée est bien là, à propos des opérateurs booléens en particulier ; et il en va de même de l'expression « recherche booléenne », qui s'articule naturellement avec le développement consacré à ces opérateurs. La situation est à peu près la même pour « recherche en plein texte », que les auteurs présentent dans le « Vocabulaire » comme quasi synonyme de « recherche en texte intégral », expression que le lecteur attentif peut rapprocher d'un passage cité plus haut³⁵ : « Chaque moteur possède un mode d'indexation particulier (titre de l'URL, en-tête, premières lignes ou *intégralité des textes* d'un document...) qu'il est utile de connaître pour mieux formuler vos requêtes. »

En revanche les notions d'« anneau Web », de « balise META » ou de « tri par clé » constituent des additions non ancrées de façon évidente dans le corps du dossier. Certains additifs de cette sorte semblent être voulus comme *complétant* le tableau que les trois premiers volets offrent au lecteur. Sans doute doit-on compter dans cette catégorie les références aux technologies *push* et *pull*, ou encore l'entrée « filtre familial ». L'inventaire que nous avons dressé porte cependant à une autre réflexion encore : l'exposé semble se cibler sur quelques grands principes *technologiques* de la recherche d'information sur Internet, et délaisser quelque peu les questions de *technique*. En témoigne la fréquence différentielle des mots employés : « base », au sens de « base de données », a 12 occurrences sous la plume des auteurs, « information » en a 15, « moteur », au sens de « moteur de recherche », en a 40, tandis que « outil » (de recherche) en a 32, « répertoire » 21 et « requête » 28. Tels sont les concepts technologiques clés de l'exposé des auteurs. En sens inverse, certains des premiers mots ordinairement utilisés avec des débutants se remarquent par leur absence dans le corps du dossier : « barre » (*bar*) n'y a aucune occurrence, non plus que « boîte » (*search box*). Les problèmes d'utilisation de majuscules et minuscules ou du paramétrage (*setting*) de l'outil de recherche utilisé n'y sont nulle part mentionnés, alors même que ces questions renvoient à des

gestes techniques qui concrétisent les notions technologiques relatives à l'emploi d'outils de recherche pour effectuer des « recherches avancées » – souci qui paraît central dans l'exposé des auteurs. En vérité, le « Vocabulaire » comporte bien des indices de ce qu'il pourrait bien n'être qu'une pièce rapportée.

Si toutes les inadéquations relevées dans ce qui précède ne peuvent être regardées comme de tels indices (certaines ouvrent l'exposé sur des développements que les auteurs peuvent juger souhaitables tout en ayant choisi de ne pas les intégrer dans leur exposé), d'autres traits ne sauraient être interprétés de cette façon. La liste des entrées contient ainsi des doublons ou quasi-doublons : l'entrée « paramétrage de langue » (language setting) apparaît ainsi deux fois dans la liste. Cette dernière comporte en outre une entrée « recherche sensible à la casse » (case-sensitive search) et une entrée « sensible à la distinction majuscule-minuscule » (case sensitive, case dependant). On aura noté que “case sensitive” est écrit tantôt avec un trait d'union, tantôt sans. Plus généralement, il manque en vérité un « toilettage » de la liste utilisée, qui aurait permis sans doute d'éliminer diverses scories : une coupure de mot en fin de ligne alors que le mot n'est pas en fin de ligne (“wild card character” est écrit “wild card cha-racter”), un infinitif écrit de façon redondante (to browse (to)), etc.

Malgré cela, et peut être pour cela à plus forte raison, une conclusion s'impose, sans vraie surprise : bien que nous en ayons interprété certains traits comme significatifs d'une volonté de se rapprocher des « choses », et non d'en rester aux mots, l'exposé est plus technologique que technique, en contraste par exemple avec les deux premiers exposés examinés dans cette deuxième partie. Mais nous en retiendrons surtout que, du fait même de préférer les mots aux choses, il ouvre sur un panorama vaste et profus, dont nous poursuivrons autrement l'exploration dans la suite de notre travail.

Conclusion

Les analyses précédentes ont vu l'émergence insistante d'une notion en vérité cruciale, celle de *frontière praxéologique*. Il y a, dans l'emploi fait ici du mot de frontière, plus qu'une simple métaphore géopolitique. Une frontière organise l'espace – ici l'espace praxéologique dans lequel se déploie l'ensemble Σ des exposés sur la RII – et y exprime de façon spécifique des contraintes qui pèsent sur les personnes et les institutions agissant dans cet espace. En règle générale, du point de vue des collectifs anciennement stabilisés, enracinés et « racinés »³⁶ – au plan praxéologique où nous nous situons ici –, on est d'un côté ou de l'autre d'une frontière, dont le franchissement ne va pas de soi. Quand celui-ci se réalise, il est souvent furtif, momentané – on ne s'installe pas à demeure, en règle générale, de l'autre côté d'une frontière. Si l'on est un frontalier, qui ne cesse de franchir telle ou telle frontière, il arrive qu'on le cache, comme un fait mal compris, mal reçu, étrange comme peut l'être un étranger. Toute frontière sépare le groupe de personnes et d'institutions qui sont d'*ici* des personnes et institutions de *là-bas*. Vue d'ici, elle sépare l'endotique de l'exotique, l'exotérique de l'ésotérique. Vue de là-bas, elle peut porter à un sentiment de supériorité qui vient compenser quelque peu le déplaisir persistant de l'état de séparation où l'on se trouve. Une frontière praxéologique ne saurait se traiter par l'ignorance, notamment au plan didactique : car une telle frontière ne s'abolit pas aisément. Son franchissement, personnel ou institutionnel, est toujours un problème didactique délicat.

La toute première frontière qu'il faut noter en matière de RII est sans doute celle qui sépare le système praxéologique Π_0 « minimal » – taper quelques mots clés (en français) dans la fenêtre de recherche de Google (par exemple) et voir si, parmi les tout premiers résultats, figure « la » réponse escomptée – d'un système praxéologique méthodiquement enrichi, procédant d'une exploration même timide de ce vaste territoire praxéologique qu'on peut noter, selon le formalisme ensembliste usuel³⁷, $\Pi \setminus \Pi_0$, c'est-à-dire *tout le reste* de Π . En fait, il semble aujourd'hui que nombre d'utilisateurs ne « sortent » pas de Π_0 et même *n'envisagent pas de le faire*. Dans une enquête par questionnaire auprès de 1210 professionnels de la santé³⁸, dont 96 % disent disposer d'une connexion à Internet (laquelle est jugée par une majorité d'entre eux indispensable à l'exercice de leur profession), la question ouverte suivante était posée : « Quelles difficultés rencontrez-vous lorsque vous cherchez des

informations sur Internet ? » La réponse la plus fréquente – qui ne réunit toutefois que 12 % des enquêtés – proclame d’abord une *absence* de difficultés, ce en quoi on peut voir ce qu’on nomme en TAD un *déni de problématique*. Les 88 % restants mettent en avant des doléances solidaires, toutes classiques : tout d’abord, le temps pour effectuer une recherche est jugé souvent excessif et peu compatible avec les ressources temporelles disponibles, au travail comme chez soi ; ensuite, on invoque la difficulté à trouver les « bons mots clés » ; enfin, on se plaint du nombre élevé de résultats affichés par le moteur de recherche, résultats qu’il est alors difficile de « trier ». À l’évidence, ce témoignage renvoie en creux à un équipement praxéologique qui ressemble à s’y méprendre à l’équipement minimal Π_0 .

La plainte contre l’excès de temps passé à la RII est sans doute en partie un symptôme de simple nouveauté praxéologique : il n’entre pas encore dans l’*habitus* culturel commun de donner plus de quelques instants à la RII, alors que les mêmes personnes peut-être trouvaient hier normal d’aller passer d’assez longs moments dans une bibliothèque pour y effectuer une recherche documentaire simple. Mais on ne peut manquer d’entendre derrière cette protestation un autre facteur, à savoir le peu de prix accordé encore aujourd’hui à ce qui, pour cette raison, sera toujours jugé *trop* prenant, alors même que, pensera-t-on, cela ne devrait pas prendre plus de *tant* de temps. On retrouve ici les effets d’une contrainte essentielle déjà soulignée : la *péjoration du monde Internet*, un monde devenu vite indispensable sans doute, dont on peut de moins en moins se passer peut-être, mais dont on confesse volontiers qu’on le tient en piètre estime et, somme toute, qu’on « n’y croit pas ». De cette contrainte imposée par la culture ambiante d’aujourd’hui, le premier chapitre de cette partie a tenté d’alléger la tyrannie en évoquant des conditions antagoniques qui exposent, elles, les bienfaits (supposés) de l’Internet et, par exemple, l’intérêt qu’il y a à disposer, *en vrac*, sur un sujet donné, d’une *foule* de documents glanés dans un monde de ressources en forte expansion, quand bien même cette abondance cacherait quelques intrus, plutôt que de se contenter d’une pauvre documentation faite de bric et de broc à partir des ressources « bien contrôlées » d’un univers limité.

Ce travail de déconstruction permet d’apercevoir du même coup qu’une attitude souvent partagée prévaut, à chaque franchissement de frontière, selon laquelle *cela ne vaudrait guère la peine d’aller plus loin* en tentant de franchir telle autre frontière encore. Tout se passe comme si, un positionnement ayant été acquis, il tendait à être conservé indéfiniment... Le phénomène doit sans doute être mis en relation avec ce fait que le passage d’une frontière est coûteux à réaliser, et parfois à assumer une fois réalisé. Lors d’un des entretiens que nous avons conduits, une jeune physicienne travaillant dans un laboratoire de pointe nous indiquait utiliser abondamment et avec confiance la RII (dans des domaines très

spécialisés, peu susceptibles selon elle d'attirer l'activité des faussaires du Net). Ce n'est qu'après d'assez longs échanges qu'elle finit par préciser qu'elle utilisait Internet presque exclusivement *en anglais*, comme si le franchissement définitif de la frontière de l'anglais devait être traité avec discrétion et tact, pour ne pas se distinguer agressivement du reste du monde. Dans le même temps, alors que l'algèbre de Boole lui était familière par ses études scientifiques, elle avouait ne jamais se servir des opérateurs booléens, fût-ce par le truchement de la « recherche avancée ».

Revenons brièvement à un épisode, évoqué dans l'introduction à cette partie, dont le protagoniste, on s'en souvient, était un chroniqueur du quotidien *Libération*. Sous le titre *Google, no limits ?* (Marcelle, 2008b), son billet du samedi suivant revient sur « l'affaire » Emil Nolde. Sans qu'il en vienne jamais à résipiscence, l'auteur reconnaît que, ainsi que divers lecteurs le lui ont conseillé (implicitement ou explicitement), il convient de pratiquer des « recherches en anglais », puisque, note-t-il à regret, « les géants Google et Yahoo sont l'un et l'autre imparablement américains ». Cette reconnaissance d'une frontière est frappante, de la part de qui n'avait assumé jusqu'ici que des mots clés en français (ou en allemand, mais bien mal choisis : le pluriel *Kinder* – ou plutôt le datif *Kindern* – au lieu du singulier *Kind*, enfant ; le substantif *Raben*, corbeau (ou plutôt : corbeaux, au pluriel), pour désigner un « gros oiseau » qui n'en est peut-être pas un, comme on le verra plus loin). En réaction à son billet du 31 mai, un lecteur se disant « ni informaticien ni amateur d'art », mais « simple paysan qui passait par là, qui se sert pas mal de google, au même titre qu'une charrue ou un poste à souder, pas plus... », avait écrit ceci, qui nous fait d'un même mouvement franchir *plusieurs* des frontières repérées jusqu'ici ³⁹.

La requête simple -Emil Nolde children- dans Google image donne un résultat dès la première page:

<http://www.terrain.org/fiction/9/images/must3-5.jpg>

-Emil Nolde child- donne un résultat supplémentaire:

http://www.artchive.com/artchive/n/nold...e_bird.jpg

Connaissant maintenant la date, on entre -Emil Nolde 1912 "kind"-

pour une recherche dans la langue allemande, ce qui nous donne:

-Kind und großer Vogel aus dem Jahr 1912-

<http://www1.karlsruhe.de/Kultur/Presse/aug0201.htm>

"kind" ou +kind permet d'éliminer les résultats avec le pluriel.

Dans les domaines techniques où je me sers de cet outil le plus fréquemment, on peut dire que presque systématiquement une recherche en anglais est indispensable.

Non seulement la frontière de l'anglais est franchie (et présentée comme *devant l'être* en bien des cas), mais en outre est franchie la frontière qui sépare l'énoncé d'un simple *résultat* de recherche de la *narration*, même résumée, d'un *parcours* de recherche, ce qui, s'agissant de la *réception* de la RII dans la culture commune (en lieu et place de son *déni*), mérite d'être amplement souligné. On aura noté encore l'usage discret d'opérateurs booléens, ce qui nous fait franchir une autre frontière encore.

Tout cela n'empêchera pas le chroniqueur, dans son billet du 7 juin, d'écrire ceci : « Avons-nous résolu la question ? En apparence, mais pas vraiment. Nolde est allemand, *Child and large Bird* sonne a minima (*sic*) comme un traducteur automatique (*sic*), et il fallut la compétence d'une spécialiste muséale pour lui restituer son titre de *Kind und grosser Vogel*. » Cette dernière assertion – à propos de la nécessité de se référer à une « spécialiste muséale » pour connaître le titre allemand de l'œuvre – est un pieux mensonge : le troisième document dont l'URL était fournie dès le 31 mai par le lecteur déjà cité mentionne en effet, dans l'œuvre de Nolde, la période qui va de „*Vor Sonnenaufgang*, das 1901 entstand, der bahnbrechenden Radierfolge *Phantasien* von 1905 bis hin zu *Kind und großer Vogel* aus dem Jahr 1912“, c'est-à-dire de « *Vor Sonnenaufgang*, qui voit le jour en 1901, en passant par les eaux-fortes novatrices de *Phantasien* en 1905, jusqu'à *Kind und großer Vogel* de l'année 1912 » ; or peut-on douter sérieusement que le titre attribué en anglais, *Child and Large Bird*, soit la traduction du titre allemand *Kind und großer Vogel* ? On peut se demander au reste s'il n'y a pas, dans la quête du titre *en allemand*, un argument opportuniste de « mauvais perdant » : l'auteur, qui connaît certainement l'œuvre de Rembrandt appelée en anglais *The Night Watch*, s'est-il jamais soucié d'en rechercher le nom dans la langue du peintre ? ⁴⁰

Dans son article du 7 juin, cet auteur indique encore que, ayant utilisé les requêtes *Emil Nolde et bird* et *Emil Nolde et child*, il a obtenu de Google « une quarantaine d'écrans d'images, mais où l'œuvre obsédante ne se retrouve qu'en deux occurrences ». Dans une note de bas de page, il apporte alors cet additif perfide (dans un français « modernisé » qui semble lui être familier) : « Pour le reste, c'est le grand bordel as usual, faisant indifféremment défiler des Emil, des Nolde et des birds. » Le même lecteur de bonne volonté réagira à cette dernière assertion ⁴¹, typique d'une certaine péjoration du monde Internet, par une tranquille indication « technique » : « On peut aisément, écrit-il, limiter le grand bordel en utilisant les guillemets, "Emil Nolde child" sort trois images dont la bonne. "Emil Nolde" ne sort que des tableaux de ce peintre. » Le contraste est frappant entre l'âpre contempteur du Net, qui en vient à nier les faits, et l'humble et sagace travailleur du Net, qui trouve.

Les frontières repérables dans les textes que nous avons examinés apparaissent généralement comme la trace spécifique de *contraintes très génériques* – l'une d'elles et non

des moindres tenant, on l'a vu, à une forte réticence, culturellement motivée, face aux innovations technologiques de masse, et en l'espèce à l'Internet. La frontière de l'anglais, quant à elle, concrétise une attitude résultant sans doute de plusieurs contraintes qui structurent la culture et la société françaises dans la longue durée ; mais tout se passe comme si cette attitude découlait *pour l'essentiel* du postulat selon lequel, dans la vie intime, sociale, professionnelle, économique, culturelle, etc., *une seule langue suffirait*, le français pouvant être, au moins autant que d'autres langues, cette langue-là. Bien entendu, un tel précepte nie l'expérience ordinaire, commune à beaucoup, selon laquelle on a généralement besoin de *plusieurs* langues pour vivre tous les jours dans les sociétés contemporaines. Un schéma globalement analogue pourrait encore être évoqué à propos des opérateurs booléens : comme la frontière de l'anglais, le regimbement qu'ils suscitent généralement se produit même lorsqu'ils ont fait autrefois l'objet d'une étude formelle et alors même que leur maîtrise est en fait *à la portée de tous*, en consonance avec le « principe d'intelligence toujours suffisante » posé dans l'introduction à cette deuxième partie. Mais la frontière la plus stérilisante peut-être, dont le non-franchissement est au reste solidaire de l'arrêt devant les frontières précédemment évoquées, est celle de *l'élaboration praxéologique* de la RII : tout semble se passer comme si, une fois accomplies de façon plus ou moins fouillée la présentation et la démonstration des principaux « outils » numériques retenus, tout le reste – soit le bon usage de ces outils – était à la charge de l'utilisateur, à qui le poids de *l'invention praxéologique* idoine incomberait ainsi presque totalement ! Là encore, on peut invoquer une grande contrainte de nos sociétés, où les artefacts destinés à l'utilisateur ordinaire sont supposés être pleinement utilisables à l'aide d'un simple mode d'emploi (Morel, 2007), selon un postulat qui nie implicitement et la problématique des types de tâches ainsi instrumentés, et la dose de didacticité que la diffusion praxéologique correspondante appelle nécessairement.

Ce qui précède montre que la construction et la diffusion d'une réponse à la question de la création d'une culture praxéologique qui nous soit bénéfique en matière de RII ne va nullement de soi. En ne franchissant pas la frontière qui permet d'accéder à la considération de *parcours* de RII, c'est-à-dire en se contentant d'évoquer des recherches sans jamais exhiber véritablement le *travail* requis de l'utilisateur et les instruments *praxéologiques* plus ou moins complexes permettant ce travail, les exposés examinés dans cette partie ressemblent à des pigeons qui n'auraient qu'*une aile*, et qui donc, selon la formule connue, ne pourraient *ni voler*. Comme le montre si peu que ce soit l'épisode autour du tableau d'Emil Nolde, franchir ce Rubicon, en revanche, c'est entrer dans un univers foisonnant de problèmes praxéologiques. Ajoutons ici un ultime détail à notre récit de cette « affaire ». Dans son second message, qui suit le billet du 7 juin, revenant sur la « recherche » par laquelle tout

avait commencé, notre paysan-internaute⁴² écrit ceci à l'adresse du chroniqueur de *Libération*, où l'ironie pointe plus qu'à demi : « Et pour votre première recherche, il aurait été bien de vérifier, sur Google, bien évidemment :-))), à quoi ressemblait le bec d'un corbeau... » Savoir que « corbeau » se dit *Rabe* en allemand est une chose ; apprendre à distinguer le bec d'un corbeau de celui du « gros oiseau » de Nolde en est une autre, et qui peut sans aucun doute faire partie de la recherche à conduire, en fonction bien sûr du parcours suivi. Cet « apprentissage » demanderait ainsi à l'usager qui a des prétentions littéraires (par exemple) mais n'a pas de lumières en fait d'ornithologie, d'oser, modestement mais résolument, franchir un instant cette frontière dont il n'avait pas pensé s'approcher un jour, pour revenir ensuite vers son petit Liré, chargé d'un léger savoir nouveau, utile à son projet. La construction d'une réponse à la question vitale qu'on a rappelée implique ainsi que l'on s'avance vers des praxéologies dont l'élaboration suppose le franchissement – et donc à terme l'affaiblissement ou le déplacement – de nombreuses frontières engendrées par les contraintes organisant la société et sa culture. L'ambition est-elle réaliste ? C'est là une question dont nous approfondirons l'examen dans la suite de notre travail.

¹ Nous préciserons ultérieurement l'ensemble Π en adoptant le point de vue selon lequel un tel élément $\pi \in \Pi$ est un « ingrédient » d'au moins une *réponse* attestée, *R*, à au moins une question *Q* relative à la RII.

² Pour voir ce tableau, aller à http://www.artchive.com/artchive/N/nolde/child_and_large_bird.jpg.html.

³ Pour le *Dictionnaire culturel en langue française* (Rey et al., 2005) un exposé est 1) un « développement oral ou écrit qui présente, expose (un ensemble de faits ou d'idées) » : en ce sens, ce dictionnaire renvoie à *analyse, description, énoncé, narration, rapport, écrit* ; 2) un « bref discours sur un sujet précis, didactique » : les renvois sont ici à *communication, conférence, discours, laïus, topo, leçon, état, mémoire, note, aperçu, résumé*.

⁴ Du premier cas relèvent ainsi les deux pages que consacre à la RII tel *Dictionnaire des nouvelles technologies en éducation* (Ameka, Béziat, Florentin, Marty et de Vulpillières, 2006, pp. 208-209) ; du second, les 315 pages d'un ouvrage intitulé *Internet. Techniques de recherche pour les professionnels* (Lenormand, 2007).

⁵ Tels par exemple Liu, 2007 ; Hock, 2007 ; Notess, 2006 ; Pomeroy, 2002 ; Witten, Gori, & Numerico, 2007 ; Boughanem & Savoy, 2008.

⁶ Nous parlerons à ce propos de *Web visible* et de *Web invisible*, même si cette terminologie est parfois discutée.

⁷ Voir, dans la première partie de ce travail, le chapitre 4, *Le cas de PageRank*.

⁸ Voir <http://pzwart.wdka.hro.nl/mdma/alumni/2006/thassine/bio/view>.

⁹ Sur Shmoogole, voir http://pzwart.wdka.hro.nl/mdma/alumni/2006/thassine/shmoogole_hmtl/view. On trouvera Shmoogole à l'adresse suivante : http://www.geuzen.org/cgi-bin/shmoogole/shmoogole_form.cgi.

¹⁰ Le 7 septembre 2008, on trouvait la vidéo de l'intervention de Marissa Meyer lors du Google Press Day 2007 à l'adresse suivante : <http://www.youtube.com/watch?v=dNRE5x-V-sQ>.

¹¹ L'enregistrement de l'émission de France-Info « Le sens de l'info » du 25 février 2007 est disponible à l'adresse <http://framablog.org/index.php/post/2006/10/19/wikipedia>.

¹² La bibliographie proposée ici ne saurait être exhaustive. Le même Henri Lilen a ainsi publié chez Vuibert les ouvrages suivants : *Cinq inventions à l'origine de l'ère électronique - Le transistor - Le circuit intégré - Le microprocesseur - Le micro-ordinateur - La carte à puce* (2007) / *Internet - Conseils et solutions pour tous* (2006) / *Une brève histoire de l'électronique* (2003) / *La saga du micro-ordinateur - Une invention française* (2003) / *Internet à l'usage des parents* (2002).

¹³ Peut-être à son insu (et par attraction du nom du logiciel de navigation utilisé), l'auteur désigne ici le navigateur utilisé par le mot d'*explorateur*.

¹⁴ Notons ici une omission de l'auteur à propos d'une possibilité que nombre de débutants semblent ignorer : la page qui s'affiche lorsqu'on lance un navigateur peut fort bien être une page blanche (et non telle ou telle page Web) ; il convient pour cela d'avoir paramétré adéquatement le navigateur utilisé.

¹⁵ À nouveau, le langage employé laisse place à une certaine imprécision : le moteur de recherche fournit les adresses de pages Web, non de sites (même si ces pages appartiennent évidemment à des sites).

¹⁶ L'auteur omet ici le cas des liens internes à une page Web.

¹⁷ L'auteur ne donne aucun exemple ; nous proposons celui-ci pour le confort de lecture de notre propre exposé.

¹⁸ Le propos de l'auteur souffre de petites imperfections qu'il lui est difficile de gommer dans une présentation aussi concise. C'est ainsi qu'il n'est pas toujours vrai que la barre d'état affiche l'adresse de la page Web à laquelle le lien sur lequel on pointe renvoie, par exemple lorsque la page en question contient des éléments en format Flash (.swf). Par ailleurs, il arrive que le pointeur de la souris, bien que positionné sur un lien, ne se transforme pas en une main, à cause par exemple d'une particularité du script sous-jacent.

¹⁹ P. 125. C'est nous qui soulignons.

²⁰ Voir <http://www.live.com/>.

²¹ Au moment où nous écrivons ces lignes, le navigateur *Mozilla Firefox* appelle *barre de navigation* ce que Internet Explorer nomme *barre d'adresses*.

²² On lit sur la quatrième page de l'ouvrage examiné que "The **teach yourself** name is a registered trade mark of Hodder Headline". La page 3 porte cette mention publicitaire : "For over 60 years, more than 50 million people have learnt over 750 subjects the **teach yourself** way, with impressive results."

²³ La question de la personnalisation (*customization*) est un thème sur lequel s'arrêtent nombre de cours de formation et d'ouvrages pour la formation relevant de la didactique du Web : notons ici que tout semble se passer comme s'il appartenait à l'utilisateur d'améliorer par là, et par lui-même, son équipement praxéologique en matière de RII...

²⁴ L'exposé ne parle pas du cas des « cartes images » mentionné comme troisième possibilité par H. Lilen, sans doute parce que ce troisième cas est intégré au second.

²⁵ Sur la notion de *texte du savoir*, voir Chevallard, 1991.

²⁶ Alors que Henri Lilen se réfère à Internet Explorer 7 (dont le développement n'a été lancé par Microsoft qu'en 2005), Moira Stephen utilise Internet Explorer 6.0 (livré avec Windows XP depuis 2001).

²⁷ Stephen, 2003, p. 232. Un *directory* est un annuaire.

²⁸ Qu'il existe un grand nombre d'outils de recherche sur le marché n'est sans doute pas étranger à l'extension donnée à ce thème. Mais cela ne saurait expliquer complètement la surface dévolue au traitement de la question des *types* d'outils, en eux-mêmes peu nombreux.

²⁹ Requête *atomiques* par contraste avec les requêtes *moléculaires*, combinaisons de requêtes atomiques à l'aide des opérateurs « booléens », sur lesquels on va revenir.

³⁰ Voir <http://web.archive.org/web/20020802035540/www.culture.gouv.fr/culture/bdd/index.html>.

³¹ Voir http://www.culture.fr/recherche/avancee/?filter_image=1&search=.

³² « Janvier 2002 : *Northern Light* ferme son moteur de recherche public et gratuit. Il focalisera ses activités sur les services dédiés aux sociétés en continuant à mettre à jour sa base de données de plus de 350 millions de pages qui sera à disposition de ses clients. » (Voir <http://www.lesmoteursderecherche.com/nlight.htm>.)

³³ Une telle rubrique figure dans 19 des 32 leçons qui composent la partie centrale de l'ouvrage.

³⁴ Le *Dictionnaire historique de la langue française* (Rey et al., 1993) date l'emprunt de l'année 1912.

³⁵ C'est nous qui soulignons.

³⁶ Detienne, 2003.

³⁷ En théorie des ensembles, on note $A \setminus B$ le complémentaire de l'ensemble B dans l'ensemble A, c'est-à-dire l'ensemble des éléments de A qui ne sont pas éléments de B ; dans le cas évoqué ici (celui de $\Pi \setminus \Pi_0$), on a $B \subset A$, mais cette condition d'inclusion n'est pas obligatoire.

³⁸ Il s'agit d'une enquête par questionnaire en ligne réalisée en février 2008 dans le cadre d'un mémoire professionnel dirigé par C. Ladage. La population examinée était celle des cadres de santé (ou des personnels faisant fonction de cadres de santé), en poste dans trente-deux CHU (centres hospitaliers universitaires) et dans trois cents quarante CH (centres hospitaliers) de France.

³⁹ Voir <http://www.ecrans.fr/forums/viewtopic.php?id=2639>.

⁴⁰ Comme on le sait, le tableau en question est appelé en français *La Ronde de Nuit* ; en néerlandais, *De Nachtwacht*. La référence à la nuit, classique depuis le XIX^e siècle, résulte d'une méprise (le lecteur peut se reporter à cet égard à l'article "Night Watch (painting)" de l'encyclopédie Wikipedia).

⁴¹ Voir <http://www.ecrans.fr/forums/viewtopic.php?id=2701>.

⁴² Nous n'avons aucune raison de douter de la présentation de soi fournie par cet intervenant.

TROISIÈME PARTIE

**DES SAVOIRS UTILES À LA RII ?
UN REPÉRAGE PRAXÉOLOGIQUE**

Introduction

Le monde caché sous le phénomène

est plus clair que le monde apparent.

Les premières constitutions nouménales sont plus solides

que les agglomérations phénoménales.

Gaston Bachelard, 1934

Dans la première partie de ce travail, nous avons montré, tout à la fois, la manière dont il est possible d'exploiter les ressources du Web afin de répondre à de multiples questions, et le fait que cette possibilité est fréquemment ignorée – un fait ubiquitaire, que nous retrouverons plus loin. Le Web apparaît ainsi aujourd'hui comme le pourvoyeur généreux, mais trop souvent dédaigné ou sous-exploité, de *milieux* appropriés à la mise en œuvre, à propos de questions fort diverses, d'une *dialectique des médias et des milieux* qui ne s'enferme pas dans les limites étroites de la seule capacité de jugement personnel de l'enquêteur ou dans le recours à des savoirs familiers mais dont l'adéquation à la tâche à accomplir est parfois très faible.

La deuxième partie, qui précède, a alors engagé notre travail dans sa perspective essentielle, celle de *l'analyse des conditions et contraintes de la définition d'une culture praxéologique de la RII offerte à tous*. Pour cela, nous avons dégagé et analysé certains éléments de l'ensemble « potentiel » Π des praxéologies de la RII présentées ou évoquées dans l'ensemble Σ des exposés existants sur la RII – en désignant sous le nom d'exposé (sur un thème donné) *toute totalité organisée, en général ouverte, de déclarations écrites* (sur ce thème). Le lecteur aura remarqué que, pour commencer d'explorer l'ensemble Π , nous avons choisi de n'examiner que des exposés en apparence peu ambitieux, sinon sommaires. C'est là un choix délibéré, qui sera maintenu autant que possible dans cette troisième partie. Même s'il ne saurait à la longue résister entièrement à notre volonté d'approfondir l'exploration de l'ensemble Π des praxéologies de la RII, ce choix a le mérite de montrer très concrètement que certains des éléments d'une culture praxéologique de la RII sont là, tout près de nous, à portée de la main dans les exposés examinés comme dans les multiples exposés de même espèce, mais qu'en même temps *ils y restent très souvent confinés*, ce qui, au reste, nous a

conduite à les donner à voir, d'une façon aussi systématique qu'il nous est possible dans ce travail.

Selon l'adage familier en TAD, « pour creuser profond, il faut creuser large ». Une culture praxéologique de la RII suppose ainsi toute une culture praxéologique *pour* la RII, soit ce que nous avons appelé, dans le titre de cette troisième partie, un ensemble de *savoirs utiles* à la RII. Pour le dire autrement, une éducation (personnelle, mais aussi bien collective) à la RII suppose toute une éducation *pour* la RII. C'est une telle culture praxéologique *proche* que nous explorerons dans cette partie. Plusieurs points doivent avant cela être nettement marqués. Le premier tient simplement à ce fait que le repérage auquel nous avons procédé – sans prétendre être exhaustive – couvre de *multiples domaines praxéologiques* (informatique, droit, économie, mathématiques, physique, histoire des sciences et des techniques, sciences politiques, etc.), dont peu d'experts sans doute peuvent se dire également familiers. Lorsqu'on déploie un système praxéologique, lorsque, comme nous le dirons, on en recherche les éléments, on découvre usuellement qu'il envoie des racines vers une foule de disciplines anciennes ou plus récentes, voire de disciplines émergentes, qui le nourrissent et que, en retour, souvent, il vivifie. Cette donnée de fait oblige, dans l'étape historique où nous en sommes objectivement, aujourd'hui, en matière de RII, à embrasser un grand nombre de champs praxéologiques, ce qui requiert une attention épistémologique ouverte, bien dans l'esprit de l'encyclopédisme des Lumières – lequel *ne consiste nullement*, ainsi qu'on le croit parfois, à empiler sans fin – et sans principe – des connaissances éparses, fussent-elles aimables ou érudites (on se reportera là-dessus à notre annexe 3, en fin de partie).

Le principe du repérage praxéologique réalisé dans ce qui suit renvoie solidairement à une autre notion clé de l'épistémologie classique : celle de l'*élémentation* d'une science, dont d'Alembert a fait, dans l'*Encyclopédie*, un exposé magistral (là encore, nous renvoyons pour plus de précision à notre annexe 3 déjà mentionnée). Élémenter une science consiste à s'efforcer de dégager des éléments praxéologiques qui permettent d'*engendrer* cette science. Bien entendu, il y a équivalence potentielle entre les *éléments* d'une science et cette science elle-même : les éléments en constituent une partie qui permet, par définition, de la « dérouler » toute entière. Cela étant, pour avancer dans la perspective indiquée plus haut, celle, donc, de *l'analyse des conditions et contraintes de la définition d'une culture praxéologique de la RII offerte à tous*, nous nous sommes donné pour horizon la mise au jour d'éléments – au sens indiqué – de la « science de la RII », en donnant au mot de science, dans ce contexte, l'acception compréhensive qui fut autrefois la sienne. Pour élémenter la science de la RII, nous devons donc élémenter une foule de domaines praxéologiques, et cela bien sûr en ne retenant de ces domaines, en principe, *que ce qui apparaît indispensable* pour engendrer

la science de la RII. Bien entendu, comme l'élémentation est un horizon qui ne saurait véritablement être atteint, ce que nous visons est, en pratique, le repérage d'éléments praxéologiques qui apparaissent, sinon strictement indispensables, du moins raisonnablement *utiles* pour engendrer une science de la RII.

Il s'agit donc, tendanciellement, de construire par élémentation une connaissance de certains domaines praxéologiques qui ne saurait se ramener exactement, sauf heureux hasard, à la connaissance déjà construite en telle institution existante : toute élémentation est en ce sens une *réélémentation*. Cette exigence devrait épargner à qui s'engage dans une telle perspective l'illusion de maîtrise, ce lot ordinaire du professeur ou du formateur servant une discipline établie. Il est cependant une façon classique de succomber à cette illusion : c'est de croire que, pour fonder une science nouvelle, il suffirait de réunir les rudiments de quelques sciences bien installées, et que, donc, il ne faudrait ici qu'un peu d'informatique, de droit, de technologie des réseaux, d'anglais aussi, une pincée de programmation, etc. Procédant d'un point de vue qui se voudrait surplombant, un tel jugement rate en vérité l'objet qu'il croit se donner parce que cet objet *reste en fait à construire*.

C'est là pourtant une erreur épistémologique qu'il est aisé de commettre. Trop souvent, en effet, en dépit de la large réception de la théorie de la transposition didactique (Chevallard, 1991), les travaux de didactique relatifs à un système praxéologique depuis longtemps répertorié dans l'institution scolaire font l'économie de préciser, autrement que par l'étiquette habituelle qui désigne ce système (« les équations du premier degré », « l'appui tendu renversé », « la respiration », etc.), la structure, la composition, le fonctionnement de ce « savoir », que l'on suppose alors « connu » tant des auteurs que de leurs lecteurs prédestinés. L'ignorance ou l'incertitude sont alors unilatéralement rejetées du côté des élèves, étudiants et autres personnes en formation, qui auraient précisément à « apprendre » ledit savoir – et qui sont pour cela supposés ne pas le connaître. On peut voir en cette posture la reprise acritique par des chercheurs de l'illusion professorale de maîtrise vis-à-vis des « savoirs à enseigner » que prescrit le « contrat social d'enseignement » actuel, et cela alors même que le « contrat social de recherche » n'exige pas d'eux qu'ils « possèdent » les connaissances existantes (lesquelles ne sont pour eux qu'un *moyen* potentiel), mais bien qu'ils *produisent* les connaissances manquantes utiles à tel ou tel projet praxéologique. En bonne épistémologie de la didactique, donc, l'examen d'un système praxéologique quel qu'il soit suppose ce qu'on a pu appeler un *modèle épistémologique de référence* (Bosch & Gascón, 2005), qui ne se réduise pas à se prévaloir formellement d'une version savante ou pseudo-savante supposée du savoir considéré, mais qui, précisément, dans son principe au moins, permette d'interroger toute version possible, y compris savante, de ce savoir. Dans le cas qui nous occupe, cette

exigence est patente : si l'on peut, semble-t-il, supposer l'existence « quelque part » d'un savoir « total » des équations algébriques ou de la respiration, au moment où nous écrivons la chose n'apparaît guère crédible s'agissant de la RII. De là ce fait heureux (au plan de l'épistémologie des recherches en didactique) qu'un *repérage praxéologique explicite*, fût-il partiel – et sans doute, à notre insu, partiel –, apparaît indispensable à quiconque souhaite soulever, ainsi que nous le ferons ensuite, le double problème de la définition d'une culture praxéologique de la RII et de l'accès de l'utilisateur non spécialisé à cette culture praxéologique. On ne s'étonnera pas, en conséquence, que, dans les chapitres qui suivent, nous ne nous soyons pas contentée d'évoquer des connaissances, des savoir-faire et autres éléments praxéologiques comme s'ils étaient « bien connus » de chacun, mais que, mettant la main à la pâte aussi souvent qu'il nous a paru nécessaire, nous en fassions la « démonstration » de la façon la plus *explicite*. Contre l'inertie praxéologique que les frontières déjà identifiées – ou qui le seront dans ce qui suit – engendrent et matérialisent, la définition d'une culture de la RII productrice d'intelligibilité et d'opérationnalité est, croyons-nous, à ce prix – qui n'est que l'une de ses conditions de possibilité.

CHAPITRE 9

Premiers repérages : histoire, économie, politique, documentation, droit

Dans les examens monographiques réalisés jusqu'ici, une donnée relative à l'écologie de l'ensemble Σ des exposés n'a guère été soulignée, tant elle paraissait évidente : nombre d'exposés sur la RII se présentent comme des *parties* d'exposés sur un objet plus vaste, qui est en général « l'Internet » (plutôt que, simplement, « le Web », d'ailleurs). La plupart des éléments du corpus étudié sont, dans la période où nous écrivons, de cette sorte. Les trois exposés étudiés précédemment en fournissent une claire illustration : dans le petit ouvrage intitulée *Windows Vista et Internet facile*, l'exposé sur la RII occupe 27 pages seulement sur les 174 que compte l'ouvrage, soit 15,5 % de la surface imprimée ; dans le manuel *Basic Computer Skills*, l'exposé analogue occupe 10 pages sur 261, soit moins de 4 % de la surface imprimée ; enfin, dans le *Guide bilingue de la communication électronique*, le dossier sur la RII s'étend sur 12 pages dans un livre qui en compte 320, soit 3,75 % de la surface imprimée. S'il existe des exposés spécialisés ¹, ils constituent une petite part de la « littérature » sur laquelle nous nous penchons ici.

Les associations textuelles que nous observons ainsi, dans lesquelles le système textuel étudié semble parfois très étroitement enchevêtré avec d'autres systèmes textuels, laissent mal apercevoir les objets qui seraient propres à un texte du savoir sur la RII ; ou, plus exactement peut-être, elles ne permettent guère de saisir le *degré de spécificité* des objets qu'elles associent *de fait* au système textuel qui nous intéresse – par exemple les leçons 17 et 18 du *Guide* examiné dans la partie précédente. Au demeurant, il est clair que cette spécificité est elle-même évolutive en fonction des pratiques sociales mettant en jeu l'Internet : un exposé sur la RII visant un groupe humain dans lequel la pratique de l'Internet *hors RII* serait tout à fait banalisée n'aurait sans doute qu'à évoquer en passant, comme instrumental mais non spécifique, le fait de disposer d'une connexion à l'Internet, si possible à haut débit, etc. – de même qu'il n'est pas nécessaire dans une recette de cuisine de préciser ce qu'est une poêle ou une cocotte, alors même que, pour au moins certains praticiens domestiques de la cuisine, il conviendrait de préciser ce qu'est un moule à manqué ou un cul-de-poule. Le caractère évolutif de la situation examinée nous conduira quelquefois à des décisions anticipant une

stabilisation encore à venir des *pratiques* « *générales* » de l'Internet, cette évolution étant bien entendu articulée aux changements affectant constamment les TIC, que l'on pourrait appeler pour cela les *toujours nouvelles* Technologies de l'Information et de la Communication. Il n'est ainsi plus aussi indispensable qu'autrefois, pour l'utilisateur qui vit en milieu urbain, de savoir se connecter à l'Internet par le truchement d'un modem en écoutant le bruit délicieux des porteuses, annonciateur d'une connexion réussie ².

Domaines praxéologiques et RII

Dans ce qui suit, toutefois, nous procèderons, non à un « zoom avant » sur la RII, mais au contraire à un « zoom arrière » permettant de situer *a priori* la question de la RII dans un champ praxéologique beaucoup plus large, par rapport auxquels les exposés sur la RII sont conduits à se situer, même lorsqu'ils en ignorent de larges cantons. Les analyses conduites jusqu'ici suggèrent en effet que les matériaux praxéologiques présentés, mobilisés ou évoqués dans les exposés sur la RII s'appuient pour l'essentiel sur un ensemble de domaines de connaissance plus ou moins étendus, que nous présenterons ci-après en les « étiquetant » ainsi : *Histoire ; Économie ; Politique ; Documentation ; Droit ; Mathématiques, Physique et Technologies numériques ; Informatique ; Anglais ; Internet ; Web*.

Cette liste appelle immédiatement un petit nombre de commentaires. La série des cinq derniers domaines qui la constituent comporte un intrus apparent : la langue anglaise. À cette exception près, cette sous-série peut être regardée comme formant un emboîtement descendant : même si des lignes de démarcation ne sauraient toujours être précisément tracées entre eux, on peut dire que la RII, ou, pour dire la chose plus simplement, mais en anglais, la *Web Search*, suppose le *Web*, la Toile, qui à son tour suppose l'*Internet*, le réseau des réseaux, lequel interconnecte des ordinateurs et suppose ainsi l'*informatique*, laquelle participe plus largement du « *monde numérique* » dont Gérard Berry (2008) a brossé récemment un riche tableau, et qu'on ne saurait évidemment séparer de ses aspects proprement *physiques et mathématiques*.

Sans doute est-il un peu plus surprenant de se référer aux premiers domaines mentionnés, l'*histoire*, l'*économie*, la *politique*, le *droit*. Quelles en sont les raisons ? Le processus historique qui a en quelque sorte « hypostasié » pour nous des entités abstraites telles que « la physique », « les mathématiques » ou « l'électronique », pour ne citer que celles-là, n'a pas encore fait son œuvre s'agissant du « monde Informatique & Internet », c'est-à-dire de l'ensemble des activités sociales liées aux technologies numériques. Cet inachèvement d'une genèse qui finira – mais quand ? – par séparer le produit du processus de

production conduit aujourd'hui encore à multiplier, en certains exposés au moins, les références à l'histoire (saisie généralement dans sa dimension événementielle), aux conditions économiques, aux effets politiques, enfin aux aspects juridiques du développement du « monde Internet ».

Le domaine de la *documentation* se situe encore autrement par rapport à la RII. Il constitue en quelque sorte, au plan des institutions comme des praxéologies, un domaine que d'aucuns regardent comme tutélaire, parce que possédant, en matière de gestion de l'information, une légitimité antérieurement établie, dans des formes parfois immémoriales (Witten, Gori, Numerico, 2006), que le développement impétueux de la RII vient partiellement valider tout en l'interrogeant sans façon, en même temps qu'il reçoit de lui une aide régulatrice souvent précieuse.

Soulignons que, s'agissant de ce qui est le point de fuite de cette recherche – les conditions et contraintes possibles de l'étude d'une question *Q* utilisant le Web comme dispositif privilégié de la dialectique des médias et des milieux, permettant de tenter de produire et de valider une réponse *R* –, les domaines praxéologiques évoqués jusqu'ici ne constituent bien entendu que la composante *commune* du système de praxéologies dont la mise en jeu pourra apparaître pertinente. En pratique, en effet, les savoirs génériques relatifs à la *Web Search* opèreront – en règle générale – en *synergie* avec des savoirs *davantage spécifiques* de la question *Q* étudiée – que ces savoirs soient, pour « l'étudiant », indigènes ou allogènes.

Personnes, institutions, dispositifs

À l'instar des autres domaines du monde créé par l'homme, la manière dont le monde Internet et, plus largement, le monde numérique est d'abord rencontré ne s'exprime pas d'emblée en termes de praxéologies. Celles-ci, en effet, sont portées par des *personnes*, ancrées dans des *institutions*, et supposent des *dispositifs* idoines. Comme l'a illustré la première partie de ce travail, quiconque est, de gré ou de force, amené à se familiariser avec certaines praxéologies du monde Internet les rencontrera d'abord à travers des *noms* plus ou moins fortement emblématiques, récurrents, insistants, que ces noms désignent des personnes, des institutions ou des dispositifs. Pour ce qui nous concerne, et pour ce qui est des institutions et des personnes, citons en vrac (volontairement), et à titre d'exemples, *Google*, *Larry Page*, *Sergey Brin*, *Marissa Meyer*, le *Googleplex* de *Mountain View*, *Mozilla*, *Wikipedia*, *Shmoogle*, *Microsoft*, *Yahoo!*, *Internet pour les nuls*, le site *Search Engine Showdown*, etc. S'agissant des praxéologies dont nous essayons de saisir l'écologie et l'économie, et pour ce qui est

maintenant des dispositifs sans lesquels elles ne pourraient guère exister, on peut de même citer – de manière tout aussi désordonnée – *l'Internet* lui-même, le *Web*, *Mozilla Firefox*, *Live Search*, *Windows Vista*, etc., soit autant de dispositifs, qui s'inscrivent dans des institutions en s'incorporant (à) des personnes.

Chaque utilisateur découpera son propre répertoire de noms, renvoyant à une multitude de « personnages », humains et non humains, et ayant pour lui valeur de repère dans le « petit monde Internet » qui est le sien. Sa « tribu » et la culture de sa tribu, notamment, pourra contenir aussi bien des noms « connus » (*Bill Gates*, *Linux*, etc.), que les noms de « proches » – tel ami féru d'informatique, tel réparateur dévoué et compétent, tel serveur proxy, etc. Bien entendu cet univers tribal et objectal sera généralement appréhendé sur le mode de la *connaissance approchée* : parmi les fans de Linux, qui par exemple sait que ce mot est une contraction de *Linus' Unix*, « l'Unix de Linus », Linus étant ici le prénom du Finlandais Linus Torvalds, le créateur, dans les années 1990, de cette version d'Unix ? Bien entendu encore, les manques s'enchaînent : parmi tous ceux pour qui Unix est un mot-emblème de l'informatique, qui saura préciser ce que c'est, et en quoi, surtout, on gagnerait, le cas échéant, à savoir ce que c'est ?

D'une manière générale, l'approche des praxéologies se fait par le truchement d'institutions et de personnes et par le moyen de dispositifs : il y a là une contrainte universelle, qu'il faut constamment avoir à l'esprit (même si, dans la suite de ce chapitre, nous ne lui concèderons qu'une présence généralement fort discrète). Dans la biographie didactique d'une personne, ainsi, la découverte de l'existence des opérateurs booléens et de leur emploi dans la formation de requêtes « efficaces » pourra s'être réalisée avec telle personne agissant comme formateur pour le compte de telle institution (de formation), ou dans tel ouvrage couplé avec tel moteur de recherche (dispositifs), étudiés en autodidaxie, selon une logique d'auto-institution permettant de se légitimer provisoirement comme son propre formateur. Les praxéologies sont ainsi indissociables des institutions, des personnes et des dispositifs qui en permettent tant l'apprentissage que la mise en œuvre.

L'irruption de l'histoire

La première partie de ce travail a montré combien les dimensions historiques, économiques et politiques de la vie de l'Internet, et singulièrement celle de Google, paraissent entremêlées. Il y a là un fait qu'on doit d'abord constater, même si on peut en relativiser la portée en soulevant la question de sa *pertinence* pour la RII. La référence à l'histoire est sans doute celle qui doit être le plus fortement soulignée. « Expliquer l'Internet », le Web, ou Google, ne

se conçoit pas – pas encore ? –, semble-t-il, sans une *certaine* mise en perspective historique. Pour ne pas multiplier des observations que la clinique du monde Internet offre à foison, nous ne prendrons ici qu'un seul exemple.

Dans le *Dictionnaire bilingue Internet et multimédia* (Benenson & Juanals, 2004), les auteurs proposent deux entrées **INTERNET (THE)**. La première est brève et presque anhistorique : nous la reproduisons ci-après. (On y observera la présence d'objets ontologiquement divers : le fondateur du Web y côtoie par exemple le protocole TCP-IP.)

INTERNET (THE) n, **NET (THE)** n Internet (P) nm, Net (le) nm (mot à mot : *internetwork*, inter-réseau)

📖 L'Internet est le plus grand réseau informatique mondial. C'est un réseau longue distance qui est à la fois numérique et hypermédia. Il désigne un ensemble de matériel informatique utilisant un même protocole de communication, TCP-IP, et interconnecté en réseau. L'Internet est souvent surnommé « le réseau des réseaux », la « toile d'araignée électronique », en référence à la multiplication des réseaux dans les télécommunications, qui ont évolué vers l'interconnexion et l'interopérabilité. 🌀 *backbone* épine dorsale, *Berners-Lee (Tim)*, Internet 2, Matrix (the), *TCP-IP*, *World Wide Web (the)* Toile (la) (p. 208-209).

La seconde entrée **INTERNET (THE)** (pp. 211-213) propose essentiellement une chronologie, sorte d'épine dorsale de l'histoire de l'Internet, dont nous ne retiendrons ici que quelques événements parmi d'autres. En 1938, rappellent d'abord les auteurs, « deux ingénieurs américains, Bill Hewlett et David Packard, s'associent pour créer une société près de San Francisco, dans la future Silicon Valley ». Dans la période 1960-1964, Leonard Kleinrock, alors doctorant au MIT, pose les bases de la commutation par paquets (*packet-switching theory*), fondement de l'Internet. En 1968, Douglas Englebart invente la souris. En 1972 a lieu la démonstration, lors de la première conférence internationale sur les communications par ordinateur (*ICCC, International Conference on Computer Communications*), d'un réseau informatique de 40 nœuds. En 1980, Tim Berners-Lee écrit « une première version de son programme World Wide Web ». L'année 1988 voit la « mise en place d'une liaison entre la France et l'Internet par l'intermédiaire de l'INRIA ». En 1993 le navigateur *Mosaic* est développé à l'Université de l'Illinois : il ouvre la voie au développement de *Netscape* l'année suivante, lequel entrera bientôt en conflit avec le navigateur *Explorer 2* lancé par Microsoft en 1995. En 1993 avait été créé le *World Wide Web Consortium (W3C)* « par le CERN, le MIT, puis l'INRIA, pour suivre la normalisation du Web ». En 2000, un groupe de travail du W3C

définit le langage XML (*eXtended Markup Language*) destiné à remplacer le langage HTML (*HyperText Markup Language*), etc.

La densité du repérage historique est, selon les ouvrages et les contextes, plus ou moins forte – ou, en sens contraire, plus ou moins allégée. Dans l'ouvrage *Et Dieu créa l'Internet...* (1996, pp. 8-9), l'auteur, Christian Huitema, qui fut l'une des chevilles ouvrières du raccordement de la France à l'Internet, et qui en raconte les grands et les petits côtés, précise la date de cet événement historique (que la chronologie citée plus haut situe en 1988) : il eut lieu exactement le 28 juillet de cette année-là. Mais cet auteur ajoute des détails qui donnent à une simple suite de dates la profondeur de champ d'un vécu personnel, qu'il n'est peut-être pas sans intérêt ni signification de connaître aujourd'hui, alors que les affres des commencements semblent dépassées sans retour³.

La mise au point fut assez laborieuse : si le 28 juillet les premiers paquets de données avaient été échangés, le 29 plus rien ne marchait. Comme d'habitude, plusieurs problèmes se mélangeaient. Aux États-Unis, MCI ne fournissait que le service transatlantique ; il fallut quelques essais avant de valider complètement le raccord fourni par un opérateur local entre cette liaison et l'université de Princeton. Le logiciel TCP-IP avait été « porté » sur les machines françaises, mais le résultat n'avait jamais été testé en face de machines américaines. Par moments, les passerelles se bloquaient, les sessions s'interrompaient. Nous dûmes travailler avec acharnement pour diagnostiquer les pannes, pour finalement insérer d'autres machines en relais afin d'éviter les configurations « problématiques ». Tout cela, bien sûr, dans l'urgence, avec force heures de travail nocturne pour effectuer des tests malgré le décalage horaire. Néanmoins, le 8 août, tout était prêt. Nous fîmes une belle démonstration, nos partenaires de la NASA purent accéder à la base de données qui, à Strasbourg, récapitulait des siècles d'observations astronomiques. Yves Devillers et son équipe avaient déjà branché sur cette liaison les échanges de messagerie de l'INRIA : les temps de transfert tombèrent de quelques heures à quelques minutes, ce qui fut immédiatement remarqué et apprécié par tous nos collègues. La liaison était opérationnelle. (pp. 9-10)

Complétons ce petit tableau par la notice consacrée, dans le *Dictionnaire bilingue*, au protocole TCP/IP, sigle qui désigne d'abord le *Transmission Control Protocol* (protocole de commande de transmission), puis l'*Internet Protocol* (protocole Internet). On y retrouvera la référence à l'histoire de l'Internet et, en l'espèce, à l'activité de l'*International Networking Group* créé en 1972 avec, à sa tête, Vinton Cerf.

TCP/IP ACR (TRANSMISSION CONTROL PROTOCOL/INTERNET PROTOCOL) TCP/IP acr protocole nm de communication TCP/IP (mot à mot : protocole de commande de transmission / protocole Internet)

📖 Protocole mis au point en 1974 pour l'*ARPAnet* par Vinton Cerf et Robert Kahn et qui est à la base de l'Internet. Les données à transporter sur l'Internet sont scindées en paquets, chacun muni d'un en-tête contenant l'adresse du destinataire. Comme il est possible que des paquets suivent des routes différentes à travers l'Internet, le protocole vérifie la bonne réception de chaque paquet arrivé à destination et restitue l'ordre des données si nécessaire. C'est le protocole de commande de transmission qui assure le découpage des données en paquets, la détection/retransmission des paquets perdus et le réordonnancement des paquets. Le protocole IP prend uniquement en charge la transmission et l'acheminement des paquets des données (datagrammes) à travers le réseau. 📧
Arpanet, data packet paquet de données, *UDP* (pp. 418-419)

Les praxéologies dans lesquelles on peut ainsi vouloir entrer sont le fruit d'un passé récent, dont la plupart des acteurs étaient encore en activité il y a peu. Ce passé est même parfois si proche que ses héros sont encore jeunes et en pleine activité, comme il en va pour Google avec Larry Page et Sergey Brin.

L'enracinement économique

Nombre de faits ne se comprennent guère si, comme dans les illustrations précédentes, le récit « historique » tend à réduire l'aventure à sa seule composante « technologique ». Dans la plupart des grands journaux, c'est d'ailleurs moins cette composante qui est mise en avant (elle n'est en général explicitée qu'à titre d'adjuvant explicatif) que la dynamique *économique* qui fait corps avec l'épopée de l'Internet. Voici par exemple un extrait d'un article signé de Cécile Ducourtieux et publié sur quatre colonnes à la une par le quotidien *Le Monde* du dimanche 3 - lundi 4 février 2008 sous le titre « Microsoft veut s'offrir Yahoo! »⁴.

« Google contrôle aujourd'hui l'univers de la recherche et de la publicité en ligne », selon Jennifer Simpson, du cabinet Yankee Group. Le groupe, créé en 1998 par deux étudiants de Stanford, Larry Page et Sergey Brin, capte entre 30 % et 40 % du total des investissements publicitaires sur Internet. Grâce notamment à une offre très populaire auprès des annonceurs : les « liens sponsorisés ».

MM. Brin et Page ont eu l'idée, au début des années 2000, de s'appuyer sur les performances de leur moteur de recherche pour proposer aux annonceurs d'acheter des mots clés. Si ces derniers font

l'objet d'une requête en ligne, leurs liens commerciaux s'affichent à droite des réponses aux requêtes.

Microsoft dispose du portail « MSN », dont une messagerie électronique (« MSN Messenger »), très populaire. Cette offre est financée, comme dans le cas de Google, par la vente d'espaces publicitaires (liens sponsorisés ou bandeaux plus classiques). Mais son moteur de recherche vient très loin derrière celui de Google. Et les revenus générés sont encore peu significatifs à l'échelle du groupe (4,7 % des 51,12 milliards de dollars de chiffres d'affaires de Microsoft en 2007).

Or, le marché de l'« epub » pourrait passer de « 40 milliards de dollars en 2007 à 80 milliards en 2010 », a martelé, vendredi, M. Ballmer. Pas question pour Microsoft de se priver d'une telle manne. Dans cette perspective, le rachat de Yahoo ! fait sens. Sur les revenus publicitaires en ligne, le portail s'est largement fait distancer par Google. Mais il a récemment beaucoup investi pour revenir dans la course. Il a sorti une nouvelle version de sa plate-forme de vente de liens sponsorisés (« Panama »). Il a acquis la société BlueLithium, considérée comme une pépite. « Cette régie est la spécialiste du ciblage comportemental. Au lieu, comme Google, d'envoyer des publicités ciblées en fonction des requêtes formulées par les internautes, elle se base sur les différents sites visités par ces derniers », explique Laurent Geffroy, du cabinet spécialisé Greenwich Consulting.

Publié dans un grand quotidien français, ce texte évoque exemplairement les liens entre les technologies de l'Internet (et de la RII en particulier), les enjeux économiques de la publicité en ligne et les comportements des internautes. L'enjeu journalistique est ici d'explicitier les raisons pour lesquelles Microsoft souhaiterait s'emparer de Yahoo! : ces raisons sont économiques, et tiennent précisément à l'énormité du marché publicitaire en ligne existant et annoncé (80 milliards de dollars en 2010), dont Microsoft voudrait s'assurer une part substantielle, face à un rival jusqu'ici incontesté, Google. En ce point, on en reste essentiellement à des variables économiques (à ceci près que tout s'effectue ici en ligne : nous sommes dans le monde Internet). Mais l'auteur de l'article entend expliquer pourquoi Google détient sur ce marché une position si avantageuse ; et c'est là que le lecteur attentif devra entrer avec elle dans certains détails touchant à la RII. Au principe, comme on le sait, il y a un moteur de recherche – Google – et les requêtes qu'on lui soumet. Ce fait de base justifie, en remontée « historique », la référence à 1998 et aux « deux étudiants de Stanford ». Car, comme on l'a vu dans la première partie de notre travail, les *Google Guys* devront attendre longtemps – du moins à l'échelle des durées d'une start-up de la Silicon Valley ! – avant de disposer d'un *business model* fructueux : comme l'écrit l'auteur, « MM. Brin et Page » n'auront l'idée qui fera le succès de Google comme régie publicitaire qu'au début des années 2000. Il est remarquable, alors, que l'explication « complète » de la mise en œuvre de

cette « idée » conduise la rédactrice à décrire (sommairement, certes) l'organisation d'une page de résultats affichée par le moteur de recherche Google. L'utilisateur naïf de Google pourra ainsi découvrir ce que, peut-être, il ignorait systématiquement : les annonces publicitaires (*sponsored links*, liens sponsorisés) affichées à droite sur la page apparaissent si les sociétés dont elles font la publicité ont acheté certains des mots clés tapés par l'utilisateur. On notera en passant que l'article examiné se positionne sur ce point *en deçà* d'une frontière que d'autres rédacteurs, dans d'autres médias (moins généralistes) auraient pu s'autoriser à franchir : l'auteure, en effet, ne va pas jusqu'à y expliquer la manière dont ces liens sponsorisés rapportent de l'argent à Google: le principe est que, chaque fois qu'un usager clique sur un lien sponsorisé, la société concernée paie une petite somme.

D'après Battelle (2006), l'idée qui fera leur extraordinaire réussite financière a été empruntée par Page et Brin à Bill Gross, le créateur de la firme GoTo (qui deviendra Overture en 2001). Cet auteur écrit ainsi (p. 93) : "The story of how Google found its business model—and its subsequent rise to glory—requires a diversion into the history of another company, GoTo.com. For while Page and Brin struggled with the notion of turning search into a business, the founder of GoTo.com, Bill Gross, saw in search the seeds of an economic revolution." Il précise la chose quelques pages plus loin dans les termes suivants (p. 109) : "Gross built not one but two entirely audacious ideas into GoTo's initial business proposition for advertisers: first was the concept of a performance-based model—one in which advertisers paid for a visitor only when a visitor clicked through an ad and onto the advertisers' sites. Instead of demanding upfront money from advertisers, the way AOL or Yahoo did, GoTo.com's model guaranteed that advertisers had to pay only when their ads were clicked upon. Of course, this is now the standard model for the multibillion-dollar paid search market. Second, and even more audacious, was how Gross priced his new engine: one cent per click, an extraordinary discount to the market. He knew his price was seven to ten times less than what every Internet marketer was paying at the time, and in an environment where traffic was crack, advertisers couldn't help but look to Gross for a fix."

On peut se demander si, inversement, le fait de n'être pas entièrement ignorant quant au modèle économique mis en œuvre par la firme Google pourrait être utile à l'utilisateur du moteur de recherche Google. Nous laisserons ici la question ouverte, non sans avancer, à titre de pure illustration, la remarque suivante : on pourrait penser que la volonté de la firme Google de gagner de l'argent risque de porter atteinte à la qualité de son moteur de recherche ; or, il

semble tout au contraire que le rendement financier du dispositif des liens sponsorisés soit d'autant plus grand que le moteur de recherche de Google apparaît de meilleure qualité aux usagers. Pour dire les choses sans ambages, la volonté de gagner de l'argent par le moyen des *sponsored links* garantirait l'investissement de Google dans l'amélioration de son moteur de recherche. Les auteurs de l'ouvrage *Web Dragons* parlent sur ce point de "strong economic loop", écrivant à cet égard (Witten, Gori, Numerico, 2006, p. 23) : "Popular [search engines] attract a greater volume of advertising and produce the highest revenues, which funds improved and diversified services."

La politique ne se fait jamais oublier

La première partie a montré que l'Internet, le Web et la RII sont soumis à des contraintes politiques, notamment étatiques : le cas de la Chine est à cet égard paradigmatique (même s'il n'est pas unique), mais nous n'y reviendrons pas ici ⁵.

À ce propos, on pourrait reprendre l'interrogation soulevée quant aux conditions économiques de la RII, en posant par exemple la question suivante : les ressources du Web auxquelles je peux accéder sont-elles conditionnées – et comment – par le fait que je me trouve en France (par exemple) ? Bien d'autres interrogations peuvent surgir – quant à diverses formes de censure notamment, ou encore à propos des risques de « big brotherisation » ⁶. À nouveau, on peut imaginer que certains usagers réduisent leur emploi des moteurs de recherche – et notamment de Google – de crainte que certaines données personnelles, par exemple quant à leurs habitudes de consultation, ne soient stockées quelque part dans le monde, dans l'une ou l'autre des nombreuses « fermes de serveurs » de Google.

Inversement, là encore, une appréhension plus exacte, mieux informée, plus approfondie, des contraintes de tel ou tel type peut permettre à l'utilisateur un emploi amélioré, plus libre parce que plus lucide des moyens d'accéder aux ressources de l'Internet. C'est là une problématique généralisable : *en savoir plus permet, toutes choses égales par ailleurs, d'être plus libre*. Dans tout ce qui suit, nous garderons ce principe en tête, en le complétant cependant par une ébauche d'analyse quantitative, car il peut arriver que *beaucoup* de savoir *en plus* ne créent que *très peu* de liberté et de puissance supplémentaires ; tandis que, à l'inverse, il pourra se faire aussi qu'*un peu* de savoir supplémentaire soit la condition d'un *gain sensible* de liberté de pensée et d'action. C'est en de tels *points critiques* que nous serons évidemment amenés à nous poser la question des frontières susceptibles de faire barrage à la diffusion des praxéologies correspondantes.

Une vieille dame saisie par la modernité : la documentation

Comme pour les domaines de connaissance précédents – histoire, économie, politique –, la première partie a montré la pesée des praxéologies documentaires sur la RII, notamment à propos du thème de l'ordre et des désordres, du « vrac » et de l'organisation.

Le mot français « vrac » semble avoir comporté dès l'origine une nuance péjorative. (Sur son étymologie, voir par exemple <http://www.cnrtl.fr/etymologie/vrac>.) Sa traduction en anglais pose problème : dans l'ouvrage de Jean-Noël Jeanneney déjà plusieurs fois cité, la deuxième partie du chapitre 6, Organiser le savoir, s'intitule « Le vrac, danger absolu ». La traductrice en anglais, Teresa Lavender Fagan, a rendu ce titre ainsi : Disorganized Bulk – an Absolute Danger (Jeanneney, 2007, p. 70). Le développement correspondant commence en français par ces mots : « L'ennemi est clair, dans ces conditions : c'est le vrac. » (p. 121) La traduction est en ce cas la suivante : “The enemy is clear: massive amounts of disorganized information.” Le monosyllabe « vrac » employé par Jeanneney claque comme un mauvais coup porté à la culture ; la traduction adoptée en anglais semble par sa lourdeur même – disorganized information – plus tranquille, même si une ambiguïté sémantique s'y glisse – s'agit-il d'une « information désorganisée » ou, simplement, non encore organisée – unorganized ? L'information serait-elle par définition « organisée » ?

Nous ne nous arrêterons ici que sur une illustration. Dans un petit livre intitulé *Les bons usages d'Internet*, les auteures, Isabelle Bréda et Carole Hourt (2007), consacrent leur deuxième chapitre, intitulé « Se documenter, s'informer », à la RII : il s'agit là, typiquement, de l'un de ces exposés qui entrent dans la population de textes que nous étudions. Nous en examinerons ici la seule introduction, que nous reproduisons ci-après.

À côté des services de communication et d'échanges, Internet propose une mine étonnante de documents, réunis sur ce que l'on appelle le Web. On compare souvent le Web à une gigantesque bibliothèque universelle. Il s'en distingue pourtant à bien des égards.

Une bibliothèque propose à ses usagers des documents validés, dont l'ensemble constitue un univers fini : on peut savoir, par son catalogue, tout ce qu'elle contient. Qui plus est, le bibliothécaire, par son travail de médiateur, facilite la rencontre entre un lecteur et les documents dont il a besoin.

Il en va autrement sur le Web où tous ceux qui en ont l'envie, le besoin et les moyens techniques, peuvent émettre des informations. Universités, institutions politiques, médias, entreprises,

associations et citoyens y publient des contenus, sans qu'aucune instance ordonne l'ensemble de ces données. Il en résulte un foisonnement d'informations hétérogènes et diverses par leur qualité.

Internet n'est donc pas une bibliothèque au sens classique du terme, mais il est un formidable gisement d'informations. La connaissance y semble à portée de clic. Néanmoins, il faut apprendre à tirer parti des outils de recherche et à exercer un regard critique sur les résultats qu'ils proposent.
(p. 38)

D'emblée, comme on le voit, le contraste est marqué entre un univers des bibliothèques, contrôlé parce que muni de contrôleurs – les bibliothécaires –, intermédiaires obligés entre l'utilisateur et l'ensemble des ressources « validées » qui lui sont en principe accessibles, et l'univers du Web, foisonnant d'informations « hétérogènes et diverses par leur qualité ». La bibliothèque serait donc d'emblée un « bon objet », continuellement bonifié, et hanté par la figure tutélaire du bibliothécaire (ou de la documentaliste). Le Web, au contraire, est désigné comme un eldorado trompeur, où une méfiance légitime s'impose et où seul un abord critique des outils de son exploration et des ressources qu'il offre permet de ne pas se perdre.

La position exprimée par les auteures citées paraît aujourd'hui relever d'une certaine *doxa* en matière d'information et de documentation : le Web peut être un lieu de perdition parce que tout y est en vrac, non classé, non contrôlé, non validé. Ainsi que nous l'avons déjà indiqué, on peut répondre à ce point de vue par un autre, presque opposé, dans lequel on considère « l'organisé » comme *un ordre parmi d'autres possibles*, que d'aucuns appelleront *désordres*, faute par exemple de savoir *qui* les aura « organisés ». À cette passion de l'organisation, on peut alors opposer la passion du « brut », le goût pour le disparate, l'intérêt pour le divers laissé en l'état, etc.

Il se pourrait – c'est là, du moins, une hypothèse à ne pas écarter – que demain cette passion contraire devienne, si l'on peut dire, obligatoire. Dans un livre au titre significatif, *Everything is miscellaneous. The power of the new digital disorder*, un spécialiste américain, se penchant sur le cas de la Librairie du Congrès (Weinberger, 2007), commence par rappeler ceci.

The Library of Congress owns 130 million items, including 29 million books on 530 miles of shelves. Every day, more books come into the library than the 6,487 volumes Thomas Jefferson donated in 1815 to kick-start the collection after the British burned the place down. The incoming books are quickly sorted into cardboard boxes by topic. Then the boxes are delivered to three to four hundred catalogers, who between them represent eighty different subject specializations. They examine each book to see which of the library's 285,000 subject headings is most appropriate.

Books can be assigned up to ten different subject headings. Keeping America's books non-miscellaneous is a big job. (p. 15)

La création des catégories nécessaires est en soi un travail : chaque année, quelque 8 000 nouvelles catégories sont créées, tandis qu'à peu près autant de catégories existantes sont redéfinies. Lorsque la nouvelle catégorie des sciences environnementales (*Environmental Sciences*) fut créée, raconte Barbara Tillett, responsable de ces opérations, "catalogers were bouncing books back and forth: 'This one is yours.' 'No, it's yours.' " (p. 16). Cette activité fébrile et incessante comporte ainsi un arbitraire quotidien, phénomène que Weinberger illustre par l'exemple suivant.

Catalogers, using their years of experience, are free to propose new subject headings for an up-or-down vote by a committee composed of senior librarians. At one typical weekly meeting, just one out of eighty proposed classifications was turned down – a subclassification of the Photographic Portraits subheading to cover Sri Lankans. The committee decided that that headings subclassifications were intended to classify people by type – children, physicians, and so on – rather than by country of origin. Such issues may not be glamorous, but they are the type faced daily by those patrolling the conceptual borders of book categorization. (p.16)

L'auteur montre alors que le déploiement de forces au service de la prestigieuse *Library of Congress* n'est que peu de chose par rapport à ce qu'il conviendrait de faire si l'on voulait traiter les ressources du Web d'une manière analogue, ce qui rend illusoire un tel traitement – et nous conduira peut-être demain à accepter pleinement le *miscellaneous*.

It takes hundreds of professionals with centuries of cumulative expertise to keep the Library of Congress well ordered. But even though the Library of Congress has itself become a standard unit of measurement for large objects – for example, NASA says it maintains information about the environment that would "fill the Library of Congress 300 times" – it's only dealing with seven thousand new books a day. The Washington Post estimates that seven million pages are added to the Web every day. Search on Google for "American history," which is just one Library of Congress subheading, and you'll get 750 million Web pages – about twenty-six times the number of books in the Library's entire book collection. The Library of Congress's carefully engineered, highly evolved processes for ordering information simply won't work in the new world of digital information. Not only is there too much information moving too rapidly, there are no centralized classification experts in charge of the new digital world we're rapidly creating for ourselves, starting with the World Wide

Web but including every connected corporate library, data repository and media player. If the Library of Congress's well-proven approach won't work as we digitize our information, ideas, and knowledge, what will? (pp. 16-17)

Ajoutons à ces conclusions liées au grand nombre une remarque au plan simplement qualitatif de la catégorisation : celle-ci est toujours délicate, et il y aura toujours une catégorie d'utilisateurs pour la trouver arbitraire sinon scandaleuse. Il y a longtemps déjà, Jean Piaget (1896-1980) rappelait l'anecdote suivante (1968, p. 25, note 1), aussi significative que savoureuse, qu'il tenait d'une de ses plus proches collaboratrices, Bärbel Inhelder (1913-1997).

À Zurich, la psychologie s'est tellement limitée aux domaines soit psychanalytiques [d'inspiration jungienne] soit philosophiques que B. Inhelder, examinant dans une grande librairie universitaire le rayon « Psychologie » et demandant « Vous n'avez rien sur l'intelligence ? », a reçu la réponse suivante : « Ah ! vous placez l'intelligence dans la psychologie ? Nous ne savons jamais où la classer exactement et l'avons mise en médecine ! » Honneur donc aux psychiatres, et attention aux individus qui ne les consultent point sous le prétexte imprudent que la possession d'une intelligence claire ne soulève pas de question nosologique, comme s'il n'y avait pas là un symptôme inquiétant au point de vue jungien...

Nonobstant les observations précédentes, il va de soi que les sciences de la documentation et de l'information, ce qu'on appelle en anglais *Library and Information Science*, proposent des praxéologies pertinentes en matière de RII ⁷.

En France, l'école chargée par le ministère de l'Éducation nationale de la formation des conservateurs et des bibliothécaires de l'État et des collectivités territoriales, ainsi que des cadres des services de documentation et d'information scientifique et technique, porte actuellement le nom d'« école nationale supérieure des sciences de l'information et des bibliothèques » (ENSSIB), et délivre un master de « sciences de l'information et des bibliothèques ».

À nouveau, nous nous contenterons d'un exemple élémentaire, celui de l'exigence reconnue de « sourcer » les documents que l'on exploite dans une recherche – dès lors du moins que celle-ci n'est pas strictement privée et doit donc se soumettre à certains critères sociaux de vérifiabilité. Dans l'exposé déjà cité extrait de la plaquette *Les bons usages d'Internet*, les

auteurs écrivent par exemple ceci (dans un encadré fournissant quelques conseils aux lecteurs).

Pour vos travaux scolaires, universitaires et professionnels, pensez à établir une bibliographie, incluant une sitographie. Vous y noterez les références de toutes les sources d'information qui vous ont été utiles. Une méthodologie vous est proposée sur le site SavoirsCDI :

<http://savoircdi.cndp.fr/pedago/Refbibliogr/presentbibliolyc.htm>. (p. 52)

Le renvoi au site *SavoirsCDI* est significatif : dans la spécialisation institutionnelle courante au sein de l'institution scolaire, « faire une bibliographie » (ou une sitographie, etc.) relèverait de savoirs *en matière de documentation*. À titre d'exemple, on a reproduit ci-après une partie de la page dont l'adresse est proposée par les auteurs citées ⁸.

The screenshot shows the 'SavoirsCDI' website interface. At the top is a navigation bar with links: 'savoircdi', 'Le blog', 'Métier', 'Espace CDI', 'Acquisitions Traitement', 'Le CDI, outil pédagogique', 'Culture professionnelle', 'International', and 'Rechercher'. The main heading is 'Citer ses sources et présenter une bibliographie' in red. Below it, the section 'Lycée' is highlighted. A paragraph explains the document's purpose for students and teachers. Another paragraph mentions a reference-taking table for students. A 'Sommaire' (Table of Contents) section lists topics like 'Citer un document dans son intégralité' and 'Citer une partie de document'. A 'Version PDF imprimable' button is visible on the right.

savoircdi Le blog Métier Espace CDI Acquisitions Traitement Le CDI, outil pédagogique Culture professionnelle International Rechercher

Citer ses sources et présenter une bibliographie

Lycée

Ce document est rédigé à l'attention des lycéens, enseignants et documentalistes pour les aider à élaborer des références bibliographiques dans les situations les plus courantes. Ils apprendront ainsi à citer les sources qu'ils utilisent pour produire un exposé, un dossier, une page web....

Les collégiens se familiariseront progressivement avec les contraintes des normes en utilisant un tableau de prise de références, plus facile à remplir et correspondant mieux aux pratiques des collèges.

Sommaire

Consulter les choix de la rédaction

● Citer un document dans son intégralité

- Documents imprimés
 - ▶ Ouvrages
- Documents audiovisuels
 - ▶ Images, photographies, dessins, peintures (diapositives, transparents)
 - ▶ Films (cassettes vidéo ou dvd vidéo)
 - ▶ Enregistrements sonores (cassettes audio ou disques compacts)
- Documents électroniques
 - ▶ Cédéroms, dévédéroms
 - ▶ Internet : sites web

● Citer une partie de document

- Chapitre d'un ouvrage imprimé
 - ▶ Si l'auteur du chapitre est celui de l'ouvrage
 - ▶ Si l'auteur du chapitre est différent de celui de l'ouvrage

Version PDF imprimable
133 ko

Comme à propos des autres domaines de connaissance répertoriés dans cette partie, il conviendrait, dans un repérage préparatoire à une élémentation des savoirs pour la RII, de saisir ce qui, dans les exposés de RII, relève des praxéologies de la documentation, que ce soit par transposition ou par réinvention. Notons ici un seul point : dans la méthodologie

« proposée sur le site SavoirsCDI » en matière de références bibliographiques, la norme retenue s'explique, pour un ouvrage imprimé, comme suit.

Auteur, Prénom. *Titre de l'ouvrage.* Mention d'édition. Nom de l'éditeur, année de publication.
Nombre de pages. Titre de la collection. ISBN

On aura par exemple la référence ci-après.

Cannat, Guillaume. *Le ciel à l'œil nu en 2005.* Nathan, 2004. 144 p. ISBN 2-09-261067-8

Pour une « page de site Internet », le format proposé est le suivant.

Auteur du site, prénom. *Titre du site* [en ligne]. Nom de l'éditeur, date de publication, date de mise à jour [date de consultation]. Titre de la page. Disponibilité et accès

Il se trouve que chacune des informations recensées ici pose parfois problème. Qui est l'auteur de tel site ? Y en a-t-il un ? Et qu'en est-il de l'éditeur ? De la date de mise à jour ? Voire, en quelques cas, du titre de la page et même du site tout entier ? L'unique exemple proposé est le suivant, où l'on voit que plusieurs de ces informations manquent.

CNIL juniors [en ligne]. CNIL, 2004 [consulté le 11 juillet 2005]. Le jeu : découvre en quelques clics si tu connais tes droits sur Internet. Disponible sur : <http://www.cnil.fr/juniors/droits/quizz.htm>

Comment, alors, identifier les informations demandées, à supposer qu'elles ne soient pas manquantes ? Il y a là une difficulté objective, qui ne participe pas *que* de la connaissance de telle ou telle norme bibliographique, mais est suscitée par elle. La causalité s'inverse : à vouloir respecter certaines exigences documentaires, on se trouve alors confronté aux problèmes de la recherche d'informations sur lesquelles il n'est pas facile de mettre la main de façon assurée.

Le droit, outil et modèle

Le domaine du *droit* est à plusieurs égards une exception. Alors par exemple que, dans leur scolarité obligatoire, les élèves reçoivent une formation en histoire, acquièrent une familiarité avec les faits économiques et politiques et sont même initiés aux problématiques et aux

techniques documentaires, ils restent en revanche largement étrangers aux questions de droit – un domaine qui, plus généralement, semble aujourd’hui encore, en France, fort mal intégré à la culture commune⁹. Mais, en sens opposé, avec le développement de l’Internet, on observe un intérêt réel, souvent opportuniste et pressé, pour certaines questions de droit. Dans la formation scolaire au monde « Informatique & Internet », concrétisé par le B2i, le « Brevet informatique et internet »¹⁰, la référence au droit n’est pas absente, même si elle s’enveloppe dans des formulations en trompe-l’œil. Dans l’ancien référentiel du B2i apparaissait ainsi une compétence intitulée « Adopter une attitude citoyenne », dont le libellé est devenu, dans le nouveau référentiel¹¹, « Adopter une attitude responsable ». Bien que la mention d’une « attitude » ait apparemment, ici, une valeur euphémisante¹², il est attendu de l’élève qu’il devienne capable, au collège, de « connaître et respecter les règles élémentaires du droit relatif à sa pratique » et, au lycée, plus pleinement, de « connaître et respecter les règles élémentaires du droit relatif à l’informatique et à l’internet ». Nous reviendrons dans la quatrième partie de notre travail sur ce type de formulations « globalisantes », qu’elles concernent le droit ou tout autre matière. Mais on doit noter à cet égard que, ainsi qu’il semble en aller pour le profane, la maîtrise des éléments du droit requise des élèves paraît être regardée, dans ce contexte, comme un réquisit qui ne soulèverait guère de problèmes épistémologiques et didactiques *spécifiques*. De même, le référentiel du « Certificat Informatique et Internet » (C2i niveau 1) préparé et délivré dans les établissements d’enseignement supérieur¹³ prescrit, au titre du domaine de compétences A2 intitulé « Intégrer la dimension éthique et le respect de la déontologie », l’acquisition d’une « aptitude » désignée par l’étiquette « Maîtriser son identité numérique ». Or la notion même d’identité numérique est, du point de vue du droit comme du point de vue anthropologique, le fruit d’une conceptualisation complexe, qu’on ne saurait « maîtriser » si facilement, ainsi que le montrent les nombreuses études qui lui sont consacrées¹⁴.

La découverte et la familiarisation collectives avec ce domaine praxéologique d’abord faussement aisé ne manque pas de se traduire par un certain nombre de lapsus et autres interprétations réductrices. Pour ne prendre qu’un exemple, le site *Comment ça marche* évoque, à propos du droit à l’image, les articles 226-1 à 226-8 du code *civil*, alors qu’il s’agit là d’articles du code *pénal* ; cette erreur n’a pas été rectifiée dans l’ouvrage imprimé *Tout sur le Webmastering* de l’encyclopédie « CommentCaMarche.net » (Pillou, 2005, p. 203), non plus d’ailleurs que sur le site Internet lui-même¹⁵. Là où des lois complexes et d’interprétation délicate se trouvent augmentées d’une jurisprudence largement dépourvue d’univocité et toujours en évolution, la « transposition » scolaire entre des mains souvent inexpertes et toujours sous la contrainte de la rareté de l’espace d’exposition peut conduire à

des approximations peu utilisables. De cela nous donnerons une illustration plus précise en ouvrant un ouvrage intitulé *Informatique et internet au collège et au lycée* paru en 2007 au CRDP de l'académie de Grenoble, sous la signature d'Alain Carrier, professeur agrégé de physique et directeur de l'ingénierie éducative dans ce CRDP¹⁶. L'ouvrage a été conçu en relation avec la mise en place, au collège et au lycée, du B2i, déjà mentionné¹⁷. On y trouve une page intitulée « Adopter une attitude citoyenne », qui, à propos du droit à l'image, contient ce message (Carrier, 2007, p. 67) : « Une personne possède le droit de s'opposer à l'utilisation de son image. Le non-respect de ce droit constitue une atteinte à la vie privée. » À quoi cela renvoie-t-il ? On a là en fait la combinaison de plusieurs assertions. L'article 9 du code civil indique ce qui suit.

Chacun a droit au respect de sa vie privée.

Les juges peuvent, sans préjudice de la réparation du dommage subi, prescrire toutes mesures, telles que séquestre, saisie et autres, propres à empêcher ou faire cesser une atteinte à l'intimité de la vie privée : ces mesures peuvent, s'il y a urgence, être ordonnées en référé.

Dans une étude intitulée *Droit à l'image et droit de l'image*, l'auteur, Philippe Gauvin, apporte ce commentaire (Gauvin, 2006).

La première condition que pose le Code civil est que l'atteinte soit portée à la vie privée de l'individu. En d'autres termes, une image captée dans le cadre de la vie publique ne peut porter préjudice à quiconque.

Cependant, la vie privée et la vie publique ne sont pas strictement séparées pour qui que ce soit. Le juge, appréciant les cas qui lui sont soumis toujours *in concreto*, doit donc définir pour chaque espèce ce qui relève de la vie privée et/ou de la vie publique. Il ne suffit pas en effet d'être dans un lieu public pour que toute image puisse être captée, non plus que dans un lieu privé pour interdire cette captation d'image.

Le Code civil pose ensuite deux conditions :

- il faut qu'un dommage soit subi ;
- il faut que soit portée atteinte non seulement à la vie privée, mais surtout à l'intimité de la vie privée.

Concernant le dommage, il est généralement constitué par une atteinte morale, un préjudice moral. Son appréciation, qui va déterminer en grande partie le montant des dommages et intérêts et la peine éventuellement infligée en cas d'intention de nuire de la part de l'auteur, ne peut être effectuée que

par le juge, qui doit évaluer cela en son âme et conscience et en « bon père de famille », en fonction des évolutions de la société.

L'atteinte doit ensuite porter sur l'intimité de la vie privée. Il faut donc que le préjudice porte sur une situation habituellement réservée au cadre privé, cachée, secrète, qu'il « endommage » en quelque sorte la continuation de la vie privée, qu'il mette en péril le déroulement normal de la vie de la victime.

Cette atteinte à un droit essentiellement moral, donc détaché des choses réelles, implique une certaine subjectivité dans l'appréciation de l'atteinte.

Cette analyse suffit à suggérer que nombre de situations ne peuvent faire l'objet, de la part d'un profane qui plus est, d'une « lecture » juridique immédiate qui risquerait de « forcer » la situation dans un sens ou dans un autre. L'article 226-1 du code pénal précise par exemple ceci, où l'on retrouve « l'intimité de la vie privée ».

Est puni d'un an d'emprisonnement et de 45000 euros d'amende le fait, au moyen d'un procédé quelconque, volontairement de porter atteinte à l'intimité de la vie privée d'autrui :

1° En captant, enregistrant ou transmettant, sans le consentement de leur auteur, des paroles prononcées à titre privé ou confidentiel ;

2° En fixant, enregistrant ou transmettant, sans le consentement de celle-ci, l'image d'une personne se trouvant dans un lieu privé.

Lorsque les actes mentionnés au présent article ont été accomplis au vu et au su des intéressés sans qu'ils s'y soient opposés, alors qu'ils étaient en mesure de le faire, le consentement de ceux-ci est présumé.

La complexité des choses conduit souvent, dans les documents adressés à l'usager, à définir un espace parfois plus restreint que nécessaire, mais où la licéité est supposée à coup sûr (ou presque) garantie. Dans un ouvrage intitulé *Le droit de la vie scolaire*, les auteurs, Yann Buttner, André Maurin et Blaise Thouveny (2002), proposent un apologue scolaire dans lequel « un enseignant décide avec ses élèves de créer une œuvre multimédia dont le contenu permettra l'accès à une base de photographies dédiées à l'illustration de son cours ». Les photographies choisies exigent un nombre considérable de démarches qui, au demeurant, pourraient fort bien ne pas aboutir ! S'agissant des *personnes* uniquement, le tableau dressé par les auteurs est le suivant.

■ Une photographie des élèves de la classe réalisant le travail :

- le respect du droit à l'image conduit à obtenir l'autorisation de chacun des lycéens figurant sur la photographie ainsi que celle de leurs représentants légaux s'ils sont mineurs ;
- la photographie constituant une donnée nominative, sa numérisation nécessite une autorisation préalable de la CNIL ;
- le respect de la propriété intellectuelle du photographe exige son accord ; ses droits de reproduction et de représentation doivent être cédés à l'auteur de l'œuvre multimédia.

■ Une photographie d'un écrivain célèbre :

- il convient d'obtenir l'autorisation de l'écrivain (droit à l'image) sauf si ce dernier est en situation de personnage public (lors de la remise d'un prix littéraire) ;
- elle implique la préservation des droits d'auteur du photographe.

■ Une photographie du ministre de l'Éducation nationale en Conseil des ministres : hormis le respect des droits du photographe, aucune autorisation n'est ici requise, l'homme d'État étant en représentation publique. À l'inverse, le cliché pris lors d'une réunion familiale implique son accord.

(p. 91)

Les mêmes auteurs précisent que, s'agissant de l'autorisation concernant les élèves, elle peut prendre la forme suivante : *Je soussigné (...) autorise (...) à diffuser et à mémoriser la photographie de (...) sur le site (...) pendant une durée de (...).* Pour ce qui est de la CNIL – la Commission Nationale de l'Informatique et des Libertés –, il existe, précisent-ils encore, « une forme de déclaration simplifiée dont certaines se téléchargent sur son site Internet : <http://www.cnil.fr> ». Au tableau précédent, il manque le volet consacré aux *biens*, pour lequel nous renvoyons à l'ouvrage cité.

D'une façon générale, la situation en la matière fait que les choses peuvent se compliquer au-delà du raisonnable. Dans une conférence prononcée dans le cadre des Rencontres photographiques d'Arles le 6 juillet 2001, un juriste, Théo Hassler, professeur à l'université Robert-Schumann (Strasbourg), indiquait ceci (Kauder, 2004).

Si je saisis au débotté la photographie d'une personnalité, la seule autorisation *a priori* envisageable est celle de la personnalité, sujet de la photo.

Mais si cette personnalité est assise dans le jardin de sa propriété, les choses peuvent se corser : supposons que le photographe fasse des efforts de cadrage de façon à ce que l'on aperçoive une belle statue de bronze sur laquelle la personnalité pose sa main ainsi que la superbe maison qu'un de ses voisins a fait construire. Outre le consentement de la personnalité, il faudra examiner le cadrage de la photo de façon à voir si, le cas échéant, il ne fallait pas aussi le consentement du photographe, de

l'artiste qui a créé la statue, du propriétaire de la maison, de l'architecte qui en a conçu les plans. Cinq autorisations seraient alors nécessaires.

Ces autorisations sont de nature différente : trois dépendraient de la catégorie des droits d'auteur (photographe, sculpteur, architecte), une des droits de la personnalité (le sujet de la photographie), une dériverait du droit de propriété (le voisin propriétaire de la maison).

À la protection des droits indiqués tend à s'opposer un autre droit fondamental : le droit à l'information, le droit à se procurer et à diffuser de l'information. Le premier alinéa de l'article 10 de la *Convention Européenne de sauvegarde des Droits de l'Homme et des libertés fondamentales* (CEDH) signée à Rome le 4 novembre 1950 précise ceci ¹⁸.

Toute personne a droit à la liberté d'expression. Ce droit comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière.

Cet alinéa comporte toutefois une restriction : « Le présent article n'empêche pas les États de soumettre les entreprises de radiodiffusion, de cinéma ou de télévision à un régime d'autorisations. » Dans cette ligne, le second alinéa du même article indique alors ce qui suit.

L'exercice de ces libertés comportant des devoirs et des responsabilités peut être soumis à certaines formalités, conditions, restrictions ou sanctions prévues par la loi, qui constituent des mesures nécessaires, dans une société démocratique, à la sécurité nationale, à l'intégrité territoriale ou à la sûreté publique, à la défense de l'ordre et à la prévention du crime, à la protection de la santé ou de la morale, à la protection de la réputation ou des droits d'autrui, pour empêcher la divulgation d'informations confidentielles ou pour garantir l'autorité et l'impartialité du pouvoir judiciaire.

Bien entendu, dans la problématique générale qui est la nôtre ici – celle de la RII –, le droit à s'informer (et, donc, à informer) sont essentiels, même si cela ne doit pas conduire à méconnaître les droits protégeant les personnes et les biens. Une proposition de loi déposée le 16 juillet 2003 par les députés socialistes Patrick Bloche et Jean-Marc Ayrault visait à « donner un cadre juridique au droit à l'image et à concilier ce dernier avec la liberté d'expression ». Dans l'attendu des motifs, les auteurs écrivaient : « Il n'est plus possible d'ignorer que le droit à l'image est devenu un droit absolu, sans restriction aucune. » Ils poursuivaient dans les termes suivants ¹⁹.

La simple utilisation de l'image d'une personne, sans préjudice particulier pour celle-ci, mais aussi depuis peu celle de l'image d'un bien, sont devenues répréhensibles. Dans ce domaine, les pouvoirs du juge sont sans limites et les procès abusifs se multiplient. Les tribunaux en sont arrivés à mettre sur un pied d'égalité l'image d'une personne et celle d'un animal, d'un immeuble, d'un bateau ou encore d'un paysage.

Sous le régime actuel du droit à l'image, l'utilisateur d'une image, qu'il en soit ou non l'auteur, qu'elle représente des biens ou des personnes, doit s'assurer que les personnes ou les propriétaires des biens représentés ont bien approuvé expressément et par écrit l'utilisation en cause. Une telle autorisation s'avère souvent extrêmement difficile à obtenir, soit que le titulaire du droit à l'image éprouve des réticences quant à l'utilisation qu'il est possible d'en faire, soit, plus fréquemment, que ce titulaire demeure introuvable.

Le risque pris par l'utilisateur d'une image en l'absence d'autorisation est, à l'heure actuelle, suffisamment grand pour être totalement dissuasif. En effet, les dispositions de la jurisprudence actuelle ouvrent la possibilité à tout titulaire d'un quelconque droit à l'image d'obtenir une compensation financière, quand bien même l'utilisation litigieuse ne lui causerait aucun préjudice. L'effet pervers de cette jurisprudence, pourtant très louable dans ses intentions protectrices des droits de la personnalité et du droit de propriété, est d'inciter nos concitoyens à marchander leur image et celle de leur bien ou, pire encore, de provoquer chez eux des réflexes procéduriers dignes des pires recours d'outre-Atlantique.

Ce passage se concluait sur un constat alarmant.

Ce droit absolu à l'image, de construction uniquement prétorienne, né de l'interprétation extensive de textes très généraux du code civil, entrave de plus en plus les missions de pédagogie, de culture et d'information qui incombent jusqu'ici aux gens de l'image. La liberté d'expression est en souffrance.

La proposition de loi déposée visait donc « tout à la fois à prendre acte de la reconnaissance du droit à l'image par la jurisprudence actuelle et à infléchir celle-ci dans un sens plus conforme au respect de la liberté d'expression ». Sa philosophie peut s'exprimer en peu de mots : « Nul ne devrait pouvoir agir en justice pour revendiquer un droit à l'image sans rapporter la preuve d'un agissement fautif et d'un réel préjudice. » Elle prévoyait donc d'ajouter au code civil un article 9-2 dont voici le libellé.

Chacun a un droit à l'image sur sa personne.

Le droit à l'image d'une personne est le droit que chacun possède sur la reproduction ou l'utilisation de sa propre image.

L'image d'une personne peut toutefois être reproduite ou utilisée dès lors qu'il n'en résulte aucun préjudice réel et sérieux pour celle-ci.

L'article 2 de cette proposition de loi, quant à lui, redressait l'article 544 du code civil en proposant un article 544-1 avec une disposition analogue concernant les biens.

Cette proposition restera une proposition : elle ne sera pas retenue en 2006 lors de la réforme du droit d'auteur. Ainsi existe-t-il une tension résultant de la tendance à la privatisation de l'espace public, qui, qu'on le veuille ou non, devient une entrave à la liberté de s'informer. Dans une étude déjà citée (Gauvin, 2006), l'auteur note ainsi le lien causal entre cette situation et une contrainte montante, qu'il faut certainement situer, dans l'échelle des conditions et des contraintes de la TAD, au niveau de la *civilisation*.

Le droit à l'image est un droit qui s'est développé récemment, sous l'influence du développement d'une conception consumériste de la société par les individus qui la composent dans les pays occidentaux. Tout devient monnayable, y compris ce qui juridiquement a été conçu pour ne pas l'être.

Que seraient alors, sinon les éléments, du moins les rudiments de droit utiles à la RII ? Dans un petit livre déjà mentionné, les auteures soulèvent ainsi deux questions (Bréda et Hourt, 2007, p. 60-61) : « Télécharger, est-ce illégal ? » et « Dans quels cas peut-on faire des copies de fichiers trouvés sur Internet ? ». À la première question, elles répondent en rappelant d'abord que le téléchargement n'est pas techniquement illégal ; et elles ajoutent alors ceci ²⁰.

Ce qui est illégal, c'est de télécharger des fichiers protégés par le droit d'auteur et de les rediffuser. Pour mieux protéger les auteurs et les sociétés qui les produisent, la France s'est dotée en 2006 d'une loi, dite « DADVSI » (...). Elle légalise notamment des systèmes de protection et de contrôle des œuvres numériques et instaure des sanctions graduées à l'encontre des internautes qui téléchargeraient des fichiers protégés par ces systèmes.

Le texte s'arrête alors sur le P2P, le *peer-to-peer*, ou système « poste à poste », « technique qui permet aux internautes de mettre certaines ressources de leur ordinateur à la disposition des autres sans passer par un serveur central ou un intermédiaire ». Cette technique a des

usages nobles et d'autres qui le sont moins ; à leur propos, les auteures formulent cette mise en garde.

[Le P2P est] un moyen très répandu pour partager très largement des fichiers qui sont protégés par le droit d'auteur. Attention : avec la promulgation de la loi DADVSI en août 2006, la mise à disposition de fichiers soumis au droit d'auteur depuis un logiciel d'échange P2P et leur téléchargement peuvent être considérés comme des délits de contrefaçon.

S'agissant de la seconde question soulevée, le texte que nous suivons évoque alors le *copyleft*.

Parfois, un auteur donne l'autorisation de rediffuser son œuvre, à la seule condition que son nom soit systématiquement mentionné. C'est la logique du *copyleft*, ou copie libérée. C'est sur ce principe que Linux et les logiciels libres ont été créés. Des artistes ont choisi de diffuser leur œuvre de cette manière, sur les réseaux P2P. On peut donc recopier et diffuser librement leur travail, en précisant toujours le nom de l'auteur.

Puis se trouve mentionné le problème de disposer de multiples copies d'un fichier musical par exemple, pour l'écouter sur l'appareil de son choix : « un lecteur MP3, un Ipod, son ordinateur portable, etc. » À cet égard, les auteures se contentent de signaler que la loi DADVSI remet la responsabilité de la clarification nécessaire à une nouvelle « Autorité », dite « de régulation des mesures techniques ». Cette loi, promulguée le 1^{er} août 2006, fait alors l'objet d'une présentation concise qui occupe tout de même une pleine page (p. 62). Le lecteur profane peut apprendre par exemple que le droit d'auteur se divise en droit de représentation, droit de reproduction et droits moraux, et que les « droits voisins » sont, eux, « liés à l'interprétation d'une œuvre » ; il est informé en outre que la loi précise les « exceptions au droit d'auteur ». Judicieusement, faute d'un espace suffisant, cette notice renvoie non seulement au texte de la loi elle-même ainsi qu'au code de la propriété intellectuelle (CPI), mais également à un utile commentaire disponible sur le blog d'un avocat parisien, maître Eolas (2006).

L'étroitesse de l'offre contenue dans l'ouvrage examiné (de même, à plus forte raison, que celle du livre *Informatique et internet au collège et au lycée*) est sans doute la conséquence d'une volonté *générale* de « resserrer » la matière à présenter (plutôt que de « l'élémenter »). Mais ce choix forcé ne fait au fond qu'exprimer un fait massif, auquel les repérages praxéologiques dont nous rendons compte dans cette partie de notre travail nous ont ramenée encore et encore : le « monde Internet » semble de prime abord inépuisable et, même

en tentant de se limiter sévèrement aux seuls éléments potentiellement utiles à la RII, le « contenant » paraît toujours trop exigü pour accueillir le contenu que l'utilisateur serait en droit d'espérer, qu'il soit collégien ou praticien semi-professionnel. Pour prendre un exemple, celui du *copyleft* d'une œuvre (d'un logiciel, etc.), il ne semble pas inutile de savoir au moins que l'œuvre qu'on en tirera éventuellement (en modifiant le logiciel, etc.) sera elle-même obligatoirement *copyleft*. Considérons de même cet autre passage du dense opuscule évoqué ici.

Lorsque vous réutilisez des documents trouvés sur Internet, il faut veiller à respecter les droits d'auteur et de reproduction. Ces droits sont définis par le Code de la propriété intellectuelle, modifié par la « loi relative au droit d'auteur et droits voisins dans la société de l'information ». Vous pouvez invoquer le droit de citation si vous reprenez un court extrait d'une œuvre protégée, à condition de mentionner clairement l'auteur et l'adresse du site où vous l'avez trouvé. Dans le doute, et pour toute reprise d'un document entier (texte, image ou son), demandez par écrit à l'auteur ou l'éditeur du site l'autorisation d'utiliser son œuvre, en précisant bien ce que vous comptez en faire. (p. 52)

Ici, on peut s'interroger sur ce qu'est un « court extrait ». Le blog évoqué par les auteures – celui de maître Eolas –, fait apparaître en chacune de ses pages la mention suivante, qui rappelle le droit : « Tous les billets de ce blog sont la propriété exclusive du maître de ces lieux. Toute reproduction (hormis une brève citation en précisant la source et l'auteur) sans l'autorisation expresse de leur auteur est interdite. » Mais qu'est-ce qu'une « brève citation » ? L'enquête sur ce point ramène à la convention de Berne (1886/1979), qui, dans son article 10, indique seulement ceci ²¹.

Sont licites les citations tirées d'une œuvre, déjà rendue licitement accessible au public, à condition qu'elles soient conformes aux bons usages et dans la mesure justifiée par le but à atteindre, y compris les citations d'articles de journaux et recueils périodiques sous forme de revues de presse.

Les conditions énoncées, certes non négligeables, ne sont pas pour autant tyranniques : la citation doit être clairement attribuée à l'auteur cité et doit être « justifiée par le but à atteindre » – ce qui exclut sans doute certaines pratiques scolaires de « copier-coller ». Quant à la longueur de la citation, la convention n'en dit rien. En France, le CPI, en son article L122-5, précise ceci ²².

Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire [...] sous réserve que soient indiqués clairement le nom de l'auteur et la source : a) Les analyses et courtes citations justifiées par le caractère critique, polémique, pédagogique, scientifique ou d'information de l'œuvre à laquelle elles sont incorporées...

L'article « Droit de courte citation » de l'encyclopédie Wikipédia (en français) explicite dans les termes suivants, dans le cas des œuvres littéraires, un statut juridique « fixé depuis longtemps ».

- La citation doit être brève, tant par rapport à l'œuvre dont elle est extraite que par rapport au nouveau document dans laquelle elle s'insère.
- Elle implique que le nom de l'auteur, son copyright et le nom de l'œuvre d'où elle est extraite soient cités, de façon à respecter le droit moral de l'auteur. Dans le cas d'un extrait de livre, le titre, l'éditeur, la date de publication doivent également être mentionnés.
- La citation ne doit pas concurrencer l'ouvrage original et doit être intégrée au sein d'une œuvre construite, pour illustrer un propos. La citation en outre doit plutôt inciter le lecteur à se rapporter à l'œuvre originelle.

Ces précisions sont alors complétées par les illustrations ci-après.

La citation s'apprécie en principe par rapport à ce qu'est l'œuvre publiée. Par exemple, si Baudelaire a publié une œuvre *Les Fleurs du mal* (composée de poèmes), la citation intégrale de tout un poème pour illustrer le recueil serait une « citation » du recueil. En revanche, si le recueil est formé de poèmes publiés séparément, chaque poème est une œuvre indépendante, et le droit à citation ne permet pas de citer le texte intégral...

Les anthologies, ne sont ainsi juridiquement pas une collection de citations, mais des œuvres dérivées dont la publication nécessite l'autorisation des ayants droit de l'œuvre originale.

Au-delà, c'est au juge de se prononcer : en matière musicale, ainsi, dans un jugement du 15 mai 2002, le tribunal de grande instance de Paris a indiqué qu'un extrait musical de trente secondes d'un morceau de trois minutes n'était pas une courte citation ²³.

Le droit d'auteur, on l'a vu en passant, comporte des droits moraux et des droits qu'on nomme patrimoniaux – droit de représentation et droit de reproduction. Rapportés aux pratiques des internautes, on peut dire par exemple que « le fait de visualiser une page est une représentation, le fait de l'enregistrer sur son disque dur est une reproduction » (article « Droit

d’auteur » de Wikipédia). Comment la loi gère-t-elle cette situation (et quelques autres, liées aux techniques de navigation sur Internet) ? Avant même la loi DADVSI, une partie des difficultés potentielles était réglée par ce qu’on nomme l’exception de représentation et de copie privées. La loi du 1^{er} août 2006 met l’accent sur certaines pratiques assimilables à la copie privée dans son article 1^{er} où on lit notamment ceci ²⁴.

[Lorsque l’œuvre a été divulguée, l’auteur ne peut interdire] La reproduction provisoire présentant un caractère transitoire ou accessoire, lorsqu’elle est une partie intégrante et essentielle d’un procédé technique et qu’elle a pour unique objet de permettre l’utilisation licite de l’œuvre ou sa transmission entre tiers par la voie d’un réseau faisant appel à un intermédiaire ; toutefois, cette reproduction provisoire qui ne peut porter que sur des œuvres autres que les logiciels et les bases de données ne doit pas avoir de valeur économique propre.

Dans son blog, l’avocat déjà cité apporte alors ce commentaire éclairant.

[Les restrictions au droit d’auteur concernent] Les reproductions temporaires faisant partie d’un processus technique : il s’agit de ce que les informaticiens connaissent sous le nom de « cache » : ainsi, l’affichage par votre navigateur préféré des pages de ce blog entraîne automatiquement une copie de mes billets dans la mémoire cache de votre navigateur ; en principe, cette copie est couverte par l’exception de copie privée. La loi ajoute expressément une exception spécifique pour prévenir tout problème lié à une utilisation professionnelle des navigateurs, par exemple. Bon, aucun procès n’ayant été intenté à ma connaissance sur ce fondement, je ne pense pas que ces nouvelles dispositions vont bouleverser le droit français.

On n’ira pas au-delà de ces premières remarques ici ²⁵. Deux enseignements nous semblent devoir être tirés de l’état observable des relations entre le droit et le monde Internet. Le premier concerne directement le repérage praxéologique amorcé ici et ce qui en constitue l’horizon, l’élémentation des savoirs auxiliaires d’une « science » de la RII. Loin que l’on doive supposer à l’internaute une formation *préalable* en droit ou à tout autre domaine praxéologique, c’est, pensons-nous, le repérage et l’étude subséquente des éléments de ce domaine utiles à la RII (et plus largement au monde Internet) qui doivent être regardés comme ouvrant une voie de pénétration adéquate dans le domaine considéré, ici le droit – même s’il ne s’agit aucunement, par là, de revendiquer le statut de « juriste » ! Le second enseignement va en sens inverse de celui-ci : l’institution juridique et ses acteurs – magistrats et avocats notamment – ont dû *apprendre* le monde Internet afin de continuer à assumer leur raison

d'être – dire le droit, rendre la justice. À cet égard, la connaissance du monde Internet dans les milieux juridiques concernés peut être regardée comme un paradigme sur lequel méditer. Sans prétendre en approfondir davantage l'analyse, mais à simple titre d'exemple pratique parlant, nous avons simplement reproduit ci-après la liste des entrées du lexique figurant dans un ouvrage intitulé *Communication et Internet : pouvoirs et droits* (Delprat & Halpern, 2007, pp. 169-188).

Accès [terme en anglais : *access*] – **Active X** – **Additif** [terme en anglais : *add-on*] – **Administrateur de site, de serveur** [terme en anglais : *webmaster*] – **Adressage dispersé** [termes en anglais : *hash code, hash coding*] – **Adresse de courrier électronique** [terme en anglais : *e-mail address*] – **Adresse universelle** [termes en anglais : *Uniform Resource Locator, Universel Resource Locator (URL)*] – **Adresse IP** – **Agenda électronique** [terme en anglais : *Organizer*] – **Agent** [terme en anglais : *agent*] – **Amorce** [terme en anglais : *boot*] – **Amorcer** [terme en anglais : *boot (to)*] – **Ancre** [terme en anglais : *anchor*] – **Anneau de sites** [terme en anglais : *webring*] – **Anticrénelage** [terme en anglais : *antialiasing*] – **Applet Java** – **Appliquette** [terme en anglais : *applet*] – **Arrobe, arobase** [terme en anglais : *at-sign*] – **Arrosage** [terme en anglais : *spamming*] – **Article de forum** [termes en anglais : *News item, News posting, News article*] – **Assistant électronique de poche (ADP)** [terme en anglais : *Personal Digital Assistant (PDA)*] – **Autoroute de l'information** [termes en anglais : *information highway, information superhighway*] – **Balise** [terme en anglais : *tag*] – **Barrière de sécurité** [terme en anglais : *firewall*] – **Bloc-notes** [termes en anglais : *blog, Web log, Weblog*] – **Bombardement** [terme en anglais : *bombing*] – **Bombe programmée, bombe logique** [terme en anglais : *logic bomb*] – **Cadre** [terme en anglais : *frame*] – **Calcul intensif** [terme en anglais : *supercomputing*] – **Canular** [terme en anglais : *hoax*] – **Cheval de Troie** [terme en anglais : *Trojan Horse*] – **CGI (Common Gateway Interface)** – **Chat** – **Cliquer-parler** [termes en anglais : *Click and Talk, Web Call Back, Web Call through*] – **Code à octets** [terme en anglais : *bytecode*] – **Code source** – **Collecter** – **Concentrateur** [terme en anglais : *hub*] – **Configurer** [terme en anglais : *configure (to)*] – **Connectabilité** [terme en anglais : *connectivity*] – **Connecteur logiciel** [terme en anglais : *socket*] – **Connexion** [termes en anglais : *log in, log on*] – **Connexité** [terme en anglais : *connectivity*] – **Contribution** – **Cookie** – **Coprocasseur** – [terme en anglais : *coprocessor*] – **Courrier électronique, courriel** – [termes en anglais : *e-mail, electronic mail*] – **Crénelage** [terme en anglais : *aliasing*] – **Cryptage** – **Dialogue en ligne (forme abrégée : dialogue)** [terme en anglais : *chat*] – **Diffusion systématique sur la toile (ou diffusion sur la toile)** [termes en anglais : *webcasting, netcasting*] – **Disque numérique polyvalent** [terme en anglais : *Digital Versatile Disk (DVD)*] – **Distribution sélective** [terme en anglais : *push technology*] – **Domaine** [terme en anglais : *domain*] – **Données personnelles** – **Dorsale** [terme en anglais :

Backbone] – **Échange de données informatisé (EDI)** [terme en anglais : *Electronic Data Interchange (EDI)*] – **En-tête de message** – **Exploration de données** [terme en anglais : *datamining*] – **Extensibilité** [terme en anglais : *scalability*] – **Extranet** [terme en anglais : *extranet*] – **Fenêtre** [terme en anglais : *window*] – **Fenêtre d'attente** [terme en anglais : *splash screen*] – **Fil (de la discussion)** [terme en anglais : *thread*] – **Firewall** – **Flux (de travaux)** [terme en anglais : *workflow*] – **Foire aux questions (FAQ)** [terme en anglais : *Frequently Asked Questions, FAQ*] – **Forum** [terme en anglais : *newsgroup*] – **Fouineur** [terme en anglais : *hacker*] – **Fournisseur d'accès** [terme en anglais : *access provider*] – **Frimousse** [termes en anglais : *emoticon, smiley*] – **FTP (File Transfer Protocol)** – **Glisser-déposer** [terme en anglais : *drag and drop*] – **Grappe** [terme en anglais : *cluster*] – **Groupe** [terme en anglais : *cluster*] – **Groupe de discussion** – **Hachage** [terme en anglais : *hash code*] – **HTML** – **HTTP (HyperText Transfert (sic) Protocol)** – **Hypertexte** [terme en anglais : *hypertext*] – **Icône** [terme en anglais : *icon*] – **Incendier** [terme en anglais : *flame (to)*] – **Internaute** [terme en anglais : *cybernaut*] – **Internet** [termes en anglais : *Internet Network, Internet, Net*] – **Intranet** [terme en anglais : *Intranet*] – **Javascript** – **Lien hypertexte** – **Liste de diffusion** [terme en anglais : *mailing list*] – **Logement** [terme en anglais : *slot*] – **Logiciel antivirus, antivirus** [termes en anglais : *antivirus program, antivirus software, vaccine program*] – **Logiciel de messagerie** – **Logiciel libre** – **Logiciel de navigation, navigateur** [terme en anglais : *browser*] – **Logiciel malveillant** [termes en anglais : *malicious logic, malicious software, malware*] – **Logiciel médiateur** [terme en anglais : *middleware*] – **Logiciel socle** [terme en anglais : *baseware*] – **Magasin de données** [terme en anglais : *datamart*] – **Message incendiaire** [terme en anglais : *flame*] – **Messagerie électronique** [termes en anglais : *e-mail, electronic mail, electronic messaging*] – **Modérateur** [terme en anglais : *moderator*] – **Module d'extension** [terme en anglais : *plug-in*] – **Morphose** [terme en anglais : *morphing*] – **Moteur de recherche** [terme en anglais : *search engine*] – **Moteur d'exécution** [terme en anglais : *runtime software*] – **Mouchard** – **Multiprocesseur** [terme en anglais : *multiprocessor*] – **Navigateur** – **Netbios** – **Nommage** [terme en anglais : *naming*] – **Norme simplifiée** – **Numériseur** [terme en anglais : *scanner, digitizer*] – **Octet** [terme en anglais : *byte*] – **Page d'accueil** [terme en anglais : *home page*] – **Pare-feu** – **Passerelle** [terme en anglais : *gateway*] – **Permutation** [terme en anglais : *swap*] – **Pièce jointe** [termes en anglais : *attachment, attachment file*] – **Pirate** [terme en anglais : *cracker*] – **Pop-up** – **Portail** [terme en anglais : *portal*] – **Raccordement numérique asymétrique (RNA)** [termes en anglais : *asymetric (bit rate), Digital Subscriber Line (ADSL)*] – **Réamorcer** [terme en anglais : *reboot (to)*] – **Recherche individuelle** [terme en anglais : *pull technology*] – **Référentiel** [terme en anglais : *repository*] – **Réinitialiser** [terme en anglais : *reset (to)*] – **Réseau informatique** [terme en anglais : *computer network*] – **Responsable du fichier** – **Retouche** [terme en anglais : *patch*] – **Riper** [terme en anglais : *rip (to)*] – **Scrutation** [terme en anglais : *polling*] – **Serveur** [termes en

anglais : *server, on-line data service*] – **Serveur DNS (Domain Name Server)** – **Serveur mandataire** [termes en anglais : *proxy server, proxy*] – **Signature de virus** [terme en anglais : *virus signature*] – **Signet** [terme en anglais : *bookmark*] – **Site** [terme en anglais : *Web site*] – **SMS (Short Message Service)** – **Spammeur** – **Système d’adressage par domaines** [terme en anglais : *Domain Name System (DNS)*] – **Système d’exploitation** – **Téléchargement** [termes en anglais : *downloading* (téléchargement à partir d’un autre ordinateur), *uploading* (téléchargement vers un autre ordinateur)] – **Témoin (de connexion)** [terme en anglais : *cookie*] – **Toile (mondiale)** [terme en anglais : *World Wide Web (WWW)*] – **Traitement de données** – **URL (Uniform Resource Locator)** – **Ver** [terme en anglais : *worm*] – **Virus** [terme en anglais : *virus*] – **Visionneur** [terme en anglais : *viewer*] – **Visite** [terme en anglais : *visit*] – **Web**

Le tableau est frappant, nous semble-t-il. À titre d’illustration, voici l’entrée *cookie* de ce lexique.

Un cookie est un enregistrement d’informations par le serveur dans un fichier texte situé sur l’ordinateur client (le vôtre), informations que ce même serveur (et lui seul) peut aller relire et modifier ultérieurement. La technique des cookies repose sur le protocole HTTP, c’est-à-dire le protocole du Web. Il ne faut donc pas voir de cookies partout – seul un serveur Web peut en envoyer. Plus précisément, un cookie se compose d’un ensemble de variables (ou de champs) que le client et le serveur s’échangent lors de transactions HTTP, lesquelles variables sont tout simplement stockées sur la machine cliente dans un simple fichier texte. Un cookie est obligatoirement rattaché à un nom de domaine et un ensemble d’URL de telle sorte que seule une requête provenant du même serveur pourra y accéder. Par exemple, grâce à un programme CGI, le serveur a la possibilité de mettre à jour ou d’effacer un cookie. Mais pour cela, il doit spécifier tous les attributs du cookie, par conséquent seul le serveur qui a créé un cookie peut le modifier ou le supprimer. Un même « client » peut stocker un maximum de 300 cookies, dont 20 maximum pour un même serveur, chaque cookie pouvant atteindre jusqu’à 4000 octets (environ 4 Ko).

Il est important de souligner la valeur *fonctionnelle* – juriste spécialisé – des connaissances du monde « Informatique & Internet » visées. Pour emprunter un ultime exemple à un autre auteur (Larrieu, 2005, pp. 51-52), la question de la protection des bases de données requiert non seulement une connaissance naïve, commune, de cette notion, mais, en l’espèce, suscite une définition légale que le code de la propriété intellectuelle, en son article L112-3, formule en ces termes.

On entend par base de données un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen.

Subtilement, la loi distingue en fait *trois* objets distincts de droits²⁶ : les *données* incluses dans la base ; la *structure* de la base ; les *informations* contenues dans la base, soit la « substance informationnelle » des données. Cette dernière est protégée par un droit *sui generis* à propos duquel Jacques Larrieu apporte ce commentaire.

Le droit *sui generis* protège le contenu de la base, sa **substance informationnelle**, pas son architecture générale (...). Entre-t-il en concurrence avec le droit reconnu au créateur de la donnée ? Non, car il n'a pas le même objet. « Il saisit la donnée, l'information dans sa substance » [écrivent Michel Vivant et Christian Le Stanc] Il s'ensuit que, paradoxalement, le producteur va pouvoir protéger les informations contenues dans les données, alors que l'auteur des données n'en a pas le pouvoir. (p. 58)

Ainsi le « droit des bases de données » exige-t-il une réflexion appropriée sur son objet même, sans que des frontières praxéologiques intangibles puissent jamais être tracées de façon congrue : ce qu'il faut savoir du monde « Informatique & Internet » dépendra indéfiniment du projet dans lequel on s'inscrit, qu'il s'agisse de dire le droit ou de rechercher de l'information sur Internet. C'est cette « leçon » que l'on retrouvera dans la suite des chapitres de cette troisième partie de notre travail.

CHAPITRE 10

Mathématiques, physique et technologies numériques

On ne peut aborder le domaine des mathématiques, de la physique et des technologies numériques sans s'arrêter un instant sur un petit nombre de considérations qui valent pour *l'ensemble* des domaines de connaissance que nous parcourons rapidement ici. Une première considération a, pour l'essentiel, déjà été mentionnée en diverses façons : elle a trait aux connaissances relevant du domaine examiné et qui *pourraient* se révéler pertinentes du point de vue de l'économie et de l'écologie des praxéologies de la RII. De ce point de vue, une attention particulière doit être portée à ce qu'on peut nommer, faute de mieux, des *connaissances* (tenues pour) *critiques*, éléments praxéologiques « ancillaires » (ou auxiliaires, même quand elles ont un caractère fondamental), relevant de domaines autres que celui de la RII *stricto sensu* et qui modifient significativement l'écologie et l'économie du rapport de l'utilisateur tant aux techniques qu'aux composants technologico-théoriques de la RII. Pour le dire en termes naïfs, il s'agit de repérer ces situations dans lesquelles d'aucuns considèrent qu'*un peu plus* de savoir « périphérique » par rapport à la RII permettrait à l'utilisateur d'améliorer *sensiblement* son « confort praxéologique » en matière de RII, par contraste avec ces situations sans doute plus courantes où même *beaucoup* de savoir *en plus* ne paraît guère accroître le bien-être praxéologique espéré. Bien entendu, à beaucoup d'acteurs du domaine de la RII, nombre de situations sembleront ne relever ni de l'un ni de l'autre extrêmes, tombant ainsi dans une zone grise où, en particulier le choix de regarder comme utile ou peu utile telle connaissance dépendra de considérations adventices.

Une seconde considération découle de ce qui précède. Ainsi qu'on a commencé de le faire, la problématique que l'on vient de rappeler pousse à envisager comme recelant d'éventuels éléments praxéologiques *critiques* de vastes domaines de connaissance mettant en jeu de nombreuses « spécialités » (au sens académique du terme). Nous en assumerons pourtant, ici, un abord résolument *généraliste* – quoique non exclusif, bien sûr, de contrôles plus « spécialisés » – afin d'être le mieux possible à l'écoute de la pluralité ouverte d'institutions et de personnes dont nous avons commencé à explorer l'équipement praxéologique en matière de RII.

Des mathématiques dans la RII ?

Qu'est-il donc utile de savoir, ainsi, à qui pratique la RII, en matière de mathématiques, de physique et de technologies numériques ? Pour ce qui est des mathématiques, nous nous limiterons ici à considérer un point de difficulté à propos de l'algorithme PageRank de Google. Dans leur article *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (1998), Page et Brin écrivent – en italique – ceci.

We assume page A has pages $T1...Tn$ which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. There are more details about d in the next section. Also $C(A)$ is defined as the number of links going out of page A. The PageRank of a page A is given as follows:

$$PR(A) = (1 - d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

On voit que si la page A pointe vers elle-même, l'une des pages $T1, T2, ..., Tn$ n'est autre que la page A elle-même ; en d'autres termes, le nombre $PR(A)$ est défini en fonction de lui-même (et d'autres nombres encore, bien sûr). Or beaucoup de personnes, en dépit de l'instruction mathématique prodiguée au collège, sont étonnées quand on attire leur attention sur le fait que dans l'équation $x = 3x - 16$, le nombre x apparaît défini *en fonction de lui-même*, et plus précisément, ici, comme étant égal à 3 fois lui-même, diminué de 16, au contraire de ce qui se passerait si l'on avait l'égalité $x = 3 \times 8 - 16$, égalité qui est une équation « toute résolue ». Découvrir qu'il est tout de même possible de définir x de cette façon que les mathématiciens appellent *implicite* (par opposition à la définition *explicite* que fournit l'égalité $x = 3 \times 8 - 16$) constitue une conquête qui devrait être réalisée au collège. Or pour bien des usagers ordinaires de Google qui, par extraordinaire, seraient mis en contact avec la formule

$$PR(A) = (1 - d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn)),$$

une telle conquête n'a jamais été faite, ou bien ses fruits en ont été perdus.

Le lecteur attentif à ce détail aura peut-être observé que, dans la formule donnée par Brin et Page dans une présentation du PageRank citée dans la première partie de ce travail (et reproduite par Barbara Cassin à la page 95 de son livre sur Google), à savoir

$$W_j = (1 - d) + d \sum_{i=1, i \neq j}^N l_{ij} \frac{W_i}{n_i},$$

le PageRank (PR) de la page j dépend éventuellement du PR de toutes les pages Web *autre qu'elle-même* : c'est là en effet le sens de l'indication selon laquelle on doit avoir $i \neq j$ dans la sommation appelée par le symbole \sum : i prend toutes les valeurs entières de 1 à N , hormis la valeur j . Pourtant, même dans ce cas, il y a apparemment une difficulté. Supposons ainsi que l'on cherche trois nombres x, y, z liés par les égalités suivantes :

$$\begin{cases} x = 3y - z \\ y = 7x - z \\ z = x + 2y \end{cases}$$

Ici, chacun des nombres x, y, z est « défini » en fonction des deux autres, et le système de trois équations à trois inconnues ci-dessus apparaît donc à nouveau comme fournissant « au mieux » une définition *implicite* de x, y et z . Si l'on remplace dans les deux premières égalités le nombre z par son expression en fonction de x et de y (expression que fournit la troisième égalité), il vient ceci.

$$\begin{cases} x = 3y - (x + 2y) \\ y = 7x - (x + 2y) \end{cases} \text{ soit encore } \begin{cases} x = y - x \\ y = 6x - 2y \end{cases}$$

Ici, chacun des nombres x et y est défini *en fonction de l'autre* : la définition reste implicite. Poursuivons. En ramenant dans le premier membre le terme en x pour la première égalité, en y pour la seconde, on obtient

$$\begin{cases} 2x = y \\ 3y = 6x \end{cases} \text{ soit encore } \begin{cases} x = \frac{1}{2}y \\ y = 2x \end{cases},$$

ce qui reste apparemment « circulaire » mais se réduit à la seule égalité $y = 2x$ (puisque de celle-ci il découle que $x = \frac{1}{2}y$). Comme $z = x + 2y$, on a $z = x + 4x = 5x$. Si l'on ajoute alors la condition de « normalisation » $x + y + z = 1$, on obtient l'équation $x + 2x + 5x = 1$, soit $8x = 1$, ce qui donne $x = \frac{1}{8} = 0,125$, $y = 2x = \frac{1}{4} = 0,25$ et $z = 5x = \frac{5}{8} = 0,625$.

D'une façon similaire, quoique à une autre échelle (au lieu de trois nombres x, y et z , il s'agit d'explicitier plusieurs milliards de nombres), on a un phénomène analogue pour « définir » (explicitement) le PageRank des pages du Web. Or cette difficulté, dont un

outillage mathématique adéquat permet classiquement de disposer (Chevallard, 2008), peut apparaître, aux yeux de l'utilisateur peu familier du passage d'une définition mathématique implicite à une définition explicite, comme une *pure impossibilité*, ce qui peut le conduire à douter de la véracité des affirmations sur les principes qui seraient au cœur du classement des résultats affichés par Google, et donc à entrer dans une logique de dubitation – sinon de suspicion pure et simple – en matière de RII. On notera, par contraste, que nombre de non-spécialistes des moteurs de recherche sont portés à « expliquer » les classements de résultats par le *nombre de visites* reçues par les pages concernées, un indicateur qui a le mérite implicite d'être défini de façon *explicite*. On a là un symptôme parmi d'autres de la difficulté à passer une frontière – celle qui court entre définitions explicites et définitions implicites – que chacun a pourtant passée au cours de ses études secondaires mais sans en général en être conscient, et dont beaucoup se sont éloignés pour se lover dans le confort cognitif des seules définitions explicites.

Les bases physiques du numérique

Dans un texte déjà mentionné, sous le titre « Le circuit intégré, moteur du numérique », Gérard Berry note ce qui suit – que complète une note infrapaginale signalant le rôle « des avancées fondamentales en physique qui ont donné le prix Nobel 2007 à Albert Fert et Peter Grünberg ».

Les trois composants matériels du monde numérique sont les circuits électroniques, les mémoires de masse (disque, flash, etc.) et les réseaux. Les mémoires de masse croissent à une vitesse fantastique. Les disques de micro-ordinateurs sont passés de quelques méga-octets à quelques centaines de giga-octets en une vingtaine d'années, soit un gain d'un facteur 100 000, et les mémoires flash s'insinuent partout. Avec la numérisation des télécommunications, la généralisation du transport des bits par fibres optiques et l'introduction d'Internet, les réseaux ont également connu une explosion exponentielle. Mais nous nous limiterons ici au composant clef qu'est le circuit électronique, qu'il faut regarder à plusieurs échelles de structure et de comportement. (pp. 32-33)

Que faut-il savoir de ce qui est évoqué ici ? À propos des circuits électroniques, l'auteur écrira un peu plus loin ceci ²⁷.

Mettre le plus grand nombre possible de composants à la plus grande fréquence raisonnable est la clef de l'augmentation de la puissance de calcul et de la diminution des coûts. On dispose à l'heure

actuelle de dizaines ou centaines de millions de transistors opérant à des fréquences de quelques centaines de mégahertz pour les circuits des objets numériques nomades, et de milliards de transistors à quelques gigahertz pour les gros microprocesseurs. Ces chiffres paraissent astronomiques, mais ils sont encore au-dessous des besoins : il faut plusieurs téra-opérations par seconde pour faire une belle image en télévision haute définition, ce qui paraît pourtant simple au novice ! (pp. 34-35)

Est-il utile de comprendre tout cela et de pouvoir s'y référer de façon opérationnelle ? C'est là une question qui concrétise la problématique déjà exposée²⁸. On notera à cet égard que le genre de connaissances présentées ou évoquées ici paraît susceptible d'accroître le confort *technologico-théorique* de l'utilisateur plutôt que son bien-être proprement *technique* : cela lui permettrait de *mieux comprendre* plutôt que de *mieux faire*. On peut toutefois imaginer que, selon une dynamique praxéologique jamais tout à fait inhibée, la connaissance d'éléments technologico-théoriques conduisent à améliorer des techniques existantes et pas seulement à donner une meilleure intelligibilité à ces techniques, distinction à laquelle il convient également d'être attentif.

Le cas de l'électricité

À titre d'illustration nous nous arrêterons brièvement sur un petit livre récemment paru (de Clerck, 2007), intitulé *Les dessous de l'Internet* et sous-titré de façon un rien énigmatique *Au fil de l'électricité, histoire de comprendre* – il s'agirait donc moins d'apprendre à *faire* que de se donner les moyens de *comprendre*. Il pourra paraître surprenant que, pour « expliquer » l'Internet, l'auteur songe à remonter jusqu'à un fait certes essentiel mais nullement spécifique de l'Internet : l'existence et la disponibilité de l'électricité – ainsi que de bien d'autres choses encore. Le texte figurant sur la quatrième de couverture est à cet égard éclairant ; le voici.

On n'y prend garde, mais si Internet fonctionne c'est d'abord grâce à l'électricité !

Et c'est en ajoutant à l'électricité un brin d'électronique, pas mal d'informatique et une bonne dose de télécommunications que l'on a abouti à cette formidable invention.

Ces technologies font aujourd'hui partie de notre vie courante, mais nous les utilisons souvent sans en connaître le fonctionnement. Or si dans le détail elles peuvent parfois se révéler complexes, les principes sont largement compréhensibles de tous, sans aucune formation scientifique.

C'est à un rappel analogue que procède Gérard Berry dans sa leçon inaugurale au Collège de France déjà citée (Berry, 2008).

... il faudra réduire de façon majeure la consommation d'énergie des dispositifs numériques. On perçoit bien que l'autonomie des appareils à pile est un problème, tant pour le confort que pour la pollution due aux batteries ; mais on sait moins que les ordinateurs branchés au secteur consomment une proportion significative et rapidement croissante de l'électricité produite. De grandes sociétés installent déjà leurs fermes de serveurs Internet près des barrages ou des centrales électriques pour disposer d'assez de courant. Diminuer l'énergie consommée est devenu le problème majeur pour la conception des circuits, des méga-ordinateurs et des méga-fermes de calcul. (p. 73)

Réfléchir – par exemple dans un souci économique et environnemental – au problème évoqué ici peut conduire l'utilisateur isolé d'un ordinateur connecté à l'Internet à « découvrir » que certaines phases d'une RII peuvent se dérouler « loin de l'ordinateur ». Ainsi pourra-t-il décider de préparer ses requêtes, de réfléchir aux moteurs de recherche qu'il envisagera de solliciter, aux sites généralistes ou spécialisés qu'il pensera bon de visiter, à certains blogs aussi qu'il lui paraîtra *a priori* judicieux de ne pas oublier d'examiner, etc., *sans allumer son ordinateur*. En un tel cas, il n'y aura pas, certes, de lien *direct* entre la connaissance (et la conscience) du *fait physique* souligné plus haut et l'évolution d'une *technique personnelle de RII*, mais bien un lien *génétique* – l'idée de modifier sa technique de RII ayant pu un jour découler d'une prise de conscience d'un phénomène physico-économique : le coût énergétique global du « tout à l'ordinateur ». La RII se sera ainsi révélée « robuste », en imagination au moins, à un changement – réduire la consommation d'énergie électrique – qui aura été, en l'espèce, évoqué plutôt que vécu.

La raison de la proximité – évoquée ci-dessus – des fermes de serveurs par rapport aux lieux de production de l'électricité peut demeurer mystérieuse pour qui a oublié le phénomène de perte en ligne : si la puissance électrique P (mesurée en watts) est transportée par dans un fil de résistance R avec une intensité I , la perte de puissance p est donnée par la formule $p = RI^2$. Si la tension (voltage) est V , l'intensité I est donnée par $I = \frac{P}{V}$: elle est donc inversement proportionnelle à la tension – plus la tension est élevée, plus l'intensité est faible. On a ainsi : $p = RI^2 = R \frac{P^2}{V^2} = \frac{RP^2}{V^2}$: la perte en ligne est inversement proportionnelle au carré de la tension électrique (et proportionnelle à la résistance). Pour diminuer les pertes en ligne, on est donc conduit à augmenter la

tension (et à diminuer la résistance du fil) : ainsi passera-t-on (grâce à des transformateurs) de la très haute tension à la sortie des centrales (THT : 400 kV), à la haute tension (HT, moins de 100 kV), puis à la moyenne tension (MT, de 5 kV à 30 kV), enfin à la basse tension (BT, de 230 V à 380 V) fournie aux utilisateurs.

L'induction électromagnétique

Le petit ouvrage d'Antoine de Clerck fait tout démarrer avec l'avènement de l'électricité. Chacun des chapitres de son petit livre est divisé en trois sections : « Pour les curieux », « Pour les passionnés », « Pour les experts ». En voici le sommaire (de Clerck, pp. 5-6).

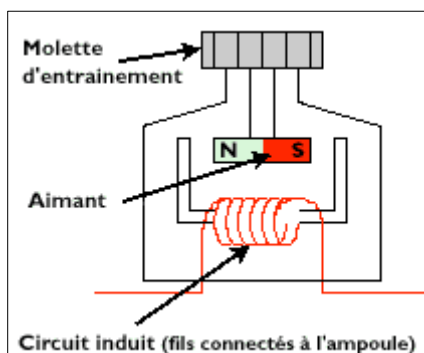
1. Au commencement était l'électricité
2. L'électronique peut tout
3. La magie des ondes
4. Les télécommunications, loin et près
5. Les ordinateurs, un concentré d'intelligence
6. Internet, un bel avenir

Conclusion

Chronologie

Glossaire

Bien entendu, ainsi que ce sommaire le suggère, la possibilité *physique* de la RII ne dépend pas que de l'emploi de l'électricité : on aura noté par exemple que l'*électronique* constitue un sous-domaine clé de notre affaire. Nous nous arrêterons cependant quelques instants sur le cas de ce que l'auteur nomme « le phénomène de l'induction », auquel il consacre tout un encadré éclairant (p. 13). L'illustration la plus simple en est classiquement fournie par l'alternateur de bicyclette, tel que le donne à voir le schéma ci-contre, que nous empruntons au site consacré



au physicien Louis Néel (1904-2000), lauréat en 1970 du prix Nobel de physique²⁹. Lorsque la *molette* est entraînée par la roue de la bicyclette, elle fait tourner l'*aimant* qui lui est assujéti. Le mouvement de l'aimant *induit* alors un courant électrique dans la *bobine* : tel est le phénomène clé. Le texte qui accompagne le schéma précise ceci : « La bobine est reliée à une ampoule qui s'allume grâce au courant induit par la rotation de l'aimant. À noter : c'est

sur ce principe qu'est produit le courant alternatif par EDF ! » Ici, tout un univers se joue : l'énergie mécanique de la roue qui fait tourner l'aimant peut être remplacée par la force de l'eau, du vent, etc.

L'*induction électromagnétique* – tel est le nom « complet » – n'est pas absente de la formation scolaire donnée dans la classe de physique-chimie : sans que son nom y figure, le phénomène apparaît actuellement dans le programme de la classe de troisième, qui mentionne d'ailleurs en passant l'alternateur de bicyclette³⁰. Le document qui accompagne ce programme comporte une sous-section intitulée « La production industrielle d'électricité » qui propose, à la suite d'un « texte destiné aux professeurs », un « texte destiné aux élèves » précédé de la remarque que, « en même temps que ce texte, on peut présenter aux élèves différentes illustrations de machines d'époque » ; nous le reproduisons ici.

En 1831, l'Anglais Michael Faraday découvre le phénomène d'induction, c'est-à-dire la création d'un courant par un aimant mobile. Dès lors on sait convertir l'énergie mécanique en énergie électrique. Dès 1832, Ampère fait construire par son compatriote Hippolyte Pixii une machine à manivelle faisant tourner un aimant en U (inducteur) devant deux bobines (induit), produisant ainsi un courant alternatif. Adjoignant ensuite à sa machine un « commutateur d'Ampère », Pixii produit un courant continu. Ensuite, les physiciens améliorent la qualité et la puissance des générateurs électriques. Par exemple, ils remplacent les aimants par des électroaimants : ce sont des bobines parcourues par un courant électrique qui se comportent comme des aimants, mais avec un champ magnétique qui peut être très fort si la bobine comporte beaucoup de spires et si l'intensité du courant est élevée. En 1871, le Belge Zénobe Gramme fait construire une machine entraînée par un moteur à vapeur, dont on peut dire qu'elle est la première à pouvoir produire industriellement de l'électricité. En 1881, une grande exposition est organisée à Paris pour présenter les différents appareils capables de produire, de transporter et d'utiliser l'électricité. Elle accueille près d'un million de visiteurs. On y voit aussi bien des générateurs de courant alternatif que des générateurs de courant continu, même si ces derniers ont la faveur des utilisateurs de l'époque (la machine de Gramme, par exemple, est à courant continu). Cependant, tous les constructeurs rencontrent des difficultés importantes pour transporter l'énergie électrique : en effet la tension chute rapidement à mesure que la longueur de la ligne augmente, à cause des pertes par effet Joule. C'est entre les années 1880 et 1890 qu'est découvert un appareil appelé « transformateur ». Cet appareil, qui ne fonctionne qu'en courant alternatif, permet à volonté d'élever ou d'abaisser la tension électrique. Cela est très intéressant car, pour une énergie transportée donnée, plus la tension est élevée, plus les pertes par effet Joule sont réduites. Le transformateur assure le succès du courant alternatif face au

courant continu. Le domaine de la construction d'appareils électriques, et en particulier de générateurs, est devenu un enjeu économique majeur : dans toute l'Europe et aux États-Unis, de nombreux constructeurs se livrent une concurrence sévère. C'est au début du XX^e siècle qu'apparaissent sur le marché la plupart des appareils électriques ménagers : grille-pain, fer à repasser, machine à coudre, ventilateur, sonnette, etc. En 1913, le réseau français de distribution est encore très morcelé : la production est faite par un grand nombre de petites usines, et le courant distribué sur des distances assez faibles ; à cette date, seules 6 000 communes, sur 36 000 environ, sont équipées. En 1936, la plupart des communes de France reçoivent l'électricité. En avril 1946, une grande part des usines de production et des réseaux de distribution sont regroupés dans la société nationale électricité de France.

Sans aller plus loin sur cette question, on doit conclure que le « phénomène d'induction » – que l'auteur invoquera à nouveau pour expliquer « la transmission du son par l'électricité » (pp. 26-27) – est un fait physique « bien connu ». Sa « redécouverte », voire sa rencontre inaugurale (pour qui aurait échappé à une première rencontre scolaire) peut alors provoquer ce que nous nommerons un *effet Baruch*, où l'enthousiasme du « découvreur » peut faire un contraste irritant avec l'attitude plus indifférente des « demi-savants », sinon des vrais savants ; ce n'est pas là le prix le plus cher à payer pour une culture adéquate au service de notre relation au monde « Informatique & Internet » et à la RII en particulier.

Le nom d'effet Baruch est liée à un épisode connu de la vie littéraire du Grand Siècle, que, dans sa biographie de Jean de La Fontaine, Roger Duchêne rapporte en ces termes (Duchêne, 1990) : « Dès qu'il [La Fontaine] avait pris goût pour quelque ouvrage, raconte le fils de Jean Racine, son esprit en était entièrement occupé. Il en parlait à tout propos. Racine, qui l'avait un jour emmené à l'office des Ténèbres, s'aperçut qu'il s'ennuyait fort. Il lui donna une Bible pour le distraire. La Fontaine tomba sur Baruch, un des petits prophètes de l'Ancien Testament. Il y lut la prière des Juifs, et il la trouva admirable. Il le dit à Racine en lui demandant qui était ce Baruch. Puis, pendant plusieurs jours, à tous ceux de sa connaissance qu'il rencontrait, il disait seulement : "Avez-vous lu Baruch ? C'est un fort grand génie." Son enthousiasme lui ôtait toute mesure. »

Si nous soulignons ici un tel phénomène, c'est que, d'une façon générale, le repérage des connaissances critiques qu'impose l'étude des praxéologies de la RII, de leur économie et de

leur écologie, ne doit en aucune façon en être gêné, et cela d'autant moins qu'il s'agit précisément de repérer des connaissances éventuellement adéquates au projet praxéologique de personnes ou d'institutions que celles-ci sont portées à censurer, voire à refouler, au motif que de telles connaissances « appartiendraient » à d'autres institutions et, pour cela, à d'autres personnes. Dans l'épistémologie *fonctionnelle* où nous nous situons, une connaissance vaut d'abord et surtout *par sa valeur praxéologique* – par ce qu'elle permet de faire et de comprendre –, choix qui peut entrer en conflit avec d'autres, liés notamment à la compétition entre institutions en vue d'un usage monopolistique (ou oligopolistique) des connaissances. Par contraste, l'indifférence au « pedigree » (sinon à la genèse) de telle ou telle connaissance sera pour nous la règle.

À propos de disques durs

L'analyse des besoins praxéologiques des institutions et des personnes en matière de RII se heurte d'une manière générale à un ordre épistémologique établi, constitué par des transpositions institutionnelles parfois fort anciennes, qui ont sélectionné des fragments de praxéologies en fonction de conditions et de contraintes devenues parfois largement étrangères à la fonctionnalité appelée par les besoins actuels de connaissance et d'action. Cette situation assez générale conduit à se défier des évidences institutionnelles ou personnelles en la matière, ce qui impose déjà de les *repérer*. Ainsi ne va-t-il pas de soi de regarder même conjecturalement comme utile, sinon indispensable, d'intégrer la connaissance du « phénomène d'induction » aux praxéologies de la RII, notamment parce que les évidences culturelles courantes ne rangent pas dans la même catégorie les deux réalités évoquées. Pourtant, en matière causale, et donc aussi au plan de l'intelligibilité, l'évident n'est jamais sûr, comme le suggère le petit exemple suivant.

L'information que l'on recherche sur Internet est située sur des *disques durs*. Dans un article paru dans le quotidien *Le Monde* le 26 janvier 2008, sous le titre « Attention, l'humanité perd la mémoire », Franck Laloë, directeur de recherche émérite au CNRS, écrit : « ... les faits sont là : tout patrimoine numérique abandonné à lui-même, ne fût-ce que cinq ou dix ans, risque d'être définitivement perdu. » Et il note en particulier ce qui suit.

Les disques durs, eux, reposent sur un principe physique un peu analogue aux bandes magnétiques : si leur rapidité d'accès est indéniable, pour la longévité, ils ne font guère mieux. De fait, jamais, dans toute son histoire, l'humanité n'a utilisé de techniques aussi instables pour enregistrer ses données.

Cette observation paraîtra inattendue à beaucoup – une réaction qui, au demeurant, a sans doute en partie motivé la rédaction de l'article cité ! Elle tend à faire apparaître « en creux » ce qui est sans doute une réalité pour nombre d'utilisateurs des produits informatiques : un disque dur est une *boîte noire* dans laquelle l'ordinateur – ou plus exactement son système d'exploitation – stocke des données de façon implacable, sauf accident majeur. Or, cette vision, nous dit-on ici, mérite tout simplement d'être répudiée. Le physicien cité écrit encore ceci.

... les supports sur lesquels est inscrite l'information numérique sont constamment rongés de l'intérieur par le temps ; même si on conservait tout ce qu'il faut pour les lire, l'information disparaîtrait tout aussi inexorablement ! Les bandes magnétiques vieillissent en une dizaine d'années, et la seule façon de conserver les informations est de les recopier sur une bande plus neuve, etc., *ad infinitum*. C'est ce que font les grandes bibliothèques pour conserver leurs données ; mais si la recopie est oubliée pendant un certain temps, tout est perdu.

Sans doute faut-il, pour intégrer cela, au moins *entrouvrir* la boîte noire qu'est – le cas échéant – un disque dur. Dans un livre intitulé *La physique par les objets quotidiens*, les auteurs, Cédric Ray et Jean-Claude Poizat (2007), consacrent leur chapitre 11 au disque dur (pp. 100-107). Qu'y apprend-on quant à l'enregistrement de données binaires du point de vue physique ? Voici un premier élément de réponse.

Imaginons à présent une surface recouverte d'une couche de matériau ferromagnétique. Initialement, le matériau n'a été exposé à aucun champ magnétique et les petits domaines qui le composent ne présentent aucune aimantation commune. Appliquons un champ magnétique : les domaines vont s'aimanter dans la direction du champ imposé, créant ainsi des zones magnétiques où l'aimantation est unique. Ainsi, pour écrire des données sous forme binaire (...), il suffit de disposer d'un dispositif capable de produire un champ magnétique de direction constante mais dont on peut choisir le sens. En imposant l'un des deux sens à différentes zones de la couche de matériau magnétique, on y enregistre les données, le 0 correspondant à un sens et le 1 à l'autre... (p. 102)

Tel est le principe ; voici maintenant le mécanisme plus concret par lequel ce principe se réalise.

Comment produire un champ magnétique ? Par exemple en faisant circuler un courant électrique dans une bobine (le champ magnétique qui en résulte sera plus intense si l'on met un morceau de fer

doux au centre de l'enroulement) : c'est le principe de l'électro-aimant (...). Si l'on inverse le sens du courant dans le bobinage, le champ magnétique change de sens. On peut ainsi définir une orientation d'aimantation spécifique pour chaque zone de la couche en fonction du sens du courant dans l'électro-aimant. C'est le principe du système d'écriture adopté pour tous les disques durs.

(p. 102)

On retrouve ainsi une situation dont l'élément clé nous est « connu » : la relation entre courant électrique et champ magnétique. Il est vrai, sans doute, que la conception et la fabrication des disques durs reposent aujourd'hui sur des prouesses techniques et des avancées technologico-théoriques exceptionnelles.

Les auteurs que nous suivons soulignent notamment le rôle des travaux sur la « magnétorésistance géante » (GMR, en anglais) conduits par Albert Fert (laboratoire de physique des solides, Orsay) et Peter Grünberg (Centre de recherche de Jülic), qui ont pour cela reçu le prix Nobel de physique en 2007. Dans le communiqué où elle présente son choix, l'Académie des sciences royale suédoise (2007) précise : « Le prix Nobel de Physique de cette année récompense la technologie utilisée pour lire l'information stockée sur des disques durs. C'est grâce à cette technologie qu'une miniaturisation radicale des disques durs a été possible ces dernières années. » Dans ce cas, l'intervalle entre la découverte fondamentale (1988) et l'apparition de nouvelles têtes de lecture (1997) aura été remarquablement court.

Malgré cela, il suffit d'un petit nombre de connaissances « critiques » pour donner (ou pour redonner) à ce type d'objets numériques une dimension concrète, physique, qui ouvre la possibilité de s'interroger à leur propos, même en simple utilisateur : aussi peu que ce soit, le « monde numérique » cesse alors d'être saturé d'évidences paresseuses et devient ainsi fructueusement problématique.

CHAPITRE 11

L'informatique

Le domaine de connaissance que nous désignons sous le nom d'*informatique* oblige à soulever une question que nous avons laissée de côté jusque-là. Quand on considère comme domaines de connaissance éventuellement pertinents pour la RII l'histoire, l'économie, le droit ou la physique, on n'imagine pas, semble-t-il, qu'une institution ou une personne confrontée au problème de la RII puisse prétendre que la solution à ce problème suppose la connaissance de *toute* l'histoire, de *toute* l'économie, de *l'ensemble* du droit, de la physique, etc. Spontanément, on s'attend à ce que l'exigence de connaissances soit formulée en disant qu'un bon équipement praxéologique en matière de RII demande de connaître *certaines éléments* d'histoire, d'économie, de droit ou de physique (et de bien d'autres choses encore : on peut penser ici, en passant, à cette partie de la physique appliquée qu'est l'électronique).

En revanche, lorsqu'on en vient à « l'informatique », peut-être parce que l'Internet, et donc le Web, et donc la RII, en sont regardés comme des sous-domaines, la tentation peut exister de poser qu'il convient de connaître « l'informatique » : *toute* l'informatique ! Cela n'a bien entendu guère de sens : on ne saurait pas plus connaître toute l'informatique qu'on ne peut connaître toutes les mathématiques, toute la philosophie, toute la physique, etc. Pourtant, dans une société donnée, la tentation peut être forte d'identifier une telle connaissance supposée quasi « exhaustive » de la matière à ce qu'en doivent savoir, en principe, les « spécialistes » patentés – tels par exemple, dans le cas français, les « agrégés de... » (de mathématiques, de philosophie, de physique, etc.), ou encore les titulaires d'un diplôme universitaire attestant une formation « poussée » dans le domaine considéré. Il arrive que les meilleurs auteurs soient à deux doigts de céder à cette tentation, qui offre l'attrait fondamental de laisser croire que le problème de la formation adéquate serait ainsi résolu en quelque sorte par avance !

C'est ainsi que, souhaitant expliquer la manière dont il est possible, à partir d'un index en texte intégral (*full-text index*), d'ordonner les documents contenant un mot clé donné, ce qui est pour l'essentiel un problème d'algorithmique, les auteurs de l'ouvrage *Web Dragons* fournissent vaillamment une longue explication, qu'ils renoncent toutefois à mener à son

terme, laissant entendre *in fine* que, si le lecteur avait fait des études universitaires complètes de *computer science*, la pièce manquante du puzzle explicatif leur serait tout à fait familière³¹.

Returning results in order of relevance, estimated using a similarity measure, is more challenging. If there are many matches, even sorting the documents into order can be a substantial undertaking. However, you don't need all the results at once, just the top 10 or 20, and there are ways of finding the top few documents without sorting the entire list. But we do need to compute all the similarity values to find the best ones—or do we? Having found a few good ones, there are ways of quickly ruling out most of the others on the basis of a partial computation. *These are things you might learn in a graduate computer science course.* (p. 110)

Le souhait subreptice, peut-être pas uniquement rhétorique, des auteurs touchant la formation informatique *préalable* de leur lecteur s'exprimera à nouveau quelques pages plus loin, en cet autre passage tout aussi significatif (p. 126) : "How search engines deliver the goods is one of the marvels of our world. Full-text search is an advanced technology, *the kind you learn about in computer science graduate school*".

Par contraste, dans l'observation et l'analyse auxquelles nous avons à procéder, nous nous garderons de toute problématique paresseuse et illusoire de la *totalisation* – qu'il est loisible aux auteurs dont nous observons les « discours praxéologiques » de faire leur. Dans ce travail, rappelons-le encore une fois, nous nous donnons au contraire pour horizon une problématique de l'*élémentation* des domaines de connaissance, dans laquelle la règle est de tenter de sélectionner, dans tous les domaines réputés pertinents, les connaissances qui, adéquatement organisés, constitueront les *éléments* de la science de la RII visée.

Théorie, pratique : une tension indépassable ?

Que retient-on, dans les diverses institutions préoccupées par la RII, en fait de connaissances informatiques « utiles » ? Comme à d'autres propos, nous nous limiterons ici à ébaucher l'étude de cette question. Le domaine informatique est immense, et objectivement divisé en de nombreux secteurs plus ou moins apparentés. Du point de vue qui est le nôtre ici, il faut surtout souligner l'existence d'une tension entre un pôle conceptuel, théorique, qui se présente comme réservé à une élite, et un pôle tourné vers le pratico-pratique, dont le point culminant est occupé par la figure mythifiée du « bidouilleur de génie ».

Le pôle conceptuel est sans doute bien représenté par la haute figure d'Edsger Dijkstra (1930-2002), dont nombre de « mots » sont restés célèbres – le plus célèbre d'entre eux peut-être étant celui-ci : *Computer science is no more about computers than astronomy is about telescopes*. Dijkstra s'illustra notamment par son rejet sans nuance du langage de programmation BASIC (qui naît en 1964), écrivant à cet égard ³² : “*the teaching of BASIC should be rated as a criminal offence: it mutilates the mind beyond recovery*”. Son combat contre l'instruction GOTO, en particulier, est rapporté dans les termes suivants dans l'article « Edsger Dijkstra » de l'encyclopédie Wikipédia ³³.

Constatant les dégâts provoqués par l'usage incontrôlé de l'instruction *goto* en programmation, il rédige en 1968 pour les *Communications of the ACM* un article qu'il nomme *A case against the GOTO statement* (« Un procès contre l'instruction GOTO »). Voulant publier rapidement l'article sous la forme d'une lettre à l'éditeur, l'éditeur Niklaus Wirth le rebaptise *Go To Statement Considered Harmful* (« L'instruction Go To jugée nuisible »). Ce nouveau titre autant que le propos de l'article devient alors célèbre dans le milieu de l'informatique. Les titres de la forme *X considered harmful* se multiplient, jusqu'à un *Dijkstra considered harmful*. En outre, l'instruction *goto* est effectivement supplantée par la programmation structurée (concept qu'on lui doit et présenté entre autres dans EWD 268). En programmation structurée, le *goto* est remplacée par des instructions comme *if... then... else..., while... do, repeat... until*, chacune contenant *une seule* entrée et *une seule* sortie.

L'orientation « théoricien » en informatique, qui est indéniablement à la base de la plupart des grandes avancées de l'informatique comme discipline intellectuelle, semble ainsi s'être développée dans une tension constante avec les « praticiens » aux fabuleux chiffres d'affaires ³⁴. L'article « Informatique » de l'encyclopédie Wikipédia illustre la chose dans cet autre passage.

L'émergence d'un aspect réellement scientifique dans la programmation elle-même (et non dans les seules applications scientifiques que l'on programme) ne se manifeste qu'avec la série *The Art of Computer Programming* de Donald Knuth, professeur à l'Université de Stanford, à la fin des années 1960, travail monumental encore inachevé en 2004. Les travaux d'Edsger Dijkstra, Niklaus Wirth et Christopher Strachey procèdent d'une approche également très systématique et elle aussi quantifiée. On demandait à Donald Knuth dans les années 1980 s'il valait mieux selon lui rattacher l'informatique (*computer science*) au génie électrique – ce qui est souvent le cas dans les universités

américaines – ou à un département de mathématiques. Il répondit : « *Je la classerais volontiers entre la plomberie et le dépannage automobile* » pour souligner le côté encore artisanal de cette jeune science.

Toutefois, la forte scientificité des trois premiers volumes de son encyclopédie suggère qu'il s'agit là plutôt d'une *boutade* de sa part. Au demeurant, la maîtrise de langages comme Haskell, Ocaml ou même APL demande un niveau d'abstraction tout de même plus proche de celui des mathématiques que des deux disciplines citées.

Même si l'univers de discours évoqué ici paraît bien éloigné des problèmes de RII que doit affronter l'utilisateur ordinaire, il est sans doute prudent, au plan épistémologique, de demeurer attentif aux types d'exigences qui s'y font entendre. Ce n'est pas un hasard si, dans sa leçon inaugurale déjà citée, Gérard Berry, note *in fine*, pour le regretter (p. 96) : « ... enseigner l'informatique consiste souvent à enseigner l'usage d'un clavier, d'une souris et d'un traitement de texte dans un cours de technologie, choses à contenu conceptuel très faible. » En vérité, c'est énoncer là, en creux, l'un des grands problèmes du processus d'élémentation auquel nous nous intéressons : le problème de la faiblesse de la « profondeur de champ technologico-théorique » des équipements praxéologiques les plus souvent envisagés.

Matériel et logiciel

En dépit de la mise en garde proférée par Dijkstra, l'entrée dans le « continent informatique » proposée aux usagers potentiels ou débutants du Web s'articule à cet objet, apparemment indépassable à ce niveau, qu'est l'ordinateur, ou, plus précisément, le *personal computer* – le PC. On se repérera ici tout d'abord à l'aide de l'exposé épuré qu'Antoine de Clerck propose dans le chapitre 5, intitulé « Les ordinateurs, un concentré d'intelligence », de son livre *Les dessous d'Internet*. Comme chacun des chapitres du livre, celui-ci est divisé en trois sections : « Pour les curieux », « Pour les passionnés », « Pour les experts ». La première section du chapitre 5 donne une information qui se veut minimale : après une ouverture consacrée à la distinction entre « calculateur » et « ordinateur », écho affaibli des problèmes que soulèvent l'usage en anglais du mot *computer*, elle conduit l'auteur à tracer pour le lecteur un premier croquis utile.

... les ordinateurs sont constitués d'une part de circuits électroniques qui forment « l'intelligence » de la machine, et d'autre part d'un clavier et d'un écran qui en permettent le pilotage. Pour exploiter un ordinateur il est également nécessaire d'utiliser des logiciels, programmes informatiques qui

pilotent l'ordinateur dans la réalisation d'une fonction. Par exemple, un logiciel de « traitement de texte » permettra à l'ordinateur de réaliser la fonction de machine à écrire. Un logiciel de conception d'images permettra à l'utilisateur de réaliser des dessins. (p. 87)

Cette section comporte également un développement sur les *réseaux* d'ordinateurs : il s'agit-là, aujourd'hui, d'un « must », même dans les présentations les plus rudimentaires. Mais au-delà du simple curieux se trouve le « passionné » : c'est à son intention que l'auteur présente alors la distinction du « hard » et du « soft » (p. 88). Le premier – le *hardware* – est fait de « l'ensemble des circuits et composants électroniques internes », que l'on trouve « généralement assemblés dans une *unité centrale* », à quoi s'ajoutent l'écran et le clavier ainsi que des éléments complémentaires (souris, imprimante, lecteur de CD, scanner) appelés *périphériques* ; le second – le *software* – est constitué des programmes informatiques qui permettent d'utiliser le « hard », tels un traitement de texte ou un logiciel de conception d'images. *Hardware* et *software*, fait observer l'auteur, relèvent d'industries généralement distinctes : IBM, Hewlett Packard ou Dell fabriquent des ordinateurs, tandis que Microsoft, dont la fortune repose sur le succès du système d'exploitation Windows, se consacre aux logiciels. Le « passionné » aura droit à quelques autres développements encore – sur les documents mémorisés par l'ordinateur et classés dans des fichiers, sur les CD-ROM (*Compact Disc – Read Only Memory*), les clés USB (*Universal Serial Bus*), etc. L'auteur passe alors rapidement en revue différentes utilisations des ordinateurs (à la maison, au bureau, dans les universités et les centres de recherche, dans les bureaux d'études et dans les usines) avant de revenir sur la question des *réseaux* informatiques, qui permettent de partager des *périphériques* (une imprimante par exemple), et qui sont au cœur de l'Internet, « immense réseau informatique à l'échelle mondiale » (p. 90). La section se termine par un encadré sur le mot et la notion de *bug*. La troisième section, « pour les experts », entre plus avant dans l'analyse des ordinateurs et des réseaux. Une première sous-section, « L'anatomie de l'ordinateur » (pp. 91-97), porte d'abord sur le matériel (le *hardware*) et s'arrête successivement sur la notion de *microprocesseur*, « véritable cerveau de l'ordinateur » et « grand horloger » ; de *mémoire* et ses différentes sortes : mémoire « morte » ou ROM, mémoire « vive » ou RAM (*Random Acces Memory*), « brouillon » de l'ordinateur, à quoi s'ajoute le *disque dur*, mémoire de type RAM où « viendrons s'enregistrer les différents logiciels installés sur l'ordinateur, ainsi que les documents créés » ; de *carte mère*, « carte électronique sur laquelle sont fixées les principaux composants de l'ordinateur (microprocesseur, mémoire, etc.) et qui permet la connexion entre eux »³⁵.

L'auteur passe alors à la description des logiciels – du *software*. Il y distingue d'une part le *système d'exploitation*, lequel « joue le rôle d'interface entre l'utilisateur et la machine », en signalant que « les systèmes d'exploitation les plus utilisés sont Windows, format propriétaire de l'entreprise Microsoft, et Unix, format libre et ouvert », et d'autre part les *logiciels applicatifs*, « programmes réellement utiles pour l'utilisateur », qui existent par centaines, « tous spécialisés dans des fonctions déterminées, telles que la conception d'image ou la gestion du stock d'une entreprise » (p. 93).

Notons ici que, parler comme le fait l'auteur cité d'un « format propriétaire de Microsoft », c'est commettre un anglicisme (en prétendant « traduire » l'anglais proprietary par le français « propriétaire ») et, au mieux, user d'un néologisme. Un « format » n'est évidemment propriétaire de rien du tout ; il peut en revanche être la propriété de... La définition rapide (quick definition) de l'adjectif proprietary proposée par le OneLook Dictionary est par exemple celle-ci : protected by trademark or patent or copyright; made or produced or distributed by one having exclusive rights (Example: “ ‘Tylenol’ is a proprietary drug of which ‘acetaminophen’ is the generic form”). Le Journal officiel du 22 septembre 2000 propose de traduire “proprietary system” par « système exclusif » et déconseille explicitement de parler de « système propriétaire ».

Protocoles

Pour l'« expert » toujours, l'auteur consacre encore trois sous-sections respectivement à la *taille des informations*, à la *compression*, enfin aux *protocoles*.

Le Free On-line Dictionary Of Computing (Howe, 1993-2008) propose la définition suivante de protocol (mot traduit en français par « protocole ») : “A set of formal rules describing how to transmit data, especially across a network. Low level protocols define the electrical and physical standards to be observed, bit- and byte-ordering and the transmission and error detection and correction of the bit stream. High level protocols deal with the data formatting, including the syntax of messages, the terminal to computer dialogue, character sets, sequencing of messages etc.” Le Journal Officiel du 22 septembre 2000 définit un protocole (au sens précédent) comme l'« ensemble des conventions nécessaires pour faire coopérer des entités généralement distantes, en particulier pour établir et entretenir des échanges d'informations entre ces entités », en précisant que ces entités « peuvent être des éléments réels ou virtuels, matériels ou

logiciels, d'un réseau de télécommunication ou d'un ensemble de traitement de l'information ».

La première des trois sous-sections évoque *bits* et *octets* (en anglais *bytes*). Le lecteur y apprendra notamment, s'il ne le savait déjà, qu'un fichier de texte « simple » représente « quelques dizaines ou centaines de kilooctets » alors qu'une chanson de trois minutes codée en MP3 représente environ deux mégaoctets, en sorte qu'une mémoire de 20 gigaoctets « peut donc contenir 20 milliards de lettres d'un octet, ou bien 10 000 morceaux de musique de 3 minutes environ ».

En fait, un kilooctet ne vaut pas 1000 octets mais $2^{10} = 1024$ octets ; de même un mégaoctet vaut 1024 kilooctets, soit 2^{10} octets ou 1 048 576 octets ; un gigaoctet vaut de même 2^{30} octets, soit $1\,073\,741\,824 = 1,073741824 \times 10^9$ octets. Ainsi 10 Go valent, non pas 10^{10} octets, soit dix milliards d'octets, mais 10 737 418 240 octets, soit 737 418 240 octets de plus ou $\frac{737\,418\,240}{2^{20}} \text{ Mo} \approx 703 \text{ Mo}$. Plus de 700 « vrais » mégaoctets, cela n'est pas tout à fait négligeable. Pour éviter ce genre d'ambiguïtés, la Commission Électrotechnique Internationale (International Electrotechnical Commission, IEC) a introduit les préfixes kibi (pour “kilo binary”), mebi, gibi, etc. Un kilooctet serait alors 10^3 octets, tandis qu'un « kibioctet » (en anglais “kibibyte”) désignerait 2^{10} octets, etc. (voir http://www.iec.ch/zone/si/si_bytes.htm). Cette nomenclature semble encore peu usitée dans l'informatique grand public.

S'agissant de la notion de compression, le lecteur est introduit à la distinction entre les méthodes de compression « sans perte » (comme il en va avec la compression ZIP), et les méthodes de compression « avec perte », qui ne permettent pas de retrouver intégralement l'objet compressé, et dont les plus connus sont désignés par les sigles JPEG (*Joint Photographic Expert Group*) pour la compression des images numériques, et MPEG (*Motion Picture Expert Group*) pour la compression de documents audio et vidéo. À cet égard, l'auteur en dit un peu plus long à propos de la compression MPEG, dans ses versions successives. La version MPEG1, note-t-il, ne s'utilise plus guère que pour la compression audio ; la version la plus utilisée aujourd'hui est la version MPEG-2 qui permet une « transmission audiovisuelle de qualité standard ». À cette occasion, il précise encore ceci.

Le fameux format MP3 ne signifie pas MPEG3 mais MPEG1/Audio Layer 3 et correspond à la spécification sonore du standard MPEG1. Ce format populaire de compression audio permet une compression d'environ 1 pour 12. Un fichier audio occupe ainsi 12 fois moins d'espace une fois transformé en format MP3. (p. 96)

Selon l'auteur toujours, la version MPEG-3 n'est pas utilisée, tandis que MPEG-4 est « une version en plein développement notamment pour la gestion de l'interactivité dans l'image ». Il observe encore que la compression est « couramment utilisée sans que l'on s'en rende compte », par exemple lors d'une conversation téléphonique (fixe ou mobile) ou à la télévision.

Le développement sur les protocoles termine le chapitre 5 et revient sur la question des réseaux. Les protocoles, précise l'auteur, constituent « le langage informatique standardisé utilisé par les ordinateurs pour communiquer les uns avec les autres et échanger des informations » (p. 96). Il compare alors l'adoption d'un protocole commun à ce qui se passe « lors d'une réunion d'hommes d'affaires de différentes nationalités, lorsqu'on décide d'utiliser une langue, l'anglais par exemple, afin que tout le monde se comprenne ». Le chapitre s'achève sur la mention du protocole Internet, « *Internet Protocol*, d'où vient le sigle IP ».

Difficulté de l'élémentation : l'exemple du BIOS

L'une des difficultés constantes de « l'échantillonnage épistémologique » sur lequel peut s'appuyer l'élémentation du domaine praxéologique de la RII est la tentation de s'arrêter aux découpages de connaissances existants, qui procèdent généralement davantage d'une « *élémentarisation* » à mobile socio-cognitif qu'à une véritable élémentation, fonctionnelle du point de vue des besoins épistémologiques éprouvés. La règle qui s'en déduit oblige d'abord à refuser par méthode le monde des organisations praxéologiques *tel qu'il est*, pour observer les mises en cause dont il peut être l'objet, de tel ou tel point de vue institutionnel ou personnel, quant à ses manques éventuels, ses excès possibles, ses silences inattendus, ses insistances immotivées³⁶. Cette même règle oblige ensuite, non à vouloir maîtriser la totalité des connaissances dans une pluralité de domaines (histoire, économie, informatique..), nous l'avons souligné plus haut, mais à avoir *le souci de la totalité* de ces domaines, en considérant que nul découpage préalable ne peut nous épargner l'obligation épistémologique d'explorer indéfiniment, par delà toute frontière existante, les domaines de connaissance en évolution permanente sur lesquels peut porter le projet jamais achevé d'élémentation.

De cela nous donnerons un exemple rapide, dépourvu de tout caractère crucial, en revenant à l'ouvrage *Informatique et internet au collège et au lycée* déjà mentionné (Carrier, 2007). Après une page d'introduction suivie de trois pages d'un historique intitulé « Le développement de l'informatique », ce manuel comporte une unité d'un peu plus de deux pages sur « Le poste informatique », où est présentée la composante matérielle (écran ou moniteur, unité centrale, clavier, imprimante...) puis – sommairement – la composante logicielle (système d'exploitation, applications) d'un tel poste et où sont abordées l'installation et la mise en marche de l'ordinateur (et son arrêt), ce qui conduit l'auteur à donner à voir et à présenter l'« arrière » de l'unité centrale (bouton marche-arrêt, prise de raccordement au secteur, sortie vidéo vers l'écran, à quoi s'ajoutent les différents « ports », USB, Ethernet, etc.). Le chapitre suivant occupe moins de quatre pages (pp. 13-16) et a pour titre « L'unité centrale » ; il s'ouvre par un « écorché » qui laisse voir les principaux organes que cache le boîtier d'une unité centrale du type « tour » : alimentation, carte mère portant le microprocesseur, carte graphique, disque dur, graveur de CD, lecteur de DVD ³⁷, etc. Les divisions et sous-divisions du chapitre sont les suivants.

Le boîtier

L'alimentation

La carte mère

Le support du microprocesseur (*socket*)

Le BIOS

Les connecteurs de barrettes mémoire vive (RAM)

L'horloge système

Les connecteurs d'extension

Les connecteurs d'entrée/sortie

Le disque dur

Le lecteur de disquette

Le lecteur/graveur de CD (cédérom et CD audio)

Le lecteur/graveur de DVD

L'ouvrage présente un choix de connaissances relatives à la phénoménologie de l'ordinateur tel que le vit l'utilisateur « ordinaire ». Bien que l'approfondissement nouménologique semble chaque fois avoir été calculé au plus juste, les références à des connaissances « adventices », quoique discrètes, sont bien réelles.

Parmi les connaissances qui relèvent de la physique, on peut ranger le fait que l'alimentation contenue dans l'unité centrale a pour fonction de transformer le courant « fourni par une prise » (donc alternatif) en un courant « aussi continu que possible » – ce qui peut éclairer l'utilisateur sur certaines anomalies de fonctionnement de son ordinateur ! Semblablement, à propos du microprocesseur, le lecteur se voit indiquer que « plus sa fréquence de fonctionnement est grande (600, 1200, 2000 MHz...), plus l'ordinateur est rapide ». Ainsi passe-t-on de la physique de l'électricité aux technologies numériques proprement dites. Il en va de même dans le développement suivant, consacré aux barrettes de mémoire vive – qui pourra s'articuler, lui aussi, avec l'expérience vécue du pur utilisateur (p. 14).

Les connecteurs des barrettes de mémoire vive (*memory slot*, en anglais) sont au nombre de 2, 3 ou 4 sur les cartes mères communes.

La mémoire vive est l'espace où l'ordinateur stocke temporairement les données et les programmes qu'il utilise. Les applications actuelles nécessitent de plus en plus de mémoire vive. Si votre ordinateur « peine », il peut être utile de rajouter de la RAM. L'opération est assez simple.

L'exercice auquel se livre l'auteur est difficile. Son « découpage » de l'unité centrale est fait en termes de dispositifs matériels, qu'il s'agisse de la carte mère, de l'alimentation, du microprocesseur, voire du logement (*socket*) du microprocesseur, sans parler de l'horloge système, du disque dur et autres lecteurs et graveurs.

Dans ce tableau quasi physique, pourtant, l'auteur a inséré une réalité qui manifestement n'est pas de cet ordre : le « BIOS », dont il écrit : « Il réside en mémoire non volatile. Lors de la mise en route, le *BIOS* vérifie les différents éléments (mémoire vive, mémoire cache, unités de stockage...) puis il lance le système d'exploitation. » La formulation adoptée ne fait pas apparaître la nature – logicielle – du BIOS ; celui-ci réside « en mémoire non volatile », nous dit-on, mais sa localisation sur la carte mère n'est pas autrement précisée. Avant d'aller plus loin sur ce point, donnons ici une définition du BIOS telle qu'on la rencontre dans l'*Encarta World English Dictionary*³⁸.

control program for basic computer operation: a small unerasable computer program that contains the instructions needed to begin operation and controls the data flow between the operating system and application programs and the hardware devices.

Full form basic input-output system

Le BIOS apparaît ainsi comme le « chaînon manquant » d'une description simplifiée qui réduit le *software* au système d'exploitation (lequel réside en principe sur un disque, par exemple le disque dur) et aux multiples logiciels applicatifs.

Pour saisir ce phénomène, revenons un instant à l'ouvrage de Moira Stephen, *Basic computer skills*, pour rechercher dans son chapitre 1, *Information and communications technology* (dont on a reproduit ci-après le sommaire), le traitement réservé à ce chaînon manquant.

- 1.1 A brief history lesson
- 1.2 Computers in everyday life
- 1.3 Computer talk
- 1.4 Hardware
- 1.5 Software
- 1.6 PC performance
- 1.7 Information networks
- 1.8 Looking after yourself
- 1.9 Hardware and software security
- 1.10 Legal issues
- 1.11 Other terms you will meet
- 1.12 A changing world

On peut supposer que ce que nous recherchons se trouve quelque part entre les sections 1.4 et 1.5, *Hardware* et *Software*. Nous reproduisons ci-après le passage pertinent, situé en fait au début de la section 1.5 (Stephen, 2003, p. 15).

Software refers to the programs that make your computer work. Software includes the operating system (essential to get your computer up and running) and the applications, e.g. software that lets you carry out specific tasks (word processing, spreadsheets, e-mail, databases, etc.).

Operating system

The OS is essential to the efficient running of the PC. It controls which operations are carried out and in what order they are done. It ensures that when you press a key, the instruction is translated into something the computer can work with.

When a computer is switched on it is said to *boot-up*. During the boot-up process it carries out a Power On Self Test (POST) to check that the hardware components and the CPU and memory are all present and functioning correctly. The next thing that the computer does when it boots-up is to locate

and load the OS (or part of the OS). The OS is usually stored on disk, e.g in your Windows folder.
The OS is loaded into RAM at this stage.

On voit que ce qui est appelé ici *software* est présenté comme *incluant* le système d'exploitation et les logiciels applicatifs, sans formellement s'y réduire. C'est alors dans le lancement du système d'exploitation que – avec d'autres mots – l'existence du BIOS va être évoquée, à travers l'idée de *boot* : l'ordinateur, écrit l'auteure, “is said to boot-up”. La description proposée montre notamment que ce logiciel résident qu'est le BIOS recherche le système d'exploitation et le charge en mémoire vive. Le BIOS serait ainsi, pour parler comme Aristote, le premier moteur – un premier moteur dont l'utilisateur aura sans doute observé la manifestation graphique à l'écran lors du démarrage de leur ordinateur.

Dans chacun des deux exposés examinés – celui d'Alain Carrier, celui de Moira Stephen –, la présentation est parcellaire, comme le fait comprendre la description suivante, plus complète, empruntée au site *Comment ça marche ?* ³⁹

Différentes mémoires de type *ROM* contiennent des données indispensables au démarrage, c'est-à-dire :

- Le **BIOS** est un programme permettant de piloter les interfaces d'entrée-sortie principales du système, d'où le nom de *BIOS ROM* donné parfois à la puce de mémoire morte de la carte-mère qui l'héberge.
- Le **chargeur d'amorce** : un programme permettant de charger le système d'exploitation en mémoire (vive) et de le lancer. Celui-ci cherche généralement le système d'exploitation sur le lecteur de disquette, puis sur le disque dur, ce qui permet de pouvoir lancer le système d'exploitation à partir d'une disquette système en cas de dysfonctionnement du système installé sur le disque dur.
- Le **Setup CMOS**, c'est l'écran disponible à l'allumage de l'ordinateur permettant de modifier les paramètres du système (souvent appelé *BIOS* à tort...).
- Le **Power-On Self Test (POST)**, programme exécuté automatiquement à l'amorçage du système permettant de faire un test du système (c'est pour cela par exemple que vous voyez le système “compter” la RAM au démarrage).

Étant donné que les ROM sont beaucoup plus lentes que les mémoires de types RAM (une ROM a un temps d'accès de l'ordre de 150 ns tandis qu'une mémoire de type SDRAM a un temps d'accès d'environ 10 ns), les instructions contenues dans la ROM sont parfois copiées en RAM au démarrage, on parle alors de *shadowing* (en français cela pourrait se traduire par *ombrage*, mais on parle généralement de *mémoire fantôme*).

Cette description peut être complétée encore par la suivante, empruntée, elle, au site *Vulgarisation scientifique*⁴⁰.

Le BIOS (Basic Input Output System) est un petit programme situé dans plusieurs types de mémoires différentes : une partie dans une mémoire ROM (Read Only Memory), cette partie est non modifiable (il s'agit du boot block). La deuxième partie du BIOS se situe dans une mémoire dont le contenu est modifiable (l'EEPROM). C'est cette partie que l'on modifie lorsqu'on parle du terme "flashage". La troisième partie du BIOS se situe dans la mémoire CMOS, cette mémoire contient tous les paramètres du BIOS. Effacer cette mémoire est sans danger (lorsque l'on enlève la pile de la carte mère, on efface la mémoire CMOS), mais tous les paramètres du BIOS devront être reconfigurés !

Pour communiquer avec les périphériques matériels, un système d'exploitation utilise les pilotes. Il charge les pilotes à chaque démarrage. Seulement comment charger un pilote de disque dur par exemple si celui-ci n'a pas été amorcé ?

C'est en fait le rôle du BIOS. Le BIOS va charger tous les périphériques de base et effectuer un test du système. Cette phase de démarrage est appelée POST (Power On Self Test)

Voici ce que le POST fait au démarrage :

- Il effectue un test du processeur en premier, puis vérifie le BIOS.
- Il va chercher à récupérer les paramètres du BIOS et va donc vérifier le contenu de la mémoire CMOS.
- Il initialise l'horloge interne et le contrôleur DMA.
- Il contrôle le bon fonctionnement des mémoires (vive et cache).
- Il vérifie les différents périphériques : carte graphique, disques durs, lecteurs de disquettes et CD-ROM. Si un problème survient, le BIOS donne la source du problème en fonction du nombre de bips qu'il émet.
- Il répartit les différentes IRQ et canaux DMA disponibles entre tous les périphériques.

Comme on le voit, l'univers d'objets mis en jeu, quoique fini, est ouvert : pour compléter les exposés précédents, il faudrait encore préciser, même de façon minimaliste, ce que sont les *CMOS*, *SDRAM*, *EEPROM*, *IRQ* et autres contrôleurs et canaux *DMA*.

On saisit peut-être mieux ici le problème du découpage des organisations de connaissances qu'une élémentation doit chercher à résoudre. Voici par exemple un texte, emprunté au site *Dico du Net*, qui se situe davantage du côté de la phénoménologie de l'utilisateur que les textes précédents, et qui consomment donc moins de connaissances

informatiques, même s'il engendre une identification coupable entre BIOS et écrans de réglage ⁴¹.

BIOS - Basic Input Output System

Définition

Ce logiciel donne les premières instructions au démarrage d'un ordinateur. Le BIOS est un ensemble de routines logicielles en mémoire morte (ROM et EEPROM).

Accessible en général au démarrage de l'ordinateur (souvent en appuyant sur la touche Suppr.), il comporte plusieurs écrans de réglages.

La plupart des utilisateurs ne font aucun réglage du BIOS ; il faut cependant s'assurer que le processeur fonctionne bien à sa vitesse de fonctionnement. Pour cela il faut régler le FSB et le coefficient multiplicateur sur les valeurs correspondantes au processeur.

Le BIOS permet également de nombreux réglages qui peut (*sic*) utiles tant que l'ordinateur fonctionne bien. Citons tout de même le réglage des timings de RAM qui peut faire gagner un peu de performances.

Ces écrans de réglage indiquent également en général la température du CPU et de la carte mère.

Que le problème soulevé – quelles connaissances informatiques « généralistes » pour une bonne efficacité et un confort réel en matière de RII ? – soit de résolution incertaine ne dispense pas de le poser et de chercher à l'affronter.

CHAPITRE 12

La fracture de l'anglais

Tout ce qui précède nous a montré à profusion la présence proche de la langue anglaise à propos de la plupart des questions étudiées jusqu'ici. Il en irait de même si, au lieu d'avoir pour langue de travail le français, notre étude se faisait en italien, en espagnol ou en néerlandais, par exemple, même si chacune de ces langues et chacune des communautés linguistiques qui les parlent peuvent avoir une « posture » différente vis-à-vis de l'anglais comme langue quasi native de « l'informatique ». La raison du rôle particulier de l'anglais en ce domaine se trouve évidemment dans la genèse historique des sciences et technologies de l'information et de la communication, qui ont trouvé aux États-Unis leur terre d'élection, en sorte que l'informatique « parle anglais » (ou plutôt américain), à la manière dont la médecine parlait autrefois latin (et, secondairement, grec).

Dans une discussion ⁴² à propos des résultats des 19^{es} Olympiades Internationales d'Informatique (en anglais *International Olympiad in Informatics*), dont les sujets étaient rédigés en anglais *uniquement*, un participant à un forum de discussion intervient en ces termes face à la mise en cause de cette disposition du concours : « Prétendre faire de l'informatique sans comprendre l'anglais... comment dire... ça frise le ridicule !!! (sans jamais en sortir). » La question du recours à l'anglais semble en fait être encore un point sensible en France (sinon dans le monde francophone en général). Nous avons noté déjà qu'il s'agit là d'une *frontière* du point de vue de la diffusion des organisations praxéologiques relatives à la RII (et plus largement au monde « Informatique et Internet »). Il paraît vraisemblable que la réticence à recourir à la langue anglaise soit corrélée à une connaissance *a priori* médiocre de cette langue et surtout se trouve associée à un *travail linguistique insuffisant*, parfois presque totalement absent.

Un faux problème ?

Dans l'abord de cette question, on notera en premier lieu un type de faits qui peut être de nature à accroître la réticence des lettrés face à ce qui apparaît pourtant davantage comme une

nécessité *fonctionnelle* que comme une imposition « idéologique ». Un petit nombre de mots ou d'expressions sont employés dans « l'anglais de l'informatique » alors même qu'ils proviennent, parfois obscurément, du français. L'un d'eux vient précisément d'être rencontré : c'est le mot même d'informatique, *informatics*. L'article « Informatique » de l'encyclopédie Wikipédia, déjà cité, en présente l'histoire ainsi.

Le terme *informatique* est un mot-valise créé en mars 1962 par Philippe Dreyfus, ancien directeur du Centre National de Calcul Électronique de Bull dans les années 1950, qui, en 1962, a utilisé pour la première fois ce terme dans la désignation de son entreprise « Société d'Informatique Appliquée » (SIA), à partir des mots « *information* » et « *automatique* ».

En France, l'usage officiel du mot a été consacré par Charles de Gaulle qui, en Conseil des ministres, a tranché entre « informatique » et « ordinateur », et le mot fut choisi par l'Académie française en 1967 pour désigner cette nouvelle discipline. En juillet 1968, le ministre fédéral de la Recherche scientifique d'Allemagne, Gerhard Stoltenberg, prononça le mot *Informatik* lors d'un discours officiel au sujet de la nécessité d'enseigner cette nouvelle discipline dans les universités de son pays, et c'est ce mot qui servit aussitôt à nommer certains cours dans les universités allemandes. Le mot *informatica* fit alors son apparition en Italie et en Espagne, de même qu'*informatics* au Royaume-Uni.

Pendant le même mois de mars 1962 Walter F. Bauer inaugura la société américaine Informatics Inc. qui, elle, déposa son nom et poursuivit toutes les universités qui utilisèrent ce nom pour décrire la nouvelle discipline, les forçant à se rabattre sur computer science, bien que les diplômés qu'elles formaient étaient pour la plupart des praticiens de l'informatique plutôt que des scientifiques au sens propre. L'Association for Computing Machinery, la plus grande association d'informaticiens au monde, approcha même Informatics Inc. afin de pouvoir utiliser le mot informatics pour remplacer l'expression *computer machinery*, mais l'entreprise déclina l'offre. La société Informatics Inc. cessa ses activités en 1985, achetée par Sterling Software.

Un autre cas moins connu se trouve dans le mot de *cache*, qu'utilise sans commentaire Alain Carrier en parlant – de façon désormais classique – de « mémoire cache » dans sa présentation du BIOS d'un ordinateur. L'histoire du mot semble un rien compliquée. L'article « Mémoire cache » de l'encyclopédie Wikipédia précise ceci : « *mémoire cache* est la traduction littérale de l'anglais *cache memory*, qui vient elle-même de “mémoire cachée”, principe inventé à Grenoble dans les années 1960. » Le mot *cache*, qui apparaît en anglais à la fin du XVIII^e siècle, y désigne, pour suivre ici le *Merriam-Webster's Online Dictionary*, “a hiding place especially for concealing and preserving provisions or implements, a secure place of storage”,

soit, en français, *une cache*. La chose ne doit guère surprendre puisque, à suivre maintenant le *Online Etymology Dictionary* (Harper, 2001), le mot de *cache*, qui serait apparu en anglais en 1787, est un emprunt “from Fr. Canadian trappers’ slang, ‘hiding place for stores’ (c. 1669), from Fr. *cache* ‘conceal’...” De tels échanges en tous sens ne sont pourtant pas la règle commune.

Quand la langue s’opacifie...

Souvent, peut-on penser, l’absence de familiarité sémantique – et pragmatique – avec la langue anglaise conduit à l’*opacification*, pour le locuteur français, des mots anglais utilisés dans le jargon informatique, phénomène que renforce au reste fréquemment, en le masquant, une « traduction » qui, sans être forcément une trahison réelle, filtre le sens de façon réductrice.

Nous avons vu, dans un chapitre précédent, l’exemple du mot *browser*, dont le sens ordinaire en anglais est masqué autant par l’officiel « logiciel de navigation » que par l’approximatif « butineur ». Le phénomène d’opacification est alors corrélatif d’un phénomène d’« hypercontextualisation », dont nous avons vu aussi un minuscule exemple, celui de la mention, dans l’ouvrage *Informatique et internet au collège et au lycée*, du mot anglais *socket* à propos du « support du microprocesseur » sur la carte mère. Cette occurrence *semble* dire que ce serait là le nom *spécifique* donné en anglais au logement du microprocesseur sur la carte mère. Or, *socket* désigne en anglais n’importe quel “receptacle where something is inserted”. Le dictionnaire étymologique déjà mentionné indique que ce mot, qui vient à l’origine de l’ancien français « soc » (de charrue), se met à signifier au XIII^e siècle « pointe d’épée » avant de désigner n’importe quelle “hollow part or piece for receiving and holding something”, acception attestée pour la première fois en 1448 – le sens de « prise électrique » n’apparaissant qu’en 1885. Le passage du *socket* de la langue anglaise au *socket* du praticien de l’ordinateur *de langue française* s’accompagne ainsi d’un appauvrissement important de la sémantique culturelle.

Ce phénomène est général. On a vu plus haut l’emploi du verbe anglais *to boot up*, qui donne, en adaptation française informatique, « booter » et « rebooter ». Là encore, ces verbes n’ont plus, si l’on peut dire, de véritable sémantique (pour le non-anglophone) : ils sont une simple *étiquette* pour désigner une opération qu’ils indexent à la manière dont, dans les cultures « modernes », le prénom d’une personne la désigne *sans rien signifier*. La situation est toute différente avec l’anglais « informatique », où le verbe *to boot* a une histoire qui, pour

n'être pas des plus anciennes, n'en est pas moins fort riche. Voici d'abord l'indication que fournit à son propos le *Online Etymology Dictionary* (Harper, 2001).

boot

“start up a computer,” 1975, from *bootstrap* (n.), 1953, “fixed sequence of instructions to load the operating system of a computer,” on notion of the first-loaded program pulling up itself by the bootstraps.

L'idée clé est ici celle de *bootstrap*, geste « inaugural » par lequel une personne ou, comme ici, un « système », devient en quelque sorte sa propre cause. Écoutons à ce sujet *The American Heritage Dictionary of the English Language* dans sa 4^e édition (Picket et al, 2000).

boot•strap

n.

1. A loop of leather, cloth, or synthetic material that is sewn at the side or the top rear of a boot to help in pulling the boot on.
2. An instance of starting of a computer; a boot.

tr.v. bootstrapped, bootstrapping, bootstraps

1. To promote and develop by use of one's own initiative and work without reliance on outside help: “We've bootstrapped our way back with aggressive tourism and recruiting high tech industries” (John Corrigan).
2. Computer Science To boot (a computer).

adj.

1. Undertaken or accomplished with minimal outside help.
2. Being or relating to a process that is self-initiating or self-sustaining.

Dans ses “Quick Definitions”, le *OneLook Dictionary* propose plus sobrement pour le verbe *to bootstrap* : “help oneself, often through improvised means.” Voici maintenant « toute l'histoire », telle du moins que la raconte l'article “Bootstrapping” de l'encyclopédie Wikipedia.

Bootstrapping (computing)

In computing, **bootstrapping** refers to a process where a simple system activates another more complicated system that serves the same purpose. It is a solution to the Chicken-and-egg problem of starting a certain system without the system already functioning. The term is most often applied to

the process of starting up a computer, in which a mechanism is needed to execute the software program that is responsible for executing software programs (the operating system).

The term “bootstrapping” alludes to a German legend about Baron Münchhausen, who claimed to have been able to lift himself out of a swamp by pulling himself up by his own hair. In later versions of the legend, he used his own bootstraps to pull himself out of the sea which gave rise to the term bootstrapping. The term is believed to have entered computer jargon during the early 1950s by way of Heinlein’s short story *By His Bootstraps* first published in 1941.

L’idée de *bootstrapping* est présente aujourd’hui en biologie (*bootstrap process*), en matière de création d’entreprises (*bootstrap funding*), en électronique, en linguistique (où l’on parle de *syntactic bootstrapping* et de *semantic bootstrapping*), en physique (*bootstrap models*), en statistique, etc. L’idée est ainsi quasi ubiquitaire. Voici alors comment la *Commission générale de terminologie et de néologie* propose de rendre en français les verbes anglais *to boot* et *to reboot* ⁴³

Journal officiel du 10/10/1998

amorcer, v.

Domaine : Informatique

Définition : Mettre en marche un ordinateur en provoquant l’exécution de l’amorce.

Voir aussi : réamorcer, amorce

Équivalent étranger : boot, v. (en)

Journal officiel du 10/10/1998

réamorcer, v.

Domaine : Informatique

Définition : Provoquer de nouveau l’exécution de l’amorce.

Voir aussi : amorce, amorcer

Équivalent étranger : reboot, v. (en)

Il est possible que, pour certains utilisateurs anglophones des verbes *to boot* et *to reboot*, l’idée de *bootstrapping* passe aujourd’hui inaperçue ; il est certain aussi que, en retenant l’idée d’amorce (on amorce un ordinateur et on le réamorce), là où le praticien parle souvent de *démarrer* ou de *redémarrer*, la Commission a préservé quelque chose de l’idée de premier moteur. En revanche, la subtile interrogation qui hante l’idée de “bootstrapping” – “*who loads the loader?*” ⁴⁴ – et la référence au baron de Münchhausen, qui vont de pair, ici, dans le monde anglophone, sont perdues dans la traduction.

Notons que l'évitement des verbes démarrer et redémarrer par la Commission est sans doute dû au fait que leur usage transitif – comme dans « Mais redémarre ton ordinateur, enfin ! » – reste proscrit dans la langue « correcte ». Jean-Paul Colin écrit à ce propos dans son Dictionnaire des difficultés du français (Colin, 1994) : « L'emploi transitif [de démarrer] est de plus en plus fréquent : démarrer une fabrication, etc. Mais la seule tournure correcte du transitif est démarrer un bateau, c'est-à-dire "en larguer les amarres". Dans d'autres situations, il faut dire faire démarrer : Ma voiture démarre mal, ou J'ai du mal à faire démarrer ma voiture. »

HTML : le préfixe hyper

On a dit plus haut que, « en 2000, un groupe de travail du W3C définit le langage XML (*eXtended Markup Language*) destiné à remplacer le langage HTML (*HyperText Markup Language*) ». Il semble bien que, pour nombre de francophones, l'expression *HyperText Markup Language* soit fortement opaque, alors que le sigle correspondant, *HTML*, leur est devenu familier. Que perd-on à cet état de fait ? Un *Hypertext Markup Language*, cela pourrait être un langage – un code – de *markup* d'un hypertexte. Arrêtons-nous d'abord sur la notion d'hypertexte, mot pour lequel le *Dictionnaire culturel en langue française* (Rey et al., 2005) offre la notice suivante.

HYPERTEXTE [ipɛrtɛkst] n. m. (1965, répandu v. 1988 ; de *hyper-* et *texte*)

Fonction qui, dans un document informatique, permet de définir des renvois directs entre un élément (texte, image...) et d'autres, situés ou non dans le même fichier. *L'hypertexte est la base du web.* – Adj. *Un document hypertexte. Lien* hypertexte* (hyperlien). *Surfer sur le web en cliquant sur des liens hypertextes.*

Dans le *Dictionnaire bilingue Internet et multimédia* (Benenson & Juanals, 2004), on trouve, à l'entrée *Hypertexte*, la définition suivante.

Un hypertexte est un regroupement de données numérisées stockées sur un support électronique et structurées de façon non linéaire en unités d'informations reliées par des liens. L'hypertexte est constitué d'une base de données, elle-même composée d'unités d'information (appelées nœuds) en interrelation. Au sein de cette base, un réseau sémantique est construit à l'aide des liens utilisés pour définir les relations (hiérarchiques, associatives, analogiques) existant entre les nœuds. Le logiciel

permet à l'utilisateur-lecteur de manipuler les données et de naviguer dans ce réseau de façon hypertextuelle, c'est-à-dire non linéaire.

Le même dictionnaire comporte également, parmi quelques autres, les entrées *HyperCard*, *hypercourse*, *hyperdocument*, *hyperlink*, *hypermedia*, *hyperobject*, *hypertrail*, *hyperword*. Le préfixe *hyper* est prolifique et mérite donc d'être interrogé pour l'emphase qu'il semble transporter avec lui. Qu'en est-il au juste ? Le *Dictionnaire bilingue* précise, dans l'article déjà cité, que, « en 1965, Theodor Holm Nelson inventa le terme hypertexte et le définit comme un outil d'accès à des documents inter-reliés ». La référence à « Ted » Nelson va de soi dans le monde de langue anglaise. C'est ainsi que, dans ses “Quick définitions”, le *OneLook Dictionary* propose cette définition illustrée du mot *hypertext*.

Machine-readable text that is not sequential but is organized so that related items of information are connected (Example: “Let me introduce the word hypertext to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper—Ted Nelson”)

Cette définition a le mérite de reprendre un passage décisif, celui où Nelson introduit, sans doute pour la première fois sous une forme imprimée, le vocable *hypertext*. En revanche, quoique se référant également à Nelson, l'entrée correspondante du *Free On-line Dictionary of Computing* (Howe, 1993-2008) se réfère à une situation qui n'existait pas en 1965.

hypertext

A term coined by Ted Nelson around 1965 for a collection of documents (or “nodes”) containing cross-references or “links” which, with the aid of an interactive browser program, allow the reader to move easily from one document to another.

The extension of hypertext to include other media – sound, graphics, and video – has been termed “hypermedia”, but is usually just called “hypertext”, especially since the advent of the World-Wide Web and HTML.

L'examen de l'article de Nelson de 1965 intitulé “A File Structure for the Complex, the Changing and the Indeterminate” (paru dans les *Proceedings of the ACM National Conference*) montre que l'idée d'hyperlien n'est pas centrale dans la notion nelsonienne d'hypertexte (Wardrip-Fruin, 2004). À côté du mot d'hypertexte, Nelson introduit le mot d'*hyperfilm*, en précisant : “The hyperfilm—a browsable or vari-sequenced movie—is only

one of the possible hypermedia that require our attention.” Hypertexte et hyperfilm sont donc des genres d’*hypermedia*. Dans un article ultérieur intitulé “No More Teachers’ Dirty Looks”, Nelson (1970) en dit plus long sur ce mot, et sur le préfixe *hyper*.

Hyper-media are branching or performing presentations which respond to user actions, systems of prearranged words and pictures (for example) which may be explored freely or queried in stylized ways. They will not be “programmed,” but rather *designed, written, drawn and edited*, by authors, artists, designers and editors... Like ordinary prose and pictures, they will be *media*; and because they are in some sense “multi-dimensional,” we may call them *hyper-media*, following mathematical use of the term “hyper-.”

Cette indication donnée en passant permet de mieux comprendre le sens originel du préfixe *hyper*. Celui-ci apparaît en mathématiques dès le XVIII^e siècle avec Leonhard Euler (1707-1783), qui parle d’une courbe « hypergéométrique », et trouve de multiples emplois tout au long du XIX^e siècle et au-delà. Ainsi parle-t-on (Miller, 2008) de *hyperdeterminant* (Cayley, 1845 : “The function *u* whose properties we proceed to investigate may be conveniently named a ‘Hyperdeterminant’”), de *hypersphere* (1858), de *hypercomplex numbers* (vers 1889), d’*hypercube* (en 1894, en français), de *hypergeometric distribution* (1936), etc. Ce à quoi se réfère Nelson dans son emploi de *hyper* est très vraisemblablement l’idée de généralisation *par passage à une dimension supérieure* : les nombres *hypercomplexes* sont des nombres complexes généralisés en cela qu’ils renvoient à des espaces de dimension supérieure à trois (qui est la dimension de notre espace « ordinaire »). Même si l’expression de nombres hypercomplexes a ensuite été peu ou prou délaissée⁴⁵, le préfixe *hyper* s’est définitivement installé dans le vocabulaire mathématique avec le mot d’*hyperplane* (en français, *hyperplan*), que connaissent tous les étudiants des filières scientifiques et qui désigne, comme le précise l’*Encarta World English Dictionary*, “a figure in hyperspace that is the three-dimensional equivalent of a plane in ordinary space”. On aura noté en passant l’emploi du mot *hyperspace*, qui – selon le même dictionnaire – désigne un espace “with more than three dimensions”. L’idée et le mot d’hyperplan se trouvent déjà dans ce passage d’un article de 1860 du mathématicien George Salmon (1819-1904) intitulé “On the relation which connects the mutual distances of five points in space” et paru dans le *Quarterly Journal of Pure and Applied Mathematics* (Miller, 2008).

It must have occurred to any person who has observed the analogy between the processes of the analytic geometry of three and of two dimensions, that as far as algebraic processes are concerned

there is nothing to limit the methods employed to the case of three dimensions. If, for example, there were any beings who could conceive space of four dimensions, we could with our present faculties write as good a book of analytic geometry for their use as if we could conceive space of four dimensions ourselves. We should take four hyper-planes at right angle to each other, and define the position of any point by four coordinates expressing its four distances from these; and so proceed in every manner exactly as we proceed in the analytic geometry of three dimensions. [...] Now the process followed in Art. (2) may be extended word for word to space of four dimensions and gives the following relation between the distances of five points on a hyper-sphere.

Ce que Salmon, au milieu du XIX^e siècle, envisage encore comme une vue de l'esprit deviendra quelques décennies plus tard une réalité banale de la culture scientifique, celle-là même sur laquelle, vraisemblablement, Nelson prend appui pour nommer (et, sans doute aussi, pour *penser*) la réalité nouvelle qu'il entrevoit : de même que l'équivalent d'un plan dans un espace de dimension supérieure ou égale à 4 est appelé un hyperplan, de même ce qui correspond à un *média* (tel un texte ou un film) dans un « espace » de dimension supérieure sera un *hypermédia* – hypertexte ou hyperfilm. Avec ce « plongement » dans un espace de dimension supérieure, le champ des possibles s'accroît : telle est l'idée de base, que nous n'analyserons pas plus avant ici.

L'idée d'espace de dimension supérieure à 3 n'était pas nouvelle. En 1754 déjà, dans l'article « Dimension » écrit pour l'Encyclopédie qu'il dirige avec Diderot, d'Alembert écrit : « Cette manière de considérer les quantités de plus de trois dimensions, est aussi exacte que l'autre ; car les lettres algébriques peuvent toujours être regardées comme représentant des nombres, rationnels ou non. J'ai dit plus haut qu'il n'était pas possible de concevoir plus de trois dimensions. Un homme d'esprit de ma connaissance croit qu'on pourrait cependant regarder la durée comme une quatrième dimension, et que le produit du temps par la solidité, serait en quelque manière un produit de quatre dimensions ; cette idée peut être contestée, mais elle a, ce me semble, quelque mérite, quand ce ne serait que celui de la nouveauté. » (Pour certains commentateurs de ce texte, l'homme d'esprit en question aurait été... d'Alembert lui-même.)

HTML : le markup

Il est probable que le préfixe *hyper* n'évoque guère, pour son utilisateur d'aujourd'hui, ce qu'il évoquait pour Nelson, et que nous avons tenté de restituer en peu de mots. Malgré cela,

l'utilisateur peut avoir le sentiment – trompeur – que le mot d'hypertexte lui est transparent. Mais, s'il est francophone, il butera vraisemblablement sur une opacité notoire de l'expression *Hypertext Markup Language* : celle qui affecte pour lui le mot de *markup*. Significativement, l'auteur de l'ouvrage paru dans la collection *Que sais-je ?* intitulé *HTML le langage du Web* écrit ceci (Augier, 2000, p. 6).

HTML est le sigle du nom anglais : *Hypertext Markup Language*. – Traduit mot à mot cela donnerait quelque chose comme Langage de Descripteurs Hypertexte ce qui reste assez obscur. Plus simplement HTML est un langage qui permet de décrire des documents hypertextes au moyen de marqueurs.

La même gêne apparaît à la page précédente, à propos du sigle SGML cette fois, que l'auteur présente ainsi : « Ces lettres sont l'acronyme en anglais de “*Standard Generalized Markup Language*” ce qui voudrait dire en français : Langage Standard Généralisé de [description par] Délimiteurs. » On voit que c'est sur le mot de *markup* que l'on bute : jamais traduit, il est rendu ici indirectement à travers la mention de « Descripteurs » et autres « Délimiteurs ». Vers où faut-il chercher ?

L'enquête est rapide. Dans ses “Quick Definitions”, le *OneLook Dictionary* propose ceci pour le mot *markup* (pour lequel le *Dictionnaire bilingue Internet et multimédia* n'a pas d'entrée) : “detailed stylistic instructions for typesetting something that is to be printed; manual markup is usually written on the copy (e.g. underlining words that are to be set in italics).” L'*Infoplease Dictionary* définit un *markup* comme “a detailed instruction, usually written on a manuscript to be typeset, concerning style of type, makeup of pages, and the like”, et l'*Encarta World English Dictionary* comme “the addition to a text that is to be printed of coding or instructions for layout and style”. Plus concis encore, le *Wiktionary* indique simplement que le mot désigne “the notation that is used to indicate how text should be displayed”. Quant à *The American Heritage Dictionary of the English Language*, il articule le passé avec le présent dans une même sous-entrée.

- 4. a. The collection of detailed stylistic instructions written on a manuscript that is to be typeset.
- b. *Computer Science* The collection of tags that describe the specifications of an electronic document, as for formatting.

Le terme de *markup* est donc d'abord un terme classique de l'édition de textes. L'encyclopédie Wikipedia (en anglais) contient un article intitulé “Markup Language”

commençant par ces lignes, qui marquent également la continuité entre deux mondes, l'un fort ancien, l'autre très récent.

A markup language is a set of annotations to text that describe how it is to be structured, laid out, or formatted. Markup languages have been in use for centuries, and in recent years have also been used in computer typesetting and word-processing systems.

A well-known example of a markup language in use today in computing is HyperText Markup Language (HTML), one of the protocols of the World Wide Web. HTML follows some of the markup conventions used in the publishing industry in the communication of printed work between authors, editors, and printers.

On trouvera ci-après le résultat du “marking-up” d’une épreuve d’imprimerie, tel d’abord que l’illustre le *Publication Manual of the American Psychological Association* (2001, p. 340), puis le *Lexique des règles typographiques de l’Imprimerie nationale* (2002, p. 59).

Figure 7.1. Marking proofs. (See section 7.06.)

Sample of a Proof Marked for Correction

Make all marks in colored pencil or ink. Mark all corrections on the proofs; never alter the manuscript when correcting proofs. Because the original proofs are used by the typesetter, mark neatly using conventional proofreader's marks (see Table 7.1). When you find an error, make two marks, one in the text in the exact place where the correction is to be made and one in the margin next to the line in which the error occurs. Together the marks show exactly what is to be done. In the margin, circle words or abbreviations that are instructions. Do not circle words that new copy, and do not circle symbols. For more than one correction in a single line, mark the corrections from left to right in the nearest margin, and separate them by a slanted line (/) for clarity. do not try to squeeze corrections between the printed lines. Include any special instructions or questions in accompanying letter; do not write them on the proofs.

Handwritten marks in the image include: (ital)/in, (lc)/them, (cap)/p, an/(c)/(ital), and various symbols like a circle with a dot, a circle with a cross, and a circle with a slash.

Vérifier l'état du caractère	d'olivier sur une <u>(pique)</u> en dessous figu-	<u>o</u>
Composer en petites capitales	raient la devise grecque <u>A</u> <u>ROI</u> <u>bon et</u>	<u>u</u> / <u>p. cap.</u>
Composer en gras	<u>VALEUREUX PAR LES ARMES</u> et la mention <u>Régis</u> <u>Typi.</u> Robert Estienne, successeur	<u>r</u> / <u>gras</u>
Composer en lettres supérieures	de Conrad Néobar comme imprimeur du <u>Roi</u> , avait incité François <u>I</u> à faire les	<u>er</u> <u>I</u>
Composer en bas de casse	frais d'une typographie grecque digne des <u>Œuvres</u> à éditer. Claude <u>G</u> <u>ARAMOND</u> ,	<u>œ</u> / <u>l.b.d.c.</u>
Supprimer et espacer	qui demeure <u>le</u> <u>premier</u> de nos « tailleurs	<u>st</u> /

Le *Online Dictionary for Library and Information Science* de Joan M. Reitz propose une entrée *mark up* (il s’agit du verbe) puis *markup* (le substantif) ; nous en reproduisons le texte ici (Reitz, 2007).

mark up

In publishing, to prepare copy for printing by specifying details of typeface, fonts, layout, etc., in a code comprehensible to the compositor or typesetter.

markup

In computing, the encoding of a textual document with information external to the document, for example, to indicate the structure of the document, the type or grammatical function of words and phrases within the document, or the way in which portions of the document are displayed on screen or page, usually accomplished by inserting tags and format codes in the text. To see the markup tags used in this hypertext dictionary, select “View – Source” or its equivalent in the toolbar of your Web browser. Also spelled mark-up. See also: markup language.

Suivons l’invitation du *Online Dictionary for Library and Information Science* et allons voir le code source correspondant à l’article *mark up* (ci-après). Après le titre – “mark up” –, on y lira par fragments la phrase reproduite plus haut, avec, notamment, sa ponctuation – “In publishing, to prepare copy for printing by specifying details of typeface, fonts, layout, etc., in a code comprehensible to the compositor or typesetter.”

```
<p>
<dt><a name="markup1"><b>mark up</b></a>
<dd>In <a href="odlis_p.cfm#publishing">publishing</a>, to prepare <a
href="odlis_c.cfm#copy">copy</a> for <a href="odlis_p.cfm#printing">printing</a> by specifying
details of <a href="odlis_t.cfm#typeface">typeface</a>, <a href="odlis_f.cfm#font">font</a>s, <a
href="odlis_l.cfm#layout">layout</a>, etc., in a code comprehensible to the <a
href="odlis_c.cfm#compositor">compositor</a> or <a href="odlis_t.cfm#typesetting">typesetter</a>.
```

Dans le cas du nom *markup* (ou *mark-up*) comme du verbe *to mark up*, un mot manque à la langue française. Les signes proposés par le *Lexique* de l’Imprimerie nationale sont présentés comme des signes de *correction* : c’est ainsi le *but* qui est désigné, non le *moyen*, le *marking-up*, qui peut servir de toutes autres fins – comme il en va dans le code source ci-dessus. Dans la seconde édition du *Merriam-Webster’s Concise Handbook for Writers* (Stevens et al., 1998), on peut lire, sous le titre “Marking Up the Manuscript”, ceci : “Copy editors communicate with typesetters by using a set of standard conventions, symbols, and codes. Most of these are described and illustrated in the following section.” C’est donc tout cela que le français peine à dire d’une façon univoque et efficace.

Comment traduire “A (Web) search” ?

Nous examinerons un autre cas, très sensible, d’un manque problématique dans la langue française – manque qui n’est pas forcément le lot des autres langues européennes. Nous avons mentionné à de nombreuses reprises l’expression *Web search*. En anglais, l’internaute “searches the Web” (“for something”) ; il se livre alors à “a search” (sur le Web). C’est là un usage tout à fait courant de *search* – le nom et le verbe – dans la langue anglaise, fait dont les “Quick Definitions” du *OneLook Dictionary* donnent une bonne illustration.

noun: an investigation seeking answers (Example: “A thorough search of the ledgers revealed nothing”)

noun: boarding and inspecting a ship on the high seas (Example: “Right of search”)

noun: the activity of looking thoroughly in order to find something or someone

noun: the examination of alternative hypotheses (Example: “His search for a move that would avoid checkmate was unsuccessful”)

noun: an operation that determines whether one or more of a set of items has a specified property

verb: search or seek

verb: try to locate or discover, or try to establish the existence of (Example: “The police are searching for clues”)

verb: subject to a search (Example: “The police searched the suspect”)

verb: inquire into

L’entrée consacrée au verbe *to search* par le *Harrap’s New Standard French and English Dictionary* (1980) restituée, au demeurant, cette diversité d’emploi du mot.

search². 1. *v.tr.* inspecter (un endroit) ; chercher dans (un endroit, une boîte, un livre) ; fouiller dans (un tiroir, une bibliothèque) ; fouiller (un suspect, les poches de qn) ; scruter, sonder (un visage) ; **we searched the whole town for him**, nous avons parcouru toute la ville à sa recherche ; *Cust* : **to s. a ship, s.o.’s suitcase**, visiter un navire, la valise de qn ; *Jur* : **to s. a house**, faire une perquisition, une visite domiciliaire ; perquisitionner dans une maison ; *Surg* : **to s. a wound**, sonder une plaie ; *Lit* : **to s. men’s hearts**, scruter, sonder, les cœurs ; **to s. one’s own heart**, rentrer en soi-même ; **to s. one’s memory**, scruter sa mémoire ; *F* : **s. me!** je n’ai pas la moindre idée ! 2. *v.i.* faire des recherches ; **to s. for new remedies**, chercher des remèdes nouveaux ; **to s. after truth**, rechercher

la vérité ; **to s. for s.o., sth.**, (re)chercher qn, qch. ; **to s. through the dictionary for a word**, scruter tout le dictionnaire pour trouver un mot.

Dire de quelqu'un qu'il "*searches the Web*" pourra se traduire en disant qu'il cherche sur (ou dans) le Web. On pourrait dire aussi, certes, qu'il *fouille* le Web, mais cette manière de dire est, semble-t-il, d'un usage peu répandu ⁴⁶. Le problème du manque de mots surgit lorsqu'on passe du verbe, *to search*, au nom correspondant ; voici l'entrée du *Harrap's New Standard* qui lui est consacrée.

search¹ [sə:tʃ], *s.* **1.** (a) recherche(s) *f* (pl) ; **to make a s.**, (i) faire des recherches ; (ii) (*in property transactions*) faire une enquête de commodo et incommodo ; **to make a s. for s.o.**, (re)chercher qn ; **the s. for, after, sth.**, la recherche de qch. ; **in (the course of) my s. I found ...**, au cours de mes recherches j'ai trouvé... ; **in s. of sth.**, à la recherche de qch. ; **to be in s. of sth.**, être en quête de qch. ; chercher qch. ; être à la recherche de qch. ; **s. party**, expédition *f* de secours ; (b) *Magn* : **s. coil**, bobine *f* d'essai, d'exploration ; *Cmptr* : **binary s., dichotomizing s.**, recherche dichotomique ; **s. cycle**, cycle *m* de recherche ; **s. key**, critère *m* de recherche. **2.** (a) *Cust* : visite *f* ; **right of s.**, droit *m* de visite ; (*at sea*) droit de recherche ; (b) *Jur* : perquisition *f* (à domicile) ; **s. warrant**, mandat *m* ordre *m* de perquisition ; (c) fouille *f* (dans un tiroir, etc.).

Le problème va surgir, on le devine, au moment de traduire en français "*a search on the Web*" (ou "*a Web search*"), si du moins on veut éviter toute ambiguïté entre "*Web search*" et « recherche scientifique », "*scientific research*". Pour cela, il faudrait pouvoir parler d'une **cherche* sur Internet – ce dont, semble-t-il, aucune autorité d'aucune sorte n'a à ce jour pris la responsabilité. En anglais, les substantifs *search* et *research* ont des significations proches mais cependant distinctes, comme le rappellent les "Quick Definitions" du *OneLook Dictionary* concernant *research*.

noun: systematic investigation to establish facts

noun: a search for knowledge (Example: "*Their pottery deserves more research than it has received*")

verb: inquire into

verb: attempt to find out in a systematically and scientific manner (Example: "*The student researched the history of that word*")

Au-delà des différences de sens (on aura noté qu'il y a *research* quand il y a "*search for knowledge*"), on voit aussi que, ici, entre français et anglais, les possibilités s'inversent : alors que, en français, on peut dire « une recherche » mais non « une cherche », en anglais on dit "*a search*" mais non "*a research*" – le français « une recherche » se traduisant en certains contextes au moins par "*a research work*"⁴⁷. Cela noté, d'autres langues ont d'autres solutions que la distinction entre *search* et *research*⁴⁸. En espagnol, ainsi, l'équivalent du verbe *to search* est *buscar* ; le substantif correspondant est *búsqueda*, mot bien distinct du terme employé pour désigner la recherche scientifique, *investigación* ; on parle donc de "*buscar en la Web*", de "*búsqueda en la Web*" sans aucune ambiguïté par rapport à la notion de recherche « scientifique ».

CHAPITRE 13

L'Internet

Venant après l'histoire, l'économie, la politique, la documentation, le droit, les mathématiques, la physique et les technologies numériques, l'informatique et l'anglais, cet objet de connaissance qu'est l'Internet nous amène – en passant par le World Wide Web, que nous aborderons ensuite – sur le seuil de ce qui est au cœur de notre recherche : la RII. Plus encore que dans les domaines praxéologiques précédemment parcourus, il est malaisé de faire plus qu'illustrer par quelques exemples – qu'on espère judicieusement choisis – la pertinence des connaissances en ce domaine pour la RII : rien *a priori* qui ne soit à écarter comme intrinsèquement inutile !

L'union du téléphone et de l'ordinateur : le cas de l'ADSL

Nous revenons ici à un ouvrage déjà sollicité, *Les dessous d'Internet* d'Antoine de Clerck (2007) : c'est au chapitre 6, intitulé « Internet, un bel avenir » qu'il faut aller voir. L'auteur, qui a traité des télécommunications au chapitre 4 et des ordinateurs au chapitre 5, propose à son lecteur une formule choc (p. 99) : « Internet, écrit-il, c'est la fusion du chapitre 4 et du chapitre 5, le mariage du téléphone et de l'ordinateur. » Il y avait le téléphone, et il y avait les ordinateurs. L'auteur explicite ainsi ce qui se serait alors passé (p. 100) : « Un beau jour on a eu l'idée de faire communiquer entre eux les ordinateurs en utilisant le téléphone, et l'Internet fut créé. Internet c'est l'interconnexion des ordinateurs du monde entier à travers les réseaux de télécommunication ». Il précise encore la chose ainsi.

En fait on utilise le réseau téléphonique pour pouvoir raccorder les ordinateurs ensemble. Le fil du téléphone devient donc le moyen de brancher son ordinateur sur le réseau de l'opérateur téléphonique, qui lui-même est connecté aux réseaux des autres opérateurs dans le monde. On a appelé Internet l'ensemble de ces réseaux interconnectés. Ainsi, en branchant son ordinateur sur le téléphone on peut accéder à Internet, en fait à tous les autres ordinateurs connectés à un réseau téléphonique dans le monde. (p. 101)

Pour accéder à Internet à partir de son ordinateur, il faut solliciter un fournisseur d'accès à Internet, « qui offre la possibilité de se raccorder à l'Internet, le plus souvent par un branchement sur une prise téléphonique ». À propos de la connexion ADSL, l'auteur précise qu'il s'agit d'une technologie permettant « un accès Internet à haut débit sur les lignes téléphoniques traditionnelles en parallèle à la téléphonie » ; et il donne les précisions suivantes.

Cette innovation permet de se connecter à Internet à haut débit en utilisant une prise téléphonique classique, c'est-à-dire que la vitesse de communication est accélérée et que la connexion à Internet se fait en parallèle de la ligne téléphonique : on peut alors utiliser simultanément le téléphone pour une communication et utiliser l'ordinateur pour se connecter à Internet. (p. 103).

La signification en anglais du sigle ADSL (*Asymmetric Digital Subscriber Line*) ainsi qu'une traduction en français de cette dénomination (« ligne d'abonné numérique asymétrique ») sont données par l'auteur sans pour autant que, par exemple, l'adjectif « asymétrique » se trouve expliqué. Sur ce point, on peut recourir au *Dictionnaire bilingue Internet et multimédia* qui, à l'entrée ADSL, indique qu'il s'agit d'une « liaison numérique à débit asymétrique », expression dont le résidu d'opacité est alors levé par cette formulation : « Technique de transmission numérique qui permet de transporter des données sur une paire symétrique de fils de cuivre du réseau téléphonique, le débit dans le sens descendant (du réseau vers l'abonné) étant très supérieur au débit ascendant. » Cette précision éclairante, mais qui pourrait être encore affinée, nous paraît significative du problème des opacités et des lacunes qui, presque toujours, « mitent » un exposé, ici relatif à l'Internet, alors même que cet exposé offre en d'autres points des lumières rarement rencontrées. Sur le caractère asymétrique, l'article "Asymmetric Digital Subscriber Line" de l'encyclopédie Wikipédia indique par exemple ce que voici.

L'ADSL est considéré par les instances de normalisation comme une technologie destinée essentiellement aux particuliers et aux petites et moyennes entreprises. Vis-à-vis des ressources disponibles sur Internet, cette catégorie de clients est en général davantage amenée à télécharger des informations (consultation d'un site Web, par exemple) qu'à envoyer des informations vers un site distant. Il a donc été décidé de favoriser le sens de la communication qui va du réseau vers l'abonné (*downstream* en anglais), au détriment du sens qui va de l'abonné vers le réseau (*upstream* en anglais). C'est pour cela que l'ADSL est qualifié d'asymétrique : le nombre de sous-porteuses

affectées au sens descendant est plus élevé que le nombre de sous-porteuses affectées au sens montant.

On aura noté que, aux verbes *download* et *upload* déjà mentionnés, répondent les dénominations symétriques *downstream* et *upstream* : à nouveau, l'anglais se montre tout à la fois concis et cohérent. Mais continuons ; le même article précise encore ceci.

Dans la pratique, quand l'ADSL est mis en œuvre sur une ligne téléphonique classique (analogique), les sous-porteuses 16 à 31 sont affectées au trafic émis de l'abonné vers le réseau, et la sous-porteuse numéro 16, qui sert de signal de référence, n'est pas modulée et ne transporte donc pas d'information. Les sous-porteuses 32 à 255, quant à elles, sont affectées au trafic reçu du réseau par l'abonné. Pour ce sens de communication, c'est la sous-porteuse d'indice 64 qui joue le rôle de signal pilote.

Que sont les « sous-porteuses » ? Voici un passage précédent du même document qui éclaire la chose.

L'ADSL fait appel à la notion de sous-porteuses : la bande de fréquences comprise entre 0 Hz et environ 1,1 MHz est divisée en 255 intervalles de 4,3125 kHz. À chaque frontière d'intervalle est associée une sous-porteuse, qui est un signal modulé. La n -ième sous-porteuse est donc matérialisée sous la forme d'un signal dont la fréquence de base vaut $[n \times 4,3125 \text{ kHz}]$. Un modem ADSL peut donc être considéré comme la mise en parallèle d'un grand nombre de modems analogiques, chacun transmettant sur une fréquence différente : 4,3125 kHz pour le premier, 8,625 kHz pour le second, 12,9375 kHz pour le troisième, et ainsi de suite.

La sous-porteuse d'indice 0 n'est pas utilisée, car elle correspond à un signal de fréquence nulle. Les sous-porteuses d'indice 1 à 255 sont théoriquement utilisables pour transmettre des données. Toutefois, les sous-porteuses d'indice 1 à 15 ne sont en général pas exploitées en raison de la présence possible de signaux téléphoniques dans une gamme de fréquences proche des fréquences utilisées par ces sous-porteuses. Dans la pratique, lorsque l'ADSL est mis en œuvre sur une ligne téléphonique classique (analogique), les sous-porteuses d'indice 16 à 255 sont donc disponibles pour la communication ADSL proprement dite.

Bien entendu, on peut encore réduire les boîtes noires contenues dans ce qui précède. Ainsi en va-t-il avec la notion de *modem* et la notion de *modulation* qui lui est (apparemment) associée.

Modem, modulation, etc.

Revenons à l'auteur de l'ouvrage *Les dessous de l'Internet*. À propos du modem, il écrit ceci.

Pour accéder à Internet, il faut équiper son ordinateur d'un modem, mot qui signifie modulateur-démodulateur, et qui désigne un circuit électronique faisant l'interface entre l'ordinateur et le réseau de télécommunication. Il conditionne les informations émises par l'ordinateur pour pouvoir les transmettre suivant le format adéquat sur le réseau. Il récupère aussi les données qui lui sont envoyées et les transforme en un format lisible par l'ordinateur. Certains ordinateurs sont équipés [d'un] modem intégré. (p. 109)

De manière vraisemblablement délibérée, l'auteur reste ici à une certaine distance du phénomène qu'il évoque. L'article « Modem » de l'encyclopédie Wikipédia en français précise, de son côté, que le modem est un appareil servant « à convertir les données numériques de l'ordinateur en données analogiques transmissibles par une ligne téléphonique classique et réciproquement ». L'article de même titre de la même encyclopédie, mais en anglais, précise les choses comme ceci.

Modem (from **mod**ulator-**dem**odulator) is a device that modulates an analog carrier signal to encode digital information, and also demodulates such a carrier signal to decode the transmitted information. The goal is to produce a signal that can be transmitted easily and decoded to reproduce the original digital data.

À l'instar de son homonyme en français, le même document ajoute cet exemple central.

The most familiar example is a voiceband modem that turns the digital 1s and 0s of a personal computer into sounds that can be transmitted over the telephone lines of Plain Old Telephone Systems (POTS), and once received on the other side, converts those 1s and 0s back into a form used by a USB, Ethernet, Serial, or Network connection.

Au-delà de ces précisions, il conviendrait de s'instruire sur la modulation (et la démodulation) d'un signal : nous n'irons pas plus loin ici sur ce sujet, qui se situe à l'arrière-plan du paysage praxéologique relatif à l'Internet.

*Rappelons que l'auteur des Dessous de l'Internet consacre deux chapitres de son ouvrage à la question des ondes (chapitre 3 : La magie des ondes) et des télécommunications (chapitre 4 : Les télécommunications, loin et près) ; il y explique le phénomène de modulation d'une onde, mais uniquement en référence à la transmission d'un son. Au plan des mystères de la terminologie, le profane peut s'interroger sur la raison du féminin dans « porteuses » et « sous-porteuses », là où l'anglais parle de « signal porteur » (carrier signal) ? On trouvera la réponse dans le passage suivant de l'article « Modulation du signal » de Wikipédia : « La **modulation** peut être définie comme le processus par lequel le signal est transformé de sa forme originale en une forme adaptée au canal de transmission, par exemple en faisant varier les paramètres d'amplitude et d'argument (phase/fréquence) d'une onde sinusoïdale appelée porteuse. »*

Questions de débit : les liaisons physiques

On se rapproche davantage de la phénoménologie vécue par l'utilisateur lorsque, comme l'auteur que nous suivons, on se soucie de la question du *débit* de la connexion utilisée, comme il en va dans le passage suivant.

Le débit est la quantité d'information que l'utilisateur peut transmettre ou recevoir par unité de temps. On parle de kilobits par seconde (kbps) ou de mégabits par seconde (Mbps). Le bit est l'information élémentaire utilisée en informatique et 1 octet (o) = 8 bits. Si un document texte représente quelques kilo-octets, une image haute définition peut, elle, « peser » un ou deux mégaoctets. Quant à un morceau de musique de 5 minutes codé en MP3, il représente facilement 3 ou 4 Mo. (p. 109)

L'auteur va sur ce point jusqu'à proposer à son lecteur des travaux pratiques utiles.

Le débit maximal que l'on peut obtenir sur une ligne téléphonique traditionnelle est de l'ordre de 50 kbps. Ce qui veut dire qu'il faudra quelques secondes pour transmettre un fichier texte, mais de longues minutes pour transmettre et recevoir des images ou sons :

- pour un fichier texte d'une page de 40 octets, soit 320 kbits, il faudra $320 \text{ kb} / 50 \text{ kbps} = 6,4$ secondes.
- pour un morceau de musique MP3 de 5 minutes de 3 Mo, soit 24 Mbits, il faudra $24000 \text{ kb} / 50 \text{ kbps} = 8$ minutes.

Les accès haut débit améliorent significativement le confort d'utilisation de l'Internet avec des débits de 500 kbps, 1 Mbps et plus avec le câble ou l'ADSL. Avec un débit de 500 kbps, cent fois plus rapide que l'accès par ligne téléphonique traditionnelle, on télécharge le même fichier texte en 6 centièmes de seconde, et le morceau de musique en 5 secondes. (p. 109)

Ces descriptions concernent ce que l'on nomme généralement la *couche physique*, c'est-à-dire l'ensemble des moyens permettant la transmission effective des signaux⁴⁹. L'auteur a mentionné l'usage d'une ligne téléphonique et, en passant, du câble ; à cela il ajoute un développement sur le *Wi-Fi* (*Wireless Fidelity*), « système qui permet de se connecter à Internet à haut débit et sans fil, lorsqu'on est à proximité d'une borne spéciale, dans un rayon de 100 mètres environ » (p. 112). Il précise encore : « On branche un petit modem sur son ordinateur qui entre en communication via une onde radio avec la borne, elle-même connectée à Internet sur un réseau filaire, par l'intermédiaire d'une prise téléphonique par exemple. » L'intérêt, ajoute-t-il, est de s'affranchir de toute connexion par fil dans les lieux les plus divers (hôtels, aéroports), et y compris... chez soi. Le problème encore non résolu, note-t-il cependant, c'est celui du *handover*, soit le fait que la connexion ne s'interrompe pas lorsque l'utilisateur passe de la zone d'influence d'une borne à celle d'une autre borne.

L'auteur, qui assume d'importantes responsabilités chez un opérateur de télécommunications, souligne que le traitement de la question du handover distingue nettement la technologie du Wi-Fi et celle de l'UMTS (Universal Mobile Telecommunications System), c'est-à-dire la téléphonie mobile de troisième génération. « Le Wi-Fi, note-t-il, vient du monde informatique qui n'avait pas intégré ce paramètre, tandis que l'UMTS vient du monde des télécoms mobiles dont la préoccupation première a toujours été de gérer la mobilité. Dans un cas c'est l'informatique, spécialiste de la transmission de données qui aborde la mobilité, dans l'autre ce sont les télécoms, spécialistes des communications mobiles qui abordent la transmission de données. » (p. 113)

Des connaissances à diffuser ?

Les développements évoqués dans ce qui précède, autour du Wi-Fi notamment, nous ramènent au cœur de notre problématique : celle des savoirs ou plus généralement des praxéologies utiles à la RII comme à d'autres usages de l'Internet. Jusqu'à un certain point, ici, phénoménologie et nouménologie se rejoignent.

Sur d'autres aspects, la question des liaisons physiques qui « font » l'Internet peut toucher l'utilisateur, sinon dans sa pratique immédiate, du moins dans son rapport à l'Internet comme fait de civilisation et fait de société. On en donnera un exemple emprunté à un ouvrage déjà mentionné (Huitema, 1996). L'expression d'*autoroute de l'information* (dans l'original américain, *Information superhighway*) eut naguère une grande fortune politico-médiatique. Aux États-Unis, notamment, le vice-président Al Gore s'en fit le champion⁵⁰. D'après l'article que l'encyclopédie Wikipedia consacre à cette expression, le numéro daté du 3 janvier 1983 de l'hebdomadaire *Newsweek* précisait ainsi que des “information superhighways being built of fiber-optic cable will link Boston, New York, Philadelphia, and Washington, D. C. in a 776-mile system on the East Coast”. De son côté, l'*Oxford English Dictionary* indique que l'expression en cause désigne “a route or network for the high-speed transfer of information; esp. (a) a proposed national fiber-optic network in the United States; (b) the Internet.” Que pouvait alors avoir d'obscurcissant l'image de « l'autoroute de l'information » ? Voici ce que répond Christian Huitema dans son livre, publié alors que prospérait cette métaphore autoroutière.

On utilise souvent, pour décrire l'interconnexion des réseaux formant l'Internet, une métaphore routière. Il y aurait de grandes « autoroutes de l'information », des bretelles d'accès, des départementales. Je trouve cette image autoroutière particulièrement détestable. Quand on parle de construire des autoroutes, on voit tout de suite des bataillons de bulldozers abattre les forêts et raser les collines, des norias de camions déverser graviers et ciment, six voies de goudron traverser le pays. Or il s'agit seulement de fibres optiques, d'un micron de diamètre. Un millième de millimètre de verre ne saurait avoir le même coût ni le même impact que cinquante mètres de béton. Mais cette référence aux autoroutes amène aussi, plus subtilement, une idée de monopole. Il paraîtrait absurde de tracer, entre Paris, Lyon et Marseille, trois autoroutes à péage concurrentes. Nos amis les écologistes ne manqueraient pas d'aller s'asseoir devant les pelleteuses pour protester. Si on parle trop vite d'autoroutes de l'information, on risque d'avaloir sans y réfléchir la même conclusion, qu'il s'agit d'un investissement gigantesque qu'on devrait obligatoirement confier à une administration monopolistique. (pp. 71-72)

La connaissance utile – et « manquante », à suivre l'auteur –, est simplement ce fait qu'une fibre optique *a un micron de diamètre* : à propos de l'Internet, nous retrouvons ici l'utilité d'une connaissance physique.

Mais poursuivons. Dans un petit livre destiné aux « curieux de 9 à 12 ans », intitulé *Internet quel drôle de réseau !*, une auteure spécialiste des réseaux imagine une jeune fille,

Lou, qui vient de recevoir un ordinateur ; son héroïne entreprend aussitôt de mettre cet ordinateur en fonctionnement, devant sa petite sœur Adèle et un copain de celle-ci, Paul, ce qui donne lieu à la scène suivante.

Lou – D'accord, mais d'abord, je dois connecter l'ordinateur à Internet.

Paul – Comment tu fais ?

Lou – Eh bien, tu vois cette petite boîte avec des lumières vertes qui clignotent ?

Paul acquiesce.

Lou – C'est un modem qui nous permet de nous connecter : d'un côté il doit être relié à l'ordinateur, de l'autre, il est relié au câble.

Paul – Le câble de la toile d'araignée dont tu viens de nous parler ?

Lou – Exactement. À Paris, et comme dans beaucoup de grandes villes, des câbles ont été installés dans les immeubles, il y a quelques années, pour que les gens puissent recevoir la télévision. Ces mêmes câbles permettent maintenant de se connecter à Internet.

Paul – La télévision et Internet passent par le même câble ?

Lou – Oui.

Paul – Mais ça ne se mélange pas ?

Lou – Non, la technologie utilisée permet d'avoir les deux types de transmission sur le même câble. Internet et la télévision ne rentrent pas en interférence, ils passent juste sur le même câble. (Virieux & Chebret, 2006, pp. 17-18)

À suivre les aventures de Paul, Adèle et Lou (auxquels va venir se joindre Anna, la mère de Lou et d'Adèle, en qui on reconnaît l'auteure elle-même), on ne peut qu'être frappé d'une double évidence : d'une part, ce que Lou – ou sa mère Anna – explique ici et là aux enfants devrait être une connaissance commune, partagée par tous ; d'autre part, ce large partage des connaissances n'existe guère aujourd'hui en la matière. Suivons donc le récit proposé.

Numéros IP

Ayant évoqué la connexion physique de l'ordinateur au câble qui donne accès à l'Internet, Lou mentionne la nécessité que l'ordinateur reçoive un numéro IP (*Internet Protocol*), avec ce commentaire.

C'est un peu comme un numéro de téléphone : pour que notre ordinateur communique avec les autres, il faut qu'il ait ce numéro, de la même manière que pour téléphoner, il faut que tu aies un numéro de téléphone unique au monde. (pp. 19-20)

Sa petite sœur lui demande si elle le connaît par cœur. Lou réplique que ça ne servirait à rien, « parce qu'il change chaque fois qu'on se connecte ». Un numéro unique mais qui change tout le temps, comment comprendre cela ? Lou répond dans les termes suivants.

Eh bien, le numéro IP que nous donne le fournisseur d'accès à Internet nous est réservé pendant toute la durée de notre connexion. Quand on éteint l'ordinateur ou qu'on se déconnecte du réseau, le fournisseur d'accès peut donner ce numéro à un autre ordinateur. (p. 22)

Le flot des questions enfantines ne tarit pas : pourquoi n'aurait-on pas, comme avec le téléphone, un numéro fixe ? Lou se lance dans une réponse où elle marque *in fine* la frontière en deçà de laquelle elle se tiendra – sa mère Anna étant présentée comme la seule capable de la franchir.

C'est juste pour faciliter la configuration des ordinateurs : il suffit de dire à l'ordinateur qu'il obtient son numéro automatiquement, et toute la configuration réseau se fait, elle aussi, automatiquement. Ainsi, si les paramètres de la configuration changent chez notre fournisseur d'accès à Internet, ils changeront aussi automatiquement sur notre ordinateur. Mais on peut aussi avoir un numéro fixe si on en fait la demande à son fournisseur d'accès. Vous pourrez demander plus de détails à maman quand elle rentrera de son travail. (p. 22)

Le détail de l'explication fait référence à une opération dont l'utilisateur serait libéré du fait d'avoir une adresse IP dynamique : le reparamétrage de la « configuration réseau » qui, dans certains cas, se fait dès lors automatiquement.

Il existe d'autres arguments en faveur d'un IP dynamique : là-dessus, on pourra consulter par exemple le document <http://www.commentcamarche.net/faq/sujet-1887-securite-ip-statique-fixe-ou-dynamique>. Chaque fournisseur d'accès reçoit un ensemble de numéros IP qu'il gère à sa convenance. L'ensemble de tous les numéros IP est géré au sommet par l'Internet Assigned Number Authority (IANA) sous l'autorité de l'ICANN (Internet Corporation for Assigned Names and Numbers). Pour cela, l'IANA attribue des blocs de numéros IP à des Regional Internet Registries (RIR), qui les

délèguent eux-mêmes à des LIR, des Local Internet Registries (voir par exemple <http://www.ripe.net/info/ncc/index.html>). Ceux-ci à leur tour attribuent des numéros IP aux niveaux inférieurs (backbones, etc.), où l'on trouve les FAI. Tel site éducatif américain (<http://www.cybertelecom.org/dns/ip.htm>) présente l'exemple que voici : "Joe gets his IP number for his computer from ACME ISP. ACME ISP gets its backbone connection and IP number block from BETA Backbone. Beta Backbone gets its number delegation from ARIN. ARIN gets its numbers from IANA." Un ISP, Internet Service Provider, est un FAI. Le sigle ARIN désigne celui des cinq RIR qui gère les numéros IP pour les États-Unis : l'American Registry of Internet Numbers. (Voir là-dessus Lubrano, 2003.)

Peut-être est-on là devant une frontière praxéologique indépassable par des enfants de 9 à 12 ans ; mais on peut se demander si la même situation se justifie aujourd'hui pour des adultes (ou des adolescents). En vérité, nous nous trouvons là devant un *écotone* praxéologique où se rencontrent et s'entremêlent une phénoménologie qui ne permet pas véritablement d'agir (que doit-on faire exactement pour connecter un ordinateur donné à l'Internet ?) et une nouménologie parcellaire, qui éclaire très inégalement l'au-delà de la phénoménologie. D'une façon générale, en écologie du vivant, un écotone, zone de transition entre deux écosystèmes, est souvent relativement pauvre en espèces ; cette « loi » s'applique, apparemment, en matière d'écologie praxéologique : pour des raisons évidentes, l'exposé dont Lou, Adèle et Paul sont les héros reste à la lisière et de la phénoménologie pratico-technique, et de la nouménologie technologico-théorique de l'Internet. On se trouve là devant deux frontières qu'il n'est guère possible de franchir l'une et l'autre d'un même pas : l'une, qui donne accès à l'écosystème nouménologique, suppose, peut-on dire, que l'on se dirige « vers l'amont » ; l'autre, qui donne accès à l'écosystème phénoménologique, suppose qu'on dirige ses pas « vers l'aval ». L'image topographique employée ici a un mérite en dépit de sa simplicité : parce qu'il est en général plus difficile de monter que de descendre, elle suggère que les parcours vers la nouménologie sont plus rares ou plus inaccessibles que ceux pointant vers la phénoménologie de l'Internet du point de vue de l'utilisateur ordinaire.

Réseaux et protocoles

Pour illustrer l'idée de « montée » nouménologique, nous nous arrêterons un moment sur un autre petit ouvrage signé de Françoise Virieux, paru en 2004 et intitulé *Comment ça marche Internet ?* Le titre même marque une distance par rapport au souci pratique immédiat :

l'utilisateur « pur » peut ne s'inquiéter que fort peu d'un tel sujet et ne voir que les gestes à accomplir afin de disposer d'un ordinateur connecté à l'Internet prêt à l'emploi. De quoi l'auteure nous parle-t-elle ? Son premier chapitre (Virieux, 2004, pp. 5-19) évoque d'abord très rapidement l'histoire de l'Internet (en mentionnant notamment l'Arpanet, le CSNnet, le Bitnet, le Cernet et l'Eunet). L'Internet, souligne-t-elle, est un « réseau de réseaux d'ordinateurs », ces réseaux étant regardés comme des « systèmes autonomes » (*autonomous systems*) organisés selon une structure hiérarchique. « Par exemple, écrit-elle, le réseau local d'une entreprise qui relie tous les ordinateurs de cette entreprise va pouvoir être relié au réseau de son fournisseur d'Internet, lequel est lui-même relié à des réseaux nationaux ou internationaux » (p. 6). L'auteure va même jusqu'à annoncer une évolution aujourd'hui saluée sous le nom de Web 3.0⁵¹.

La tendance actuelle est de connecter non plus des ordinateurs, mais des objets quelconques, pourvus qu'ils soient pourvus d'une carte électronique capable de communiquer avec Internet : des téléphones portables, des téléviseurs ou même des appareils ménagers qui peuvent ainsi être commandés à distance (p. 6).

La connexion entre objets du réseau permet d'échanger des informations de nature diverse (texte, son, image, etc.). Une telle communication, note l'auteure sans s'y attarder (elle y reviendra), suppose des règles qu'on appelle en l'espèce des *protocoles*. Internet, souligne-t-elle encore, est le lieu d'un paradoxe : il offre un mode de communication fiable utilisant des canaux qui ne le sont pas. Cette remarque introduit à un développement sur les deux modes possibles de fonctionnement d'un réseau d'ordinateurs : le *mode connecté* et le mode *non connecté*. Le premier, qui reprend le modèle de la communication téléphonique (la communication mobilise un canal pendant le temps où elle se fait) est fiable mais peu robuste (si la ligne de communication est coupée, la connexion s'interrompt) ; le second, où les données à transmettre sont découpées en paquets et où chaque paquet peut avoir son propre parcours sur le réseau entre expéditeur et destinataire, peut manquer de fiabilité, mais il est robuste (la connexion ne pourrait cesser que si toutes les routes possibles entre expéditeur et destinataire devenaient impraticables). Les concepteurs de l'Internet, poursuit l'auteure, ont tenté de concilier « la robustesse du mode non connecté et la fiabilité du mode connecté » (p. 10) : de là le procédé de la *commutation de paquets* dont les règles sont constitutives de l'*Internet Protocol* (IP), mode non connecté « sur lequel on a reconstruit un mode connecté », celui du *Transmission Control Protocol* (TCP, « protocole de contrôle de la transmission »).

Dans l'ouvrage commenté un peu plus haut, Antoine de Clerck consacre un assez long développement (pp. 110-111) à la distinction entre « commutation de circuits » (mode connecté) et « commutation par paquets » (mode non connecté). Son exposé a ce mérite de commencer par la précision que voici : « la commutation est le terme employé pour désigner les techniques de télécommunications qui permettent d'établir une communication entre l'appelant et le destinataire. » Une telle technique « détermine comment va être réalisé l'aiguillage de l'information dans le réseau depuis la source jusqu'à la destination ».

Le modèle client/serveur et le DNS

Les deux protocoles fondamentaux TCP et IP feront l'objet d'un traitement plus détaillé dans la suite de l'ouvrage. Après avoir précisé ce qu'est l'Internet, puis évoqué le mode de communication des données entre les objets qui composent ce réseau, l'auteure précise alors une structure essentielle de son fonctionnement, le *modèle client/serveur*. Qu'appelle-t-on *serveur* au juste ? L'exposé que nous suivons répond dans les termes suivants.

Un **serveur** d'Internet n'est rien d'autre qu'un ordinateur sur lequel « tourne » un programme serveur qui exécute la séquence suivante : 1) attendre la requête d'un client ; 2) exécuter la requête du client ; 3) envoyer la réponse au client ; 4) attendre la requête d'un client, etc. Et cela indéfiniment, c'est-à-dire tant que l'ordinateur est en marche. (pp. 10-11)

Il existe, souligne l'auteure, de multiples types de serveur. Ainsi un serveur Web « fournira les pages web que lui demanderont ses clients ». De la même façon un serveur de fichiers « fournira les fichiers demandés ». À propos des serveurs de messagerie, elle précise ceci.

... un serveur de messagerie se tiendra prêt à recevoir des messages électroniques de ses utilisateurs. Dans ce cas, exécuter la requête du client pourra être : « déposer » le message dans la boîte aux lettres du destinataire, et la réponse renvoyée au client pourra être : « OK message envoyé. » (p. 11)

On a là un bon exemple d'une description nouménologique à propos de phénomènes connus de tout usager du courrier électronique mais interprétés ici à l'aide des métaphores constitutives du modèle client/serveur. Si cette nouménologie ne permet certes pas de reconstituer les phénomènes en cause – elle nous laisse sur le seuil d'une

phénoménotechnique, pour parler comme Bachelard⁵² –, elle leur fournit un arrière-plan supposé créateur d'intelligibilité sinon d'opérationnalité.

La situation devient plus complexe avec la section suivante – la dernière du chapitre 1, intitulée « Le domain name system ». Ce qui devrait être premier du point de vue technologico-théorique dans l'exposé proposé est ici la notion de *numéro IP*, que la même auteure mettra d'ailleurs en avant, nous l'avons vu, dans son mini-exposé, publié ultérieurement, cité plus haut (Virieux & Chebret, 2006). Dans le cas présent, cette notion n'est mentionnée, de façon liminaire, que pour introduire la notion de *nom de domaine*, ce que l'auteure fait en peu de mots, en indiquant simplement ceci : « les ordinateurs ne se connaissent entre eux que par leur **numéro IP** unique qu'on peut voir dans un premier temps comme un numéro de téléphone ». La problématique de l'introduction de la notion de nom de domaine (et du système de nommage) est ici toute « classique » : elle met en avant le fait qu'il serait plus simple pour un opérateur humain de retenir (par exemple pour consulter le site correspondant) « le nom du site web du CNRS, www.cnrs.fr, plutôt que son numéro : 193.51.136.4, et, pour le consulter, de taper <http://www.cnrs.fr> que <http://193.51.136.4> » dans la boîte d'adresse de son navigateur préféré. Cette dernière affirmation ne va guère de soi⁵³.

*Le numéro IP donné ici ne renvoyait plus, début avril 2008, à la page indiquée. Le 9 avril 2008, l'utilisation de la commande ping renvoyait le numéro 193.55.76.228. La commande **ping** s'obtient sur Windows XP en cliquant sur **démarrer** puis sur **Exécuter...** Apparaît alors une boîte de dialogue où l'on saisit **command** : en cliquant sur **OK**, on fait apparaître une fenêtre où l'on saisit **ping www.nom_de_domaine.tld** : on voit s'afficher alors le numéro IP du serveur Web sur lequel se trouve le site (voir par exemple http://www.infoprati.net/astuces/internet/astuces/divers_001.php). Cette commande peut être utile à l'internaute qui ne parvient pas à se connecter à un site : en tapant dans la boîte d'adresse de son navigateur l'adresse IP renvoyée par la commande **ping**, il obtient normalement, sinon le site, du moins le serveur où celui-ci est hébergé ; s'il n'en est pas ainsi, le problème vient vraisemblablement de l'hébergeur du site.*

La suite est plus encore problématique : l'auteure écrit de façon imprécise que l'on peut « associer à chaque adresse IP un nom unique lui aussi », ces couples – formés d'une adresse IP et du nom correspondant – constituant un vaste annuaire mondial organisé d'une façon arborescente au sein du *Domain Name System* (DNS, système de noms de domaine). Le traitement d'une telle question – la question des « noms de domaine » – s'avère en vérité

extrêmement délicat. La bijection alléguée entre adresses IP et noms de domaine, en particulier, procède d'une simplification qui vise à éclairer le lecteur en lui dissimulant en partie la notion de nom de domaine. Ayant évoqué rapidement l'arbre de nommage, la racine de cet arbre, qui est « le domaine père de tous les autres », ainsi que le premier niveau de noms de domaines (com, edu, net, org, fr, etc.), l'auteure précise que le « schéma de nommage est basé sur les organisations et non pas sur les réseaux », précisant à ce propos : « Ce sont les organisations – entreprise, université, fournisseur d'accès – qui peuvent demander un nom de domaine [...]. Depuis quelques années on voit même apparaître des noms de domaine correspondant à une personne. » La situation évoquée a évolué depuis, avec la multiplication de « sites personnels » ou de sites créés par une personne ou un collectif sur un thème donné.

Dans ce dernier cas notamment, le nom du site peut faire penser – de façon éventuellement induite – que celui-ci émane d'une institution dépositaire d'une certaine autorité, interprétation qui peut être renforcée encore par le choix du domaine de premier niveau. Par exemple, s'il s'agit du domaine .com, le site pourra être regardé comme procédant d'une entité commerciale contrôlée (et donc « sérieuse »), situation qui peut porter à accorder au site un crédit ou, du moins, une prétention à être cru sur parole que le ou les créateurs du site ne revendiquaient nullement a priori. (L'article « .com » de l'encyclopédie Wikipédia indique ceci : « .com est un nom générique du domaine de premier niveau d'Internet. Il est l'un des premiers domaines de premier niveau, établis en janvier 1985. Initialement prévu pour désigner les entités commerciales, aucune restriction n'était prévue pour l'enregistrement, et rapidement il est devenu le domaine le plus commun. » Chacun peut donc obtenir un « point-com ».)

Ce que nous noterons surtout, c'est que, pour faire entendre l'usage des noms de domaine, l'auteure se tourne alors vers la phénoménologie de leur emploi par l'utilisateur ordinaire (ou presque), comme on le voit dans le passage suivant.

La création d'un nom de domaine est assujettie à l'autorisation du domaine père. Si l'on veut ajouter un sous-domaine à univ-paris7.fr le nom de domaine de l'université Paris 7, il faut s'adresser au gestionnaire des noms de domaine de cette université. On pourra par exemple créer math.univ-paris7.fr pour le département de mathématiques de cette même université. (p. 14)

Cela fait, elle revient à des aspects nouménologiques en consacrant un assez long développement (quelque 5 pages sur 58, soit environ 8 % du texte imprimé) au rôle des noms de domaines dans le fonctionnement de l'Internet. Elle introduit la notion de *serveur de noms* et celle de *résolveur*, « petit programme » dont l'interaction avec les serveurs de noms est illustrée de façon suggestive par la narration suivante, que nous reproduisons *in extenso*.

Imaginons qu'Aline, étudiante parisienne, souhaite consulter un site web, par exemple www.google.fr, depuis son ordinateur situé sur le campus de Jussieu. Son butineur (Netscape, Mozilla ou autre) va donc demander au résolveur de lui donner le numéro IP de www.google.fr.

- 1) le résolveur interroge d'abord le serveur de noms du campus de Jussieu (shiva.jussieu.fr) ;
- 2) Shiva ne connaît pas l'adresse IP de www.google.fr, mais connaît la liste des serveurs de noms de la racine et il interroge donc l'un des treize serveurs de noms de la racine ;
- 3) le serveur de noms de la racine ne connaît pas l'adresse IP de www.google.fr, mais il connaît la liste des serveurs de noms du domaine fr. Il envoie donc cette liste à Shiva ;
- 4) Shiva envoie sa requête concernant www.google.fr à l'un des serveurs de fr, par exemple à dns.inria.fr ;
- 5) celui-ci ne connaît toujours pas l'adresse IP de www.google.fr, mais il connaît la liste des serveurs de noms du domaine google.fr et l'envoie à Shiva ;
- 6) le serveur de nom du domaine google.fr connaît l'adresse IP de www.google.fr et l'envoie à Shiva ;
- 7) Shiva envoie cette réponse au résolveur de l'ordinateur d'Alice et enregistre cette information dans sa mémoire cache (nous reviendrons sur cette notion), de sorte que si une autre personne du campus de Jussieu souhaite consulter le site www.google.fr, Shiva pourra lui en donner immédiatement le numéro IP, sans repasser par toutes les étapes précédentes. (pp. 16-18)

L'auteure note à cet égard l'importance de la *mémoire cache*, dont le fonctionnement « allège énormément le trafic sur le réseau », en soulageant fortement les treize serveurs de la racine qui, sinon, « seraient saturés de requêtes DNS et ne pourraient plus fournir ».

Selon l'article « Serveur racine du DNS » de Wikipédia citant une étude de 2003, seulement 2 % des requêtes adressées aux serveurs de la racine sont légitimes : « Les mises en cache incorrectes ou inexistantes sont à l'origine de 75 % des demandes. 12,5 % concernent des demandes pour des domaines de premier niveau inconnus, 7 % parce qu'elles traitent des adresses IP comme des noms de domaines, etc. Certains ordinateurs de bureau mal configurés tentent même de mettre à jour les enregistrements

des serveurs racine, ce qui n'est pas normal. » (Pour plus d'information, voir <http://www.bind9.net/dnshealth/>.)

La référence aux treize serveurs « de la racine » relève à l'évidence, pour l'utilisateur ordinaire d'aujourd'hui, de la nouménologie de l'Internet. Notons à cet égard un point de détail par où passe la frontière actuelle entre phénoménologie et nouménologie : pour beaucoup d'utilisateurs informés, l'écriture d'un nom de domaine s'arrête à celle du *top level domain* (TLD), c'est-à-dire à l'un des domaines de premier niveau évoqués ci-dessus ; en fait, il conviendrait de noter aussi le domaine racine, qui se marque par un... « espace vide » (*null space*) que l'on note donc en ajoutant simplement un point (*dot*) – « suivi », si l'on peut dire, d'un *null space* – à la suite du nom du TLD. On écrira donc par exemple « *www.cnrs.fr.* » au lieu de l'habituel « *www.cnrs.fr* »⁵⁴.

Le cas des noms de domaine

Le positionnement par rapport à la frontière au tracé parfois incertain, mouvant, entre phénoménologie et nouménologie est un paramètre important de détermination du contenu d'un exposé relatif à telle ou telle question touchant la structure de l'Internet ou son fonctionnement. L'attribution de noms de domaines du type *.fr* est confiée (dans le cadre du RIR RIPE NCC) à l'AFNIC – l'Association Française pour le Nommage Internet en Coopération (laquelle gère aussi les noms de domaine du type *.re*, relatifs à l'île de la Réunion). Voici la présentation que son site offre à l'utilisateur intéressé – client potentiel d'un *registrar*, bureau d'enregistrement permettant le dépôt d'un nom de domaine – de la notion même de nom de domaine : on y reste constamment dans la phénoménologie de la chose (Afnic, 2007).

Initiation

Qu'est-ce qu'un nom de domaine ?

Définition

Le nom de domaine est une des clés indispensables de l'internet. Il est utilisé dans l'ensemble des services (web, messagerie, etc.).

Prenons l'exemple d'une société nommée « AB-CD ».

Cette société est présente sur l'internet :

- l'adresse de son site web est : <http://www.ab-cd.fr> ;
- sa messagerie électronique est du type : contact@ab-cd.fr.

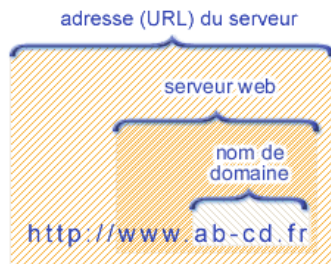
Dans toutes ces adresses, il y a une partie commune :

ab-cd.fr = le nom de domaine.

À quoi correspond... ?

- **www.** : c'est ce qu'il faut rajouter au nom de domaine pour obtenir le **nom du serveur web** (www.ab-cd.fr)

- **http://** : suivi du nom du serveur web, c'est **l'adresse (URL) du serveur web** (<http://www.ab-cd.fr>)



Structure

Un nom de domaine est constitué :

- d'un **nom** (celui d'une société, d'une marque, d'une association, d'un particulier...) ;
- et d'une **extension** (appelée aussi suffixe) : .fr, .de, .ca, .jp, .net, .com, .org, etc.

Pour en savoir plus sur ces différentes extensions.

Syntaxe

Le nom de domaine doit être uniquement composé à partir des caractères suivants :

- A à Z ;
- 0 à 9 ;
- le tiret ;
- aucune différence n'est faite entre les lettres majuscules et minuscules ;
- tous les autres caractères ne sont pas admis (accents, apostrophe, etc.) à ce jour.

Comme on le voit, la présentation se tient au plus près de ce qui apparaît à l'utilisateur désireux de se doter d'un nom de domaine – qui lui fournira la base de l'adresse de son site Web s'il en a un, d'une part, et de l'adresse de sa messagerie électronique, d'autre part.

En théorie pourtant, les situations qui pourraient exister – et qui, pour certaines d'entre elles, existent – sont plus complexes. Dans une section intitulée *How DNS works in theory* de l'article “Domain Name System” de Wikipedia, on lit d'abord en effet ceci : “a domain name usually consists of two or more parts (technically *labels*), separated by dots. For example

example.com.” Mais la suite de ce passage montre alors la relative richesse du système de nommage théoriquement disponible.

- The rightmost label conveys the [top-level domain](#) (for example, the address `www.example.com` has the top-level domain `com`).
- Each label to the left specifies a subdivision, or [subdomain](#) of the domain above it. Note: “subdomain” expresses relative dependence, not absolute dependence. For example: `example.com` comprises a subdomain of the `com` domain, and `www.example.com` comprises a subdomain of the domain `example.com`. In theory, this subdivision can go down to 127 levels deep. Each label can contain up to 63 characters. The whole domain name does not exceed a total length of 255 characters. In practice, some [domain registries](#) may have shorter limits.

Théoriquement, donc, un nom de domaine pourrait comporter par exemple 4 labels de 63 caractères (ce qui ferait 252 caractères en tout), ou, à un autre extrême, 127 labels de 2 caractères (ce qui ferait donc 254 caractères en tout). On peut trouver sur Internet des amateurs de “longest domain names” ; l’un d’eux propose ainsi le nom suivant, qu’il ne faut évidemment pas prendre au sérieux, mais dont le label de deuxième niveau comporte exactement 63 caractères :

`www.thelongestdomainnameintheworldandthensomeandthensomemoreandmore.com.`

Le chapitre que nous suivions s’achève sur la mention de l’attaque du 21 octobre 2002 contre les treize serveurs de la racine (Virieux, 2004, p. 19) : deux d’entre eux seulement résistèrent. Depuis, au reste, une attaque de semblable amplitude a eu lieu de nouveau, le 6 février 2007 ; elle n’a cependant mis à mal, sur les six serveurs attaqués, que deux d’entre eux, situés l’un et l’autre aux États-Unis et gérés, le premier par le département américain de la défense, le second par l’ICANN (ICANN, 2007).

Le modèle TCP-IP

L’auteure poursuit par un chapitre intitulé « Les cinq défis d’Internet », ce qui la conduit alors à introduire son lecteur, non sans courage, à une réalité nouménologique que les exposés pour le grand public n’évoquent, en règle générale, que de fort loin : *le modèle TCP/IP*, « du nom de ces deux plus célèbres protocoles », soit le modèle nommé en anglais *the five-layer TCP/IP model*, formé donc de cinq « couches » (*layers*) que l’on présente habituellement comme suit.

5. Application layer
4. Transport layer
3. Network/Internet layer
2. Data link layer
1. Physical layer

L'auteure procède de haut en bas, en commençant donc par la couche 5, « le niveau application ». Ici les protocoles principaux sont le protocole HTTP (*HyperText Transfer Protocol*) pour consulter les pages Web, le protocole SMTP (*Simple Mail Transfer Protocol*) pour la messagerie électronique, le protocole FTP (*File Transfer Protocol*), etc. Tout en notant que les applications les plus populaires sont la messagerie et le Web, l'auteure ajoute ceci ⁵⁵.

Il en existe beaucoup d'autres : le transfert de fichiers, déjà évoqué, le *chat* (dialogue en direct entre plusieurs ordinateurs, très prisé des ados), la mise à l'heure automatique des ordinateurs à partir d'un serveur de temps, etc. Ces applications classiques sont en voie d'être rattrapées, en termes de volume des données transférées, par des applications comme Kazaa, e-donkey, limewire. Ces logiciels fonctionnant selon le modèle du *peer-to-peer* (traduit en français par le « paire à paire », car les ordinateurs sont tour à tour serveur et client) permettent des échanges de données, en général de la musique et des films, entre internautes anonymes. (pp. 24-25).

La description du modèle TCP/IP se poursuit avec le niveau 4, « le niveau transport ». À ce niveau, il s'agit de « faire en sorte que deux ordinateurs distants de plusieurs kilomètres sachent établir entre eux une communication puis s'envoyer des données sans perte ni altération » (p. 21). Le protocole répondant aux exigences propres à ce niveau est le protocole TCP (*Transmission Control Protocol*), utilisé tant par la messagerie que pour la consultation de pages Web, le transfert de fichiers ou le *peer-to-peer*. Un autre protocole du niveau transport, l'UDP (*User Datagram Protocol*), est mentionné : au contraire du protocole TCP, il ne vise pas la fiabilité avant tout, mais la rapidité de transfert, recherchée pour le transport de la voix en temps réel ou les images fournies par une webcam, opérations qui donnent lieu à une perte de données supposée moins gênante pour l'utilisateur que ne le serait un allongement de la durée du transfert. Les couches suivantes, à savoir les couches 3, 2 et 1, sont passées en revue de façon beaucoup plus rapide. On y apprend surtout que le niveau 3, « le niveau réseau », est le niveau où se situe le protocole IP (*Internet Protocol*) : c'est le

niveau où, grâce à des *routeurs*, il faut « faire en sorte qu'un paquet de données trouve son chemin à travers Internet » (p. 21). Le chapitre se termine par un de ces tableaux très vivants dont l'auteure a le secret : à nouveau elle y met en scène Aline (de Paris) s'efforçant d'envoyer un message à son amie Yoko, étudiante à Tokyo.

Son ordinateur dispose d'un logiciel qui connaît parfaitement le protocole SMTP. Pour dialoguer avec l'ordinateur qui va réceptionner le message, ce logiciel s'appuie sur le logiciel de la couche transport qui implémente le protocole de transmission TCP, que nous nommerons TCP pour une lecture plus facile. Donc SMTP sait que les données qu'il confie à TCP seront transmises sans erreur et les lui transmet. TCP place les données dans un segment TCP, format qui lui convient, avec des en-têtes. Puis il transmet ce segment TCP à IP. Celui-ci découpe ce segment en un ou plusieurs datagrammes IP en fonction de la taille maximale de datagramme, puis envoie chaque datagramme à ses couches liaison de données et physique qui les « emballent » de nouveau dans un format de données qui convient au réseau local (trame ethernet pour un réseau ethernet par exemple). Chaque datagramme ainsi emballé est envoyé au routeur le plus proche. Ensuite chaque datagramme est envoyé de routeur en routeur à travers Internet jusqu'à l'ordinateur de destination, traversant ainsi des réseaux de toutes sortes. Arrivés à destination, les datagrammes sont réceptionnés par la couche IP de la machine de destination et transmis à TCP. TCP vérifie que tout est bien arrivé, demande à la machine source de renvoyer ce qui a été perdu puis, quand tout est complet, passe les données à l'application. (pp. 26-27).

On remarquera l'approximation explicitement signalée par l'auteure : ce qui est « le logiciel de la couche transport qui implémente le protocole de transmission TCP » devient ici, par abus de langage, le « TCP » de l'ordinateur. Bien entendu, le TCP – *transmission control protocol* – est, comme son nom l'indique, un protocole, c'est-à-dire un ensemble de règles formelles⁵⁶. La même remarque peut évidemment être faite à propos du SMTP, etc. On voit ici se tracer une portion de la ligne de démarcation entre phénoménologie et nouménologie : évoquer TCP ou SMTP comme des protocoles lorsqu'on est un « simple utilisateur » relève de la nouménologie ; la réduction phénoménologique en fait des applications présentes sur l'ordinateur de l'utilisateur. Notons aussi que le long passage cité ci-dessus, sur lequel s'achève le chapitre examiné, se conclut par une notation en suspens : quel est en effet cette mystérieuse application à laquelle « TCP » devrait « passer les données » ? La réponse ne viendra que dans le chapitre suivant, intitulé « La messagerie », où sont mentionnés « les protocoles Pop et Imap » (p. 35) – chapitre sur lequel nous ne nous arrêterons pas davantage

ici ⁵⁷. Ce chapitre en précède un autre, intitulé « Le web », et qui est lui-même suivi de quatre chapitres d'approfondissement que nous parcourons rapidement ci-après.

Le contrôle de la transmission des données

Le premier des quatre chapitres annoncés est intitulé « TCP, le protocole de contrôle de la transmission ». On a vu que le TCP devait assurer une transmission fiable des données (lesquelles circulent par paquets, en mode non connecté). Ce contrôle de fiabilité utilise, lui, on l'a noté aussi, le mode connecté, parce que « l'établissement d'une connexion permet de réserver des ressources (essentiellement de la mémoire) sur les machines clients et serveur afin de pouvoir contrôler la transmission des données » (p. 44) ⁵⁸. Mais comment le logiciel implémentant le protocole TCP contrôle-t-il la transmission des données ? Chaque octet envoyé par la machine est numéroté, ce qui permet de l'identifier. Les données à transmettre sont découpées en segments de taille convenable pour les deux machines, chaque segment étant identifié par le numéro du dernier octet qui le compose. Laissons alors la parole à l'auteure.

Les données de l'application, par exemple le message électronique, sont découpées en segments de taille convenable pour les deux machines. Un segment est identifié de manière unique par le numéro du dernier octet qu'il contient. Lorsqu'une machine envoie un segment, elle le garde en mémoire jusqu'à ce que la machine réceptrice lui envoie un acquittement pour ce segment. L'acquittement, c'est un segment TCP spécial. Si la machine émettrice ne reçoit pas un acquittement pour le segment envoyé avant un certain temps, elle réémet ce segment. Cela suppose donc que pour chaque segment envoyé, la machine enclenche un minuteur.

Puisque les segments sont numérotés de façon unique, on peut savoir s'ils sont reçus dans l'ordre et les réordonner le cas échéant. On peut aussi repérer les segments envoyés en double et les rejeter : en effet, si l'acquittement d'un segment se perd, le destinataire recevra ce segment en double. (pp. 44-45)

Le contrôle de la transmission est alors décrit à gros traits. Une erreur consiste à transformer un 0 en 1 ou l'inverse. Pour contrôler que rien de tel ne s'est produit, une « somme de contrôle » calculée par l'émetteur figure dans l'en-tête de chaque segment TCP. Le destinataire refait ce calcul : s'il ne trouve pas la même valeur, « il rejette le segment et attend que l'émetteur le lui renvoie ». L'auteure ne dit rien de la structure de la somme de contrôle utilisée (qui ne se réduit pas à être la somme des bits du segment considéré !), ce

qu'on peut comprendre⁵⁹. Le développement précédent est alors complété par de rapides considérations sur les échanges entre les deux machines afin de gérer la vitesse des flux en fonction des disponibilités en mémoire (*buffer*). Enfin, l'auteure introduit le chapitre suivant par une affirmation et une question : « Pour circuler sur le réseau, un segment TCP est emballé dans un datagramme IP. Comment les datagrammes IP trouvent-ils leur chemin à travers Internet ? » (p. 46) C'est cela qu'elle va révéler à son lecteur dans le chapitre intitulé « Le routage des données sur Internet ».

Adresses IP et routage des données

Le chapitre commence par reprendre et compléter les informations données à propos des adresses IP, en indiquant d'abord la structure d'une telle adresse, selon le protocole Internet dans sa version 4 (IPv4).

La structure d'une adresse IP est *a.b.c.d*, où *a*, *b*, *c* et *d* sont des nombres pouvant prendre des valeurs comprises entre 0 et 255 (par exemple : 217.12.3.11). Les premiers octets représentent le numéro du réseau, et le ou les derniers octets représentent le numéro de machine. Par exemple, tous les ordinateurs de l'université de Paris 6 ont des adresses IP qui commencent par 134.157.

Au passage, on remarque que le nombre maximal d'adresses IP dont on puisse disposer est de plus de quatre milliards. Ce nombre, qui semblait gigantesque au départ, se révèle aujourd'hui trop juste. C'est en partie pour cette raison qu'une nouvelle version du protocole IP (IPv6) remplacera prochainement la version actuelle (IPv4). De nouvelles fonctionnalités ont été ajoutées et les adresses seront codées sur cent vingt-huit bits au lieu de trente-deux. (pp. 47-48)

Ce passage mérite un commentaire particulier si on le regarde comme adressé à un profane doté de curiosité plus que de science. L'auteure semble identifier les entiers compris entre 0 et 255 (par exemple 217, ou 12, ou 3, ou 11, ou 134, ou 157) à des *octets*, c'est-à-dire à des séquences de 8 bits. En fait il s'agit des entiers codables sur 8 bits, et donc s'écrivant dans le système binaire avec 8 chiffres 0 ou 1. Par exemple 134 s'écrit, en base 2, sous la forme : 10000110 ; de même, 217 s'écrit 11011001 ; et 157 s'écrit 10011101. Les entiers s'écrivant avec 8 chiffres en binaire sont les entiers compris entre 0 et un maximum qui, en base 2, s'écrit donc 11111111, nombre qui, en base 10, vaut $1 + 2 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 + 2^7 = 255$. D'où l'identification entre un entier compris entre 0 et 255 et un octet ; mais la chose n'est pas évidente pour un lecteur que peuvent dérouter tant les habitudes de langage

« informatiques » que les aspects mathématiques laissés implicites, même quand ils sont élémentaires.

On a là, au demeurant, un exemple très simple qui rappelle que les « savoirs du numérique » restent dans la proximité de tout travail un peu précis sur le fonctionnement d'Internet. Par exemple, il n'est pas déplacé de savoir calculer que les « plus de quatre milliards » de possibilités offertes par le protocole IPv4 sont en vérité au nombre de $256^4 = 4\,294\,967\,296$, soit presque 4,3 milliards.

Dans le cas du protocole IPv6, où les adresses sont codées sur 128 bits, le nombre d'adresses disponibles est énormément augmenté : par rapport au système IPv4, ce nombre est en effet multiplié par

$$2^{128}/2^{32} = 2^{96} = 79\,228\,162\,514\,264\,337\,593\,543\,950\,336,$$

soit plus de $7,9 \times 10^{28}$. Le nombre total, soit 2^{128} , s'écrit, quant à lui :

$$340\,282\,366\,920\,938\,463\,463\,374\,607\,431\,768\,211\,456.$$

Il est donc supérieur à $3,4 \times 10^{38}$.

La suite du chapitre a trait aux « datagrammes IP ». Chaque datagramme comporte, outre le segment TCP à transmettre, un en-tête précisant l'adresse IP de l'ordinateur qui envoie le datagramme et l'adresse IP de l'ordinateur qui est appelé à le recevoir. Le point essentiel sur lequel l'attention du lecteur est attirée est que, une fois créés ces « petits paquets de l'Internet » que sont les datagrammes IP, une aventure commence : « Les datagrammes comprenant les données d'un même message peuvent [en effet] prendre des routes différentes. Rien ne garantit qu'ils arrivent tous à destination, ni qu'ils arrivent en bon état. » (pp. 48-49) C'est à cause de ce genre d'incertitudes que le protocole IP est dit « non fiable ». Son mérite essentiel est en fait de permettre le *routage* des données. L'auteur explicite le rôle des routeurs reliés à plusieurs réseaux et de leur *table de routage*, ainsi que la notion de « route par défaut » (lorsque le destinataire n'est pas directement accessible à partir du routeur considéré), avant de distinguer le *routage statique* – comme il en va pour un réseau local connecté à l'Internet par un unique routeur (appelé passerelle : *gateway*, en anglais), muni d'une table de routage qui ne varie pas, et qui sert souvent de pare-feu entre l'intérieur et l'extérieur du réseau local – et le *routage dynamique*, nécessaire pour les réseaux plus vastes. Dans ce dernier cas, les routeurs vont « calculer en permanence la meilleure route pour attendre les autres réseaux » (p. 51) et modifier en conséquence leur table de routage.

Quoique voulu élémentaire, l'exposé est sur ce point explicite : l'auteure précise ainsi que grâce à un protocole de routage dynamique, « les routeurs échangent en permanence (toutes les trente secondes) des informations sur les destinations qu'ils peuvent atteindre et le temps nécessaire pour les atteindre ». Elle précise de même que les protocoles de routage ne sont pas les mêmes selon que l'on est à l'intérieur d'un même système autonome, où le protocole le plus utilisé est l'OSPF (*Open Shortest Path First*), ou que le routage doit se faire d'un système autonome à un autre. Dans ce cas le protocole de routage le plus souvent utilisé est le BGP (*Border Gateway Protocol*). Ce qu'il importe de souligner, par contraste avec le routage intra-système, c'est que des problèmes de « frontière » ont de fortes chances de surgir : « si le plus court chemin entre deux machines A et B est un routeur R situé dans un système autonome différent de ceux de ces deux machines, le trafic entre A et B ne passera pas forcément par lui », dans la mesure où « le système autonome auquel appartient R peut décider de refuser le trafic de transit... » (p. 52).

La couche liaison de données

Le chapitre qui suit s'intitule « Les protocoles de la couche liaison de données ». Cette couche, on s'en souvient, est la couche numéro 2 du modèle TCP/IP ; en fait, le chapitre envisage tant la couche 2 que la couche 1, soit la *couche physique*. L'auteure consacre d'abord un assez long développement aux *réseaux Ethernet*, qui sont « des réseaux locaux d'ordinateurs reliés entre eux par un même câble de cuivre grâce à leurs cartes Ethernet » (p. 54). Sur un tel réseau, qui peut fort bien ne pas être relié à Internet, les ordinateurs ont une « adresse Ethernet », qu'il ne faut pas confondre avec l'adresse Internet attribuée à l'ordinateur en cas de connexion au réseau des réseaux.

Cette adresse Ethernet, précise l'auteure, est « de la forme 00:03:ba:5e:45:3b ». À nouveau, ici, le lecteur profane se trouve devant un profond mystère quant à la signification de la chaîne alphanumérique présentée. En fait, la forme indiquée renvoie à une séquence de 6 entiers écrits ici en base 16, c'est-à-dire en numération hexadécimale, où les « chiffres » sont les 16 symboles suivants 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. Les nombres désignés par les lettres A, B, C, D, E, F sont les entiers qui, en notation décimale, s'écrivent respectivement 10, 11, 12, 13, 14, 15. L'entier maximal qu'on peut écrire est FF, qui désigne l'entier $15 \times 16^1 + 15 \times 16^0 = 255$. Il y donc en principe 256^6 adresses Ethernet possibles, alors que, selon le protocole IPv4, il n'y a que 256^4 adresses Internet possibles⁶⁰.

Si l'auteure a choisi ne pas s'attarder sur ce sujet, elle est plus diserte, en revanche, à propos de ce qui semble être son sujet de prédilection, la circulation des datagrammes IP.

Lorsqu'un tel datagramme doit circuler entre deux ordinateurs sur un réseau Ethernet, note-t-elle, il est « emballé » dans une « trame Ethernet » (en anglais, *Ethernet frame*). Toujours attentive à cet aspect essentiel du fonctionnement de réseaux, l'auteure précise alors ceci.

S'il doit parcourir plusieurs réseaux avant d'arriver à destination, un même datagramme pourra être emballé dans plusieurs trames Ethernet sans être affecté par ces emballages et déballages successifs. Avant d'émettre une trame Ethernet, un ordinateur doit connaître l'adresse Ethernet de la machine destinataire ; en général il ne connaît que son adresse IP. Pour ce faire, il envoie une requête à toutes les machines du réseau en demandant quelle est l'adresse Ethernet de la machine dont l'adresse IP est « tant ». La machine en question reçoit alors la requête et y répond en envoyant son adresse Ethernet. (pp. 55-56)

Elle poursuit en examinant rapidement le cas d'un ordinateur connecté à Internet par le réseau téléphonique (par le truchement d'un modem). Le protocole de liaison entre l'ordinateur de l'utilisateur et les ordinateurs du fournisseur d'accès, qui doivent lui fournir une adresse IP, est généralement le protocole PPP, le *protocole point à point* (en anglais, *point to point protocol*). Dans ce cas on ne peut plus utiliser la ligne téléphonique pour téléphoner dès lors que l'ordinateur est connecté. Il en va autrement avec l'ADSL – nous en avons déjà parlé –, dont le seul démerite, comparativement, est d'exiger un équipement spécifique non seulement chez l'utilisateur mais aussi chez l'opérateur téléphonique (en sorte que l'accès par ADSL n'est aujourd'hui pas encore disponible en tout point du territoire).

Une culture appropriée est une culture large

Le chapitre suivant, l'avant-dernier de l'ouvrage, est intitulé *L'organisation humaine autour d'Internet* : il aborde la question des organisations gestionnaires de l'Internet, question sur laquelle nous ne reviendrons pas ici. Enfin, le dernier chapitre, « Un outil à préserver », s'arrête d'abord sur les « misères » de l'Internet dues aux virus et aux spams, en évoquant un certain nombre de parades à ces formes de délinquance ou d'intrusions abusives. L'auteure termine par une exhortation à la vigilance, pour que l'Internet « ne nous soit pas confisqué » (p. 62).

Nous arrêterons là notre repérage des praxéologies de l'Internet dont on peut soupçonner l'utilité pour la RII. On imagine aisément que l'exploration des cinq couches du modèle TCP/IP – ou des sept couches du modèle OSI (*Open Systems Interconnection*) – pourrait continuer presque indéfiniment, au triple niveau théorique, technologique et

technique. Or, même si ce vaste domaine comporte des parties dont l'examen peut être ardu et laborieux pour le néophyte, il est surtout remarquable de constater que nombre de connaissances qu'il serait loisible de rajouter à notre petit inventaire ne paraissent pas *a priori* étrangères à ce que pourrait être une culture riche, équilibrée, et pour cela efficace, de la RII, *sans pour autant être plus complexes* que la plupart des éléments praxéologiques inventoriés jusqu'ici.

Donnons un seul exemple : celui du *bit error rate* (BER, dénommé aussi *bit error ratio*), le « taux de bits erronés » (TBE), nommé aussi en français « taux d'erreur binaire » (TEB). Ce taux est défini comme le rapport du nombre de bits erronés reçus au nombre de bits total reçus : il mesure ainsi la qualité d'un canal de communication. Sans doute peut-on en faire une étude plus ou moins approfondie⁶¹ ; mais on sera déjà quelque peu éclairé si l'on sait que, en téléphonie, un taux d'erreur binaire supérieur à 10^{-4} gêne les conversations téléphoniques et qu'il faut maintenir en conséquence ce taux au-dessous de 10^{-5} , tandis que, pour la transmission des données sur Internet (par exemple), il convient de descendre en dessous de 10^{-12} (ce qui représente en moyenne un bit erroné sur mille milliard de bits). Comme la transmission de données utilise des réseaux ayant des taux d'erreur binaire « téléphoniques », ces erreurs doivent absolument être détectées et corrigées : on retrouve alors les considérations déjà rencontrées à propos de la « somme de contrôle » (*checksum*)⁶².

D'une manière générale, une bonne maîtrise des environnements praxéologiques de la RII peut supposer une culture *large* (sinon toujours approfondie) relative à la structure et au fonctionnement de l'Internet. À propos d'un thème sur lequel nous nous sommes arrêtée brièvement plus haut, celui des noms de domaines, un spécialiste américain connu, Danny Sullivan, cité par Greg Notess, indique ceci.

... I find many people don't understand how the domain name system works. That's essential for making better use of site-specific searching, so it can be beneficial to take more time to teach Internet basics so they can do better searching. (Notess, 2006, p. 86)

Que sont donc ces « bases » de l'Internet appropriées à la RII ? Nous verrons en fait dans les chapitres suivants qu'un certain nombre de « notions » restent à préciser. Mais nous voudrions ici, pour terminer, formuler une remarque très simple, qui nous ramènera au début de cette exploration. Presque à chaque instant, dans le trajet qui conduit de la mise à disposition d'un ordinateur à son utilisation pour la RII, il est possible, en pratique, soit d'opérer par soi-même, soit d'appeler à l'aide, dans un cadre amical ou commercial, un « spécialiste » qui prendra en charge l'opération à réaliser. Ainsi en ira-t-il pour la configuration d'un poste de travail et sa

connexion à l'Internet, opération pour laquelle il n'est pas difficile de trouver des « modes d'emploi » ⁶³, ou encore une société privée proposant un service à domicile approprié ⁶⁴. Mais même dans ce dernier cas, où l'utilisateur potentiel n'aurait pas à mettre la main à la pâte, une culture de l'Internet est nécessaire, sinon dans un but d'opérationnalité phénoménologique, du moins dans une visée d'intelligibilité nouménologique – laquelle, bien entendu, ne saurait être un obstacle, mais au contraire une aide importante à l'opérationnalité phénoménologique – à la « phénoménotechnique ». C'est en ayant ce principe en tête que nous aborderons la partie peut-être la plus brûlante de notre exploration, celle relative au *World Wide Web*.

CHAPITRE 14

Le World Wide Web

Pour amorcer le repérage du domaine praxéologique correspondant au *World Wide Web*, nous reprendrons ici le petit ouvrage de Françoise Virieux qui, au chapitre précédent, nous a accompagné dans le repérage du domaine de l'Internet. Dès la première page de son exposé, alors qu'est évoquée la mise en place progressive de ce qui deviendra l'Internet, le Web est mentionné par l'auteure, qui écrit : « Le web, l'application qui a fait sortir Internet du monde de la recherche, est né en 1994 au CERN. » (p. 5) L'ayant mentionné à nouveau dans sa description de la couche 5 du modèle de référence TCP/IP, où elle évoque le protocole HTTP « pour la consultation des pages web » (p. 24), elle lui consacre ensuite tout un chapitre, le quatrième, après avoir traité dans le troisième de « La Messagerie ». Ce chapitre, intitulé « Le web », nous l'avons vu, s'ouvre par la présentation suivante : « Le web est un système qui permet d'accéder facilement à une multitude de documents reliés entre eux et répartis sur des millions d'ordinateurs dans le monde. » (p. 37) L'auteure note ensuite le contexte institutionnel dans lequel le Web a vu le jour, celui du CERN, autour duquel s'affairaient un grand nombre de chercheurs, d'ingénieurs et de techniciens entre lesquels devaient s'échanger des documents situés sur des centaines d'ordinateurs répartis « dans toute l'Europe ». Elle mentionne alors le nom de Tim Berners-Lee, puis l'expression de « toile d'araignée », en signalant à cet égard que cela se dirait *web* en anglais.

Rappelons que, étymologiquement, le mot anglais web désigne a woven fabric, une toile obtenue par tissage. Le sens de spider's web apparaît au XIII^e siècle. Le mot désigne encore (à partir de la fin du XVI^e siècle), la membrane qui joint les orteils d'un canard et autres animaux aquatiques. Dans tous les cas on trouve l'idée de lien entre des entités distinctes formant par cela une entité de niveau supérieur.

Le propos est rapide, allusif plus qu'explicatif : la technique pour relier des documents sur le Web, note sobrement l'auteure, « s'appelle de l'hypertexte » ; et le premier programme qui a permis « de consulter la Toile », est devenu opérationnel en février 1993. Très vite, des sites

Web ont été créés, d'abord par des institutions d'enseignement ou de recherche, ensuite par des organisations à but commercial. Tout cela, bien entendu, est fort succinct.

Le Web et le modèle client-serveur

Le chapitre comporte ensuite trois sections intitulées respectivement « Butineur et serveur web », « Le protocole http », « Les moteurs de recherche ». L'auteure retrouve en ce point une plus grande latitude d'explicitation. Elle rappelle d'abord que le Web relève, à l'instar d'autres applications de l'Internet, du modèle *client-serveur*, les clients étant ici les ordinateurs des usagers – à condition qu'ils soient munis d'un « butineur » (ou navigateur). Pour l'utilisateur, la phénoménologie est fort simple (lorsque le système fonctionne correctement) : connaissant l'adresse d'un document, il sollicite son navigateur, lequel affiche bientôt ce document à l'écran ; et il en va de même si l'utilisateur clique sur un lien pointant vers le document attendu.

La nouménologie, elle, est un peu plus complexe. Ainsi l'auteure s'arrête-t-elle un instant sur la structure d'une URL, sigle dont elle précise la signification en anglais – *Uniform Resource Locator* – et pour lequel elle propose la traduction de « localisateur de ressources unifié ». À propos de l'adresse <http://www.jussieu.fr/index.html> qu'elle prend pour seul exemple, elle indique que *http* y désigne le protocole utilisé, le HTTP, elle précise que *www.jussieu.fr* est le nom du serveur Web où se trouve le document dont on a l'adresse, et que, enfin, *index.html* est le nom du document sur le serveur.

Le fait qu'il en soit ainsi appelle un commentaire. On lit sur le site Scotsmist (http://www.scotsmist.co.uk/glossary_i.html) l'indication suivante adressée aux internautes : “When you type in the URL of a web site, but do not specify an actual file name (e.g. <http://home.talkcity.com>, the server will look for a file named ‘index.html’ and will display it on your browser. This is the home page for the web site. You can display the same page by typing <http://home.talkcity.com/index.html>.” Une autre indication, adressée, elle, aux créateurs de sites, complète cette notation et l'éclaire d'un autre point de vue (Levine, Baroudi & Levine Young, 2006, p. 413) : « Il est indispensable, ou fortement recommandé, de nommer votre page d'accueil index.html. Ainsi, lorsque quelqu'un tente d'accéder au site sans spécifier le nom de la page, si le fichier est index.html, la page d'accueil apparaîtra. C'est une sécurité pour vous : sans cette page, de nombreux hébergeurs retournent une page d'erreur. Évitez cela en pensant à nommer votre page d'accueil index.html. » Par ailleurs, à

propos d'une particularité des URL très rarement commentée, les auteurs de l'ouvrage Web Dragons écrivent (Witten, Gori, & Numerico, 2007, p. 66) : "Incidentally, people often wonder about the meaning of the mysterious double slash (//). After a lecture on the evolution of the web in 1999, Tim Berners-Lee was asked, 'If you could go back and change one thing about the design, what would it be?' His characteristically low-key response was that he would use http: rather than http://. Originally he thought the two slashes looked nice, but now he regrets the countless hours that humanity wastes every year typing them."

Un serveur Web, précise ailleurs l'auteure que nous suivons, est un ordinateur « où sont stockés des documents à publier sur le Web et sur lequel tourne un programme en attente des requêtes d'un client » (p. 39). La communication entre client et serveur qui va utiliser un logiciel implémentant le protocole HTTP. À nouveau, l'auteure gratifie son lecteur d'une description vivante autant qu'éclairante : nous la reproduisons ici.

Les commandes que peut envoyer le client au serveur sont par exemple :

- GET : pour demander l'affichage d'une page ;
- POST : pour envoyer des données saisies dans un formulaire ;
- HEAD : pour demander juste l'en-tête de la page (pour connaître la date de dernière modification de la page par exemple).

Lorsque Aline tape l'URL <http://www.cnrs.fr>, son navigateur établit une connexion avec la machine www.cnrs.fr puis lui envoie une commande GET. Le serveur renvoie alors la liste des objets à afficher sur la page (un objet peut être du texte, une image, un film). Le client établit alors plusieurs connexions parallèles avec le serveur pour accélérer le transfert des objets. (pp. 40-41)

La manœuvre n'a pas à être réalisée à chaque fois : grâce à la *mémoire cache*, précise l'auteure, si l'utilisateur demande l'affichage d'une page, il n'est plus nécessaire (sauf bien sûr si la mémoire cache a été vidée entre temps) que le navigateur aille solliciter le serveur ⁶⁵.

Le Web en lui-même

La dernière section du chapitre, qui occupe une page sur les cinq que compte ce bref exposé, a trait, on l'a noté, aux moteurs de recherche, question sur laquelle nous reviendrons dans la quatrième partie de notre travail. C'est ici l'occasion de noter un phénomène que le repérage auquel nous nous livrons ne permet pas d'ignorer : s'il existe un certain nombre d'exposés sur

l'Internet, s'il en existe aussi sur la RII, s'il en existe pareillement sur les moteurs de recherche (et plus généralement sur les outils de la RII), il en existe peu sur le Web seul. Le petit ouvrage de Françoise Virieux relève, de ce point de vue, d'un cas de figure où l'Internet est mis en avant, le Web n'en constituant qu'une application possible parmi quelques autres. Un autre cas de figure existe encore, plus fréquent sans doute, où, au lieu d'arriver au Web, si l'on peut dire, ainsi que le faisait l'auteure mentionnée, on part du Web en quelque sorte, c'est-à-dire qu'on le *suppose* et que, à partir de là, on entreprend non pas de l'étudier, mais de découvrir comment « surfer » sur ce réseau de pages plus ou moins reliées entre elles. C'est ainsi que, dans un gros ouvrage de presque huit cents pages intitulé *Le livre de l'Internet*, voulu très opérationnel pour l'utilisateur de base (et, pour cela, plus phénoménologique que nouménologique), Henri Lilen (2008) offre à son lecteur, après une première partie de quelque 120 pages intitulée « Se connecter à Internet », une deuxième partie, de presque 100 pages, intitulée « Surfer sur le Web ».

Sans ignorer de tels exposés – qui relèvent du vaste ensemble des exposés relatifs au Web en général –, nous prendrons d'abord pour outil de repérage des exposés plus proprement centrés sur le Web lui-même, en commençant par les articles “World Wide Web” consacrés à notre sujet par l'encyclopédie Wikipedia, dans sa version en anglais d'une part, dans sa version en français de l'autre. À titre d'illustration, nous avons reproduit ci-après la table des matières de l'article en anglais : on verra ainsi d'un simple coup d'œil que la matière traitée fait peu de place, cette fois, aux techniques permettant de surfer sur le Web ⁶⁶.

1. How the Web works / 2. History / 2.1 History in literature / 3. Standards / 4. Java and JavaScript /
5. Publishing Web pages / 6. Statistics / 7. Speed issues / 8. Caching / 9. Link rot and Web archival /
10. Academic conferences / 11. Security / 12. Web Accessibility / 13. WWW prefix in Web addresses / 14. Pronunciation of “www” / 15. See also / 16. References / 17. External links

D'emblée, l'article propose cette définition du Web : “The **World Wide Web** (commonly shortened to **the Web**) is a system of interlinked hypertext documents accessed via the Internet.” L'exposé se poursuit alors ainsi : “With a Web browser, a user views Web pages that may contain text, images, videos, and other multimedia and navigates between them using hyperlinks.” Avec la section “How the Web works”, on quitte la phénoménologie de l'utilisateur – qui se réduit pour l'essentiel à taper une URL dans un navigateur Web ou à suivre un hyperlien – et on retrouve alors un paysage déjà familier grâce à l'ouvrage de F. Virieux (2004) : le texte examiné ici le présente quant à lui dans les termes suivants ⁶⁷.

First, the server-name portion of the URL is resolved into an IP address using the global, distributed Internet database known as the domain name system, or DNS. This IP address is necessary to contact and send data packets to the Web server.

The browser then requests the resource by sending an HTTP request to the Web server at that particular address. In the case of a typical Web page, the HTML text of the page is requested first and parsed immediately by the Web browser, which will then make additional requests for images and any other files that form a part of the page. Statistics measuring a website's popularity are usually based on the number of "page views" or associated server "hits", or file requests, which take place.

Having received the required files from the Web server, the browser then renders the page onto the screen as specified by its HTML, CSS, and other Web languages. Any images and other resources are incorporated to produce the on-screen Web page that the user sees.

Dans une version *antérieure* de l'article que nous suivons – par exemple la version du 22 novembre 2007 –, on trouvait, entre la première section, "How the Web works", et la deuxième, "History", une section intitulée "Caching" – qui se retrouve, dans la version précédemment citée, on l'a vu, en huitième position. Inchangée dans son contenu d'une version à l'autre, que contient-elle ? L'exposé qui y est consigné prolonge, en le complétant quelque peu, celui que nous avons commenté plus haut en suivant l'exposé dû à F. Virieux. Nous sommes ici dans la nouménologie du fonctionnement des navigateurs Web. Le texte indique par exemple que "almost all Web browsers cache recently-obtained data, usually on the local hard drive", avant de préciser : "HTTP requests sent by a browser will usually only ask for data that has changed since the last download. If the locally-cached data is still current, it will be reused". Le texte mentionnera plus loin l'existence de "facilities built into Web servers that can determine when files have been updated and need to be re-sent". Bien entendu toutes les requêtes ne sont pas traitées de la même façon : une requête GET est "likely to be cached if other conditions are met" ; en revanche une requête POST "is assumed to depend on the data that was POSTed and so is not cached." La question du *caching* n'est pas sans importance du point de vue de la phénoménologie vécue par l'utilisateur : il n'est pas indifférent que l'utilisateur du Web accède à cet égard à un point de vue nouménologique adéquat, non seulement vis-à-vis de son propre ordinateur, mais aussi à propos des moteurs de recherche qui, eux aussi, "store cached content from websites".

Notons pour le contraste que si, dans Le livre de l'Internet déjà mentionné, Henri Lilen (2008) se réfère aux fichiers Internet temporaires dans son chapitre 10 intitulé

« Utiliser l'historique de navigation », en revanche les notions de cache et de caching sont absentes de son exposé, dont l'opérationnalité phénoménologique évidente va ainsi de pair avec une profondeur nouménologique souvent des plus réduites.

En vérité, la nouménologie proposée dans l'article *World Wide Web* ne permet guère d'expliquer complètement des phénomènes pourtant fréquents, et parfois irritants pour l'utilisateur, lequel peut aussi bien, pour cela, être induit en erreur, par exemple par les décalages souvent rencontrés entre version de la page affichée et version réellement disponible sur le serveur. Avec le navigateur Internet Explorer, ainsi, un paramétrage permet à l'utilisateur de choisir entre quatre options : celui-ci peut demander au navigateur de vérifier s'il existe une version plus récente des pages mises en cache 1) à chaque visite d'une telle page Web, 2) à chaque démarrage d'Internet Explorer, 3) automatiquement, 4) jamais. La troisième option (« automatiquement ») paraît, au reste, obscure : le navigateur est muet à propos des paramètres et des modalités de fonctionnement de l'automatisme qui entre en jeu lorsque cette option est choisie.

On trouve, sur le site « Aide et support » de Microsoft, un article relatif à Internet Explorer 4.x et Internet Explorer 5 intitulé « Impact des paramètres de mise en cache Internet Explorer sur la navigation Web » (<http://support.microsoft.com/kb/263070/fr>). L'option « Automatiquement », qui n'existe que sur Internet Explorer 5 et les versions ultérieures, repose sur « un algorithme logique qui permet à Internet Explorer de comprendre les habitudes de comportement de navigation de l'utilisateur ». Internet Explorer, précise-t-on à l'adresse de l'utilisateur, « vérifie la présence de nouveau contenu uniquement lorsque vous revenez sur une page que vous avez déjà visualisée dans une session précédente d'Internet Explorer ou un autre jour. Avec le temps, si Internet Explorer détermine que des images sur la page sont rarement modifiées, il vérifie encore moins souvent la présence de nouvelles images. » Comprenez qui pourra !

La genèse du Web : concepts et outils

La version en français de l'article « World Wide Web » commence par indiquer que le Web « est un système hypertexte public fonctionnant sur Internet et qui permet de consulter, avec un navigateur, des pages mises en ligne dans des sites ». Cette présentation concise, mais qui

a le mérite de nommer les personnages principaux du monde du Web, est suivie d'une mise en garde préalable que nous reproduisons ici.

Le Web n'est qu'une des applications d'Internet. D'autres applications d'Internet sont le courrier électronique, la messagerie instantanée, Usenet, etc. Le Web a été inventé plusieurs années après Internet, mais c'est lui qui a rendu les médias grand public attentifs à Internet. Depuis, le Web est fréquemment confondu avec Internet ; en particulier, le mot *Toile* est souvent utilisé dans les textes non techniques sans qu'il soit clairement exprimé s'il désigne le web ou l'Internet.

L'un et l'autre article sur le World Wide Web – en français et en anglais – consacre des développements à l'histoire du Web. La version en anglais fait commencer cette histoire en 1980 avec les premiers essais, au CERN, de Tim Berners-Lee. Le travail alors engagé trouvera sa conclusion fin 1990 : Berners-Lee dispose à cette date du premier *navigateur* Web (qui, à l'instar de ses successeurs immédiats, est aussi un *éditeur* de pages Web), du premier *serveur* Web et des premières *pages Web* – qui, en l'espèce, décrivent le projet de World Wide Web lui-même⁶⁸. Les événements s'enchaînent : le 6 août 1991, Berners-Lee publie dans le newsgroup `alt.hypertext` un résumé du “WorldWideWeb project” où il distingue le point de vue de celui qui prend connaissance de l'information (*reader view*) et le point de vue de celui qui fournit l'information (*information provider view*). En ne retenant ici que le point de vue de l'utilisateur (au sens restreint : celui qui « lit » l'information), voici *in extenso* la description par son créateur de la situation fondamentale du Web⁶⁹.

The WWW world consists of documents, and links. Indexes are special documents which, rather than being read, may be searched. The result of such a search is another (“virtual”) document containing links to the documents found. A simple protocol (“HTTP”) is used to allow a browser program to request a keyword search by a remote information server.

The web contains documents in many formats. Those documents which are hypertext, (real or virtual) contain links to other documents, or places within documents. All documents, whether real, virtual or indexes, look similar to the reader and are contained within the same addressing scheme.

To follow a link, a reader clicks with a mouse (or types in a number if he or she has no mouse). To search and [*sic*] index, a reader gives keywords (or other search criteria). These are the only operations necessary to access the entire world of data.

En décembre 1991, le premier serveur non européen est créé au *Stanford Linear Accelerator Center* (SLAC). Le 30 avril 1993, les autorités du CERN annoncent que le World Wide Web

est mis dans le domaine public, le CERN renonçant à tout droit de propriété intellectuelle sur le code, chacun étant autorisé à l'utiliser, y compris en le modifiant ⁷⁰. L'article de Wikipedia ajoute ⁷¹ : "Coming two months after the announcement that the Gopher protocol was no longer free to use, this produced a rapid shift away from Gopher and towards the Web." Après le navigateur *ViolaWWW*, qui devint populaire dans un public encore très restreint, paraît en 1993 le navigateur Mosaic, qui constitue un tournant, notamment parce qu'il est le premier navigateur *graphique*, c'est-à-dire capable d'intégrer les images au texte au lieu de les présenter dans une fenêtre séparée ⁷². Mais bientôt, en quelques mois, un nouveau venu, *Netscape Navigator*, supprime Mosaic. Le 24 août 1995, Microsoft lance le système d'exploitation Windows 95. Celui-ci ne comporte pas de navigateur Web intégré ; aussi, fin 1995, Microsoft lance *Internet Explorer*, qui aura finalement raison de Netscape Navigator : celui-ci n'est plus développé depuis le 1^{er} mars 2008, ses utilisateurs étant invités à utiliser le navigateur *Firefox* ⁷³.

Comme d'autres exposés sur le sujet, l'article de Wikipedia souligne que l'idée cruciale de Berners-Lee fut de marier deux entités déjà existantes – l'*hypertexte*, d'une part, l'*Internet*, d'autre part – alors même que les acteurs du développement de chacune de ces entités n'étaient en général pas intéressés par l'autre, selon le témoignage de Berners-Lee lui-même dans son livre *Weaving the Web* (2000). Par rapport à d'autres systèmes hypertextes, le Web ne comportait que des liens unidirectionnels, ce qui "made it possible for someone to link to another resource without action by the owner of that resource". En outre, contrairement à ce qui se passait avec le logiciel HyperCard, propriété de la société Apple Computer ⁷⁴ –, le World Wide Web "was non-proprietary, making it possible to develop servers and clients independently and to add extensions without licensing restrictions", tout en rendant un tel développement beaucoup plus simple. L'article souligne toutefois une difficulté apparemment indépassable, celle liée au *link rot*, c'est-à-dire au phénomène par lequel les liens présents sur un site Web se périssent, par exemple parce qu'ils renvoient à une page Web dont le contenu a été profondément modifié ou qui a tout simplement disparu – situation que l'irritant message 404 (*not found*) a fait connaître à nombre d'utilisateurs du Web. On notera à ce propos la création en 1996 de l'*Internet Archive*, qui est sans doute le plus connu des systèmes d'archivage du Web et permet de retrouver des états antérieurs de celui-ci.

Un glossaire du Web

L'article en français « World Wide Web » de l'encyclopédie Wikipédia propose un utile inventaire de « termes rattachés au Web ». Ce tableau fait d'abord mention de mots ou

expressions qui relèvent du lexique des réseaux ou, plus particulièrement, de l'Internet, et pas seulement du Web : ainsi de l'expression *en ligne* qui « signifie “connecté à un réseau”, en l'occurrence le réseau informatique internet ». Ainsi en va-t-il encore du mot *hôte*, qui désigne un ordinateur en ligne : « Chaque hôte d'Internet est identifié par une adresse IP à laquelle correspondent zéro, un ou plusieurs noms d'hôte. Cette terminologie n'est pas propre au Web, mais à Internet. » Voici alors l'inventaire proposé au lecteur : il fournit une table d'orientation nécessairement incomplète mais significative par rapport au domaine examiné.

- Une **ressource du World Wide Web** est une entité informatique (texte, image, forum Usenet, boîte aux lettres électronique, etc.) **accessible** indépendamment d'autres ressources. Une ressource en accès **public** est librement accessible depuis Internet. Une ressource **locale** est présente sur l'ordinateur utilisé, par opposition à une ressource **distante** (ou en ligne), accessible à travers un réseau.
- On ne peut accéder à une ressource distante qu'en respectant un **protocole de communication**. Les fonctionnalités de chaque protocole varient : réception, envoi, voire échange continu d'informations.
- **HTTP** (pour *HyperText Transfer Protocol*) est le protocole de communication communément utilisé pour transférer les ressources du Web. **HTTPS** est la variante sécurisée de ce protocole.
- Une **URL** (pour *Uniform Resource Locator*) **pointe** sur une ressource. C'est une chaîne de caractères permettant d'indiquer un protocole de communication et un emplacement pour toute ressource du Web.
- Un **hyperlien** (ou **lien**) est un élément dans une ressource associé à une URL. Les hyperliens du Web sont orientés : ils permettent d'aller d'une source à une destination. Seule la ressource à la source contient les données définissant l'hyperlien, la ressource de destination n'en porte aucune trace. Il existe deux types d'hyperlien : ceux du premier type doivent être activés pour accéder à la destination ; ceux du second causent un accès automatique à la destination.
- **HTML** (pour *HyperText Markup Language*) est un langage informatique permettant de décrire le contenu d'un document (titres, paragraphes, disposition des images, etc.) et d'y inclure des hyperliens. Un **document HTML** est un document décrit avec le langage HTML. Les documents HTML sont les ressources les plus consultées du Web.
- Dans un mode de communication **client-serveur**, un **serveur** est un hôte sur lequel fonctionne un logiciel serveur auquel peuvent se connecter des logiciels clients fonctionnant sur des *hôtes clients*.
- Un **serveur web** est un hôte sur lequel fonctionne un **serveur HTTP** (ou **serveur web**). Un serveur web **héberge** les ressources qu'il dessert.

- Un **navigateur web** est un logiciel client HTTP conçu pour accéder aux ressources du Web. Sa fonction de base est de permettre la consultation des documents HTML disponibles sur les serveurs HTTP. Le support d'autres types de ressource et d'autres protocoles de communication dépend du navigateur considéré.
- Une **page web** (ou **page**) est un document destiné à être consulté avec un navigateur web. Une page web est toujours constituée d'une ressource centrale (généralement un document HTML) et d'éventuelles ressources liées automatiquement accédées (typiquement des images).
- Un **éditeur HTML** (ou **éditeur Web**) est un logiciel conçu pour faciliter l'écriture de documents HTML et de pages web en général.
- Un **site web** (ou **site**) est un ensemble de pages web et d'éventuelles autres ressources, liées dans une structure cohérente, publiées par un propriétaire (une entreprise, une administration, une association, un particulier, etc.) et hébergées sur un ou plusieurs serveurs Web.
- **Visiter** un site web signifie « consulter ses pages ». Le terme **visite** vient du fait que l'on consulte généralement plusieurs pages d'un site, comme on visite les pièces d'un bâtiment. La visite est menée par un **utilisateur** (ou **visiteur** ou **internaute**). La mesure d'audience est obtenue en copiant le code en javascript d'un lien vers le site d'un prestataire spécialisé suivant la technique du marqueur à distance.
- Une **adresse web** est une URL de page web, généralement écrite sous une forme simplifiée limitée à un nom d'hôte. Une adresse de site web est en fait l'adresse d'une page du site prévue pour accueillir les visiteurs.
- Un **hébergeur web** est une entreprise de services informatiques hébergeant (mettant en ligne) sur ses serveurs web les ressources constituant les sites web de ses clients.
- Une **agence web** est une entreprise de services informatiques réalisant des sites web pour ses clients.
- L'expression **surfer sur le Web** signifie « consulter le Web ». Elle a été inventée pour mettre l'accent sur le fait que consulter le Web consiste à suivre de nombreux hyperliens de page en page. Elle est principalement utilisée par les médias ; elle n'appartient pas au vocabulaire technique.
- Un **annuaire web** est un site web répertoriant des sites Web.
- Un **portail web** est un site web tentant de regrouper la plus large palette d'informations et de services possibles dans un site Web. Certains portails sont thématiques.
- Un **service web** est une technologie client-serveur basée sur les protocoles du Web.

Dans ce qui suit, nous commenterons un certain nombre des entrées précédentes.

Les RFC et la notion de ressource

Une première notion clé, et en fait fondatrice, est la notion de *ressource* (en anglais, *resource*). Avant de la commenter, toutefois, il convient de dire quelques mots de documents qui ont joué et continuent de jouer un rôle clé dans la vie de l'Internet et (donc) du Web : les *Requests For Comments* ou RFC : un tel développement, qui, en bonne méthode, aurait dû trouver place lors du repérage du domaine Internet, fournit un exemple (mineur) de besoin de connaissance qui, émergeant en un domaine donné, relève en fait d'un domaine antécédent ou connexe.

Un request for comments, littéralement une « demande de commentaires », est un document d'une longue série de documents, dont le premier – le RFC 1 – a été publié, sous la signature de Steve Crocker et la responsabilité du Network working group, le 7 avril 1969. En mai 2008 a paru le RFC 5280, toujours à l'enseigne du Network working group, référence dont le site du RFC Editor – entité officielle financée par l'Internet Society pour publier et gérer les RFC – précise qu'elle est aujourd'hui “historical in form and symbolic in content” et doit désormais être interprétée comme désignant “the set of users, vendors, and researchers who are working to improve and extend the Internet, in particular under the ISOC/IETF umbrella” (<http://www.rfc-editor.org/rfcfaq.html#wg>). Tout « expert » peut en principe rédiger un Internet Draft (brouillon Internet) et le soumettre à l'IETF ; le chemin, alors, sera long qui, à travers généralement la constitution et le fonctionnement d'un groupe de travail, en fera – dans certains cas seulement – un full Internet Standard (STD). Le STD 66, l'un des plus récents, est ainsi le RFC 3986 publié en janvier 2005 par T. Berners-Lee, R. Fielding et L. Masinter sous le titre Uniform Resource Identifier (URI): Generic Syntax.

La notion de ressource a subi depuis son introduction un processus de généralisation rythmé par une succession de RFC. Dans le RFC 1630 (juin 1993) est introduite la notion générale d'URI, *Universal Resource Identifier*, accompagnée de deux variantes, l'*URL*, sigle qui signifie alors *Universal Resource Locator*, et l'*URN*, ou *Universal Resource Name*. Est ressource tout ce qui a une identité, cette dernière notion se scindant elle-même en deux, l'*adressage* (soit la localisation) et le *nommage* (soit l'identité au sens strict). Le RFC 1738 (décembre 1994) marque deux modifications. Tout d'abord l'adjectif *universal* est remplacé par l'adjectif *uniform* : on parlera de “*Uniform Resource Locator*”, etc. Ensuite, de façon plus

implicite, le mot de ressource est davantage employé hors des sigles déjà mentionnés – alors que le RFC 1630 parlait simplement des « objets du Réseau » (*objects on the Network, objects on the Internet*). Mais ce n'est en fait qu'en août 1998 que le RFC 2396 propose une définition explicite de « ressource » (ou plutôt de l'anglais *resource*), définition que le RFC 3986, mentionné plus haut, viendra, en janvier 2005, confirmer tout en le précisant dans les termes suivants⁷⁵.

... the term “resource” is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with a consistent purpose (e.g., “today's weather report for Los Angeles”), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., “parent” or “employee”), or numeric values (e.g., zero, one, and infinity).

La notion de ressource désigne ainsi tout ce qu'on peut être amené à rechercher sur le Web afin d'en extraire une information.

Une notion subtilement complexe : la notion d'URL

Dans le glossaire auquel nous nous référons, l'entrée *Ressource* est suivie par une entrée relative au HTTP, sur lequel nous ne reviendrons pas ici. Puis vient la notion d'URL, sur laquelle nous nous arrêterons maintenant. Dans l'article “Uniform Resource Locator” de Wikipédia (en français, malgré le titre), on précise que « les URL sont les seules URI ayant trouvé une application pratique », du moins jusqu'ici. Le même article rappelle qu'une URL « est une chaîne de caractères combinant les informations nécessaires pour indiquer à un logiciel comment accéder à une ressource d'Internet ». Les informations en question, y précise-t-on encore, « peuvent notamment comprendre le protocole de communication, un nom d'utilisateur, un mot de passe, une adresse IP ou un nom de domaine, en numéro de port TCP/IP, un chemin d'accès, une requête ». En outre, « un identificateur de fragment peut être ajouté à la fin d'une URL pour identifier un élément à l'intérieur de la ressource ». Cette description comporte sans doute, pour beaucoup d'utilisateurs attentifs, une surprise qui ne devient évidente que lorsque l'on donne un exemple d'une telle URL ; dans l'article en question l'exemple choisi est le suivant :

<http://Jojo:1ApIn@www.exemple.com:8888/chemin/d/acc%C3%A8s.php?q=req&q2=req2#signet>

On voit que, après la classique double barre oblique //, là où on s’attendrait à rencontrer le tout aussi classique « triple w » (www), on trouve tout à fait autre chose, avec même une apparente incongruité : l’occurrence du signe *arrobe* ou *arobase* (@) ⁷⁶. En l’espèce, une enquête express révèle que ce type de syntaxe est largement inconnu parmi les utilisateurs raisonnablement éclairés de l’Internet. Nous rencontrons ici une situation d’un type que, en vérité, nous exploiterons au plan méthodologique dans la troisième partie de notre travail : elle conduit à se demander s’il est possible de trouver des « exposés » concernant le Web qui répondent à la question de savoir si un tel format d’adresse existe bien, ou si, par exemple, il s’agirait d’une bévée, voire – pourquoi pas ? – d’une facétie malicieuse de quelque rédacteur anonyme de l’encyclopédie Wikipédia ⁷⁷. Nombre d’exposés, en vérité, sont muets sur la question : ainsi en va-t-il par exemple de l’article de Wikipedia en anglais intitulé “Uniform Resource Locator”. Dans le cas particulier envisagé, cependant, il est un type d’exposés qui ne peut manquer de répondre à la question : les *Requests For Comments* (RFC). En l’espèce, c’est le RFC 1738 daté de décembre 1994 et déjà mentionné qui répond à notre question, ainsi que le montre le passage reproduit ci-après.

While the syntax for the rest of the URL may vary depending on the particular scheme selected, URL schemes that involve the direct use of an IP-based protocol to a specified host on the Internet use a common syntax for the scheme-specific data:

`//<user>:<password>@<host>:<port>/<url-path>`

Some or all of the parts “<user>:<password>@”, “:<password>”, “:<port>”, and “/<url-path>” may be excluded. The scheme specific data start with a double slash “//” to indicate that it complies with the common Internet scheme syntax. (Berners-Lee, Masinter & McCahill, 1994, p. 4)

On notera encore cette explicitation : “The user name (and password), if present, are followed by a commercial at-sign ‘@’. Within the user and password field, any ‘:’, ‘@’, or ‘/’ must be encoded.”(p. 4) La notion de “scheme” couvre un ensemble de protocoles qui ne se réduisent pas, bien entendu, au seul protocole HTTP, ainsi que le montre la liste ci-après.

ftp	File Transfer protocol
http	Hypertext Transfer Protocol
gopher	The Gopher protocol
mailto	Electronic mail address

news	USENET news
nntp	USENET news using NNTP access
telnet	Reference to interactive sessions
wais	Wide Area Information Servers
file	Host-specific file names
prospero	Prospero Directory Service

Nous ne considérerons ici – très sommairement – que ce qui concerne le protocole HTTP, à propos duquel le document que nous suivons indique ceci.

An HTTP URL takes the form:

`http://<host>[:<port>]/<path>?<searchpart>`

where <host> and <port> are as described in Section 3.1. If <port> is omitted, the port defaults to 80. No user name or password is allowed. <path> is an HTTP selector, and <searchpart> is a query string. The <path> is optional, as is the <searchpart> and its preceding “?”. If neither <path> nor <searchpart> is present, the “/” may also be omitted.

Ce qu’on désigne par [host] est le “fully qualified domain name” (FQDN, nom de domaine complètement spécifié), qui a la forme d’une “sequence of domain labels separated by ‘.’...”

Tous ces éléments d’information, pris à bonne source, se retrouvent au reste, pour l’essentiel, dans un autre exposé, celui du site *Comment ça marche*, où l’on peut lire ce qui suit⁷⁸.

Qu’est ce qu’une URL ?

Une URL (Uniform Resource Locator) est un format de nommage universel pour désigner une ressource sur Internet. Il s’agit d’une chaîne de caractères ASCII imprimables qui se décompose en cinq parties :

Le nom du protocole : c’est-à-dire en quelque sorte le langage utilisé pour communiquer sur le réseau. Le protocole le plus largement utilisé est le protocole HTTP (HyperText Transfer Protocol), le protocole permettant d’échanger des pages Web au format HTML. De nombreux autres protocoles sont toutefois utilisables (FTP, News, Mailto, Gopher...).

Identifiant et mot de passe : permet de spécifier les paramètres d’accès à un serveur sécurisé. Cette option est déconseillée car le mot de passe est visible dans l’URL.

Le nom du serveur : Il s'agit d'un nom de domaine de l'ordinateur hébergeant la ressource demandée. Notez qu'il est possible d'utiliser l'adresse IP du serveur, ce qui rend par contre l'URL moins lisible.

Le numéro de port : il s'agit d'un numéro associé à un service permettant au serveur de savoir quel type de ressource est demandée. Le port associé par défaut au protocole est le port numéro 80. Ainsi, lorsque le service Web du serveur est associé au numéro de port 80, le numéro de port est facultatif.

Le chemin d'accès à la ressource : Cette dernière partie permet au serveur de connaître l'emplacement auquel la ressource est située, c'est-à-dire de manière générale l'emplacement (répertoire) et le nom du fichier demandé.

On aura noté en passant la remarque sur le « danger » de faire apparaître ce par quoi l'utilisateur s'identifie auprès du serveur Web (à savoir <user:password>), crainte qui peut expliquer l'absence généralisée de ce segment facultatif d'une URL. Le même exposé complète l'explicitation précédente en synthétisant la situation dans le tableau que voici ⁷⁹.

Protocole	Mot de passe (facultatif)	Nom du serveur	Port (facultatif si 80)	Chemin
http://	user:password@	www.commentcamarche.net	:80	/glossair/glossair.php3

Les deux documents déjà mentionnés – le RFC 1738 et l'exposé de *Comment ça marche* – ne sont bien sûr pas les seuls où l'on puisse rencontrer l'indication qu'une URL peut comporter, après la séquence relative au protocole utilisé, l'indication de l'ordinateur d'où l'adresse est activée. L'article de l'encyclopédie Wikipédia intitulé « Adresse Web » contient ainsi le développement suivant, que nous reproduisons pour l'exemple ⁸⁰.

Soit par exemple un document HTML enregistré dans un fichier informatique nommé `page.html` qui se trouve dans le répertoire `/home/tim/` d'un ordinateur appelé `www.example.com`. Selon le moyen utilisé pour y accéder, cette ressource peut être connue sous diverses adresses :

`file:///home/tim/page.html`

La ressource est accessible en tant que fichier local `page.html` dans le répertoire `/home/tim/`. Cette adresse ne fonctionne que sur l'ordinateur où la ressource est enregistrée.

`http://www.example.com/tim/page.html`

La ressource est accessible sur le serveur HTTP `www.example.com`, sous le chemin d'accès `/tim/page.html`. Le serveur HTTP est configuré pour faire correspondre le chemin d'accès `/tim/` au répertoire `/home/tim/`.

`ftp://tim:sEcReT@www.example.com/page.html`

La ressource est accessible en utilisant le protocole FTP (File Transfer Protocol), avec le compte `tim` et le mot de passe `sEcReT`, sur le serveur `www.example.com`. Le serveur FTP est configuré pour donner au compte `tim` le répertoire `/home/tim/`.

Comme nombre d'autres, l'article dont nous extrayons ce passage fait l'objet de la mise en garde suivante : « La pertinence du contenu de cet article est remise en cause. Considérez le contenu de cet article avec précaution. » Si l'on prend cet avertissement à sa valeur faciale, et si l'on s'en tient à la pertinence *stricto sensu*, on peut en conclure que l'élucidation des « mystères » de l'adressage sur le Web paraît à certains superflue – ce qui est consonant avec la rareté de l'information à propos du point de syntaxe que nous avons examiné. Mais ce n'est pas là, nous semble-t-il, toute l'affaire. La page où se trouve discuté l'article « Adresse Web » dans sa version du 17 mars 2008 voit un discutant rappeler que « personne ne devrait écrire quoi que ce soit de substantiel sur le sujet des URI sans avoir personnellement lu les RFC essentielles sur le sujet », ce qui semble en effet raisonnable⁸¹. Pourtant, contestant le titre même de l'article qu'il discute – « Adresse Web » –, cet intervenant écrit aussi : « D'où sort cette notion "d'adresse Web" ? Je n'ai jamais lu en anglais de "web address". » Or, si l'on tape sur Google la requête "**Web address**", le tout premier résultat affiché (le 8 septembre 2008 encore) renvoie à une page Web du site *Learn the Net.com* intitulée... "Understanding Web Addresses"⁸². Ainsi que nous l'avons déjà observé, on voit ainsi un supposé « spécialiste » du Web en ignorer grossièrement les ressources. Mais plus largement, les diverses remarques critiques (et pas uniquement les anomalies ou les bévues qu'elles comportent parfois) appendues à l'article examiné inspirent le sentiment qu'on tient là, avec la question des adresses Web, un point de difficulté dont le « traitement » dans la culture commune de l'Internet *reste actuellement insatisfaisant*. On retrouve ici une donnée centrale en didactique : contre la fiction scolaire de la transparence du « savoir à enseigner », c'est d'abord ce savoir même qui est et reste *problématique*.

Ajoutons à cela un autre fait, mieux connu semble-t-il, certes, mais rarement mentionné dans la littérature sur le Web, concernant l'utilisation du préfixe `www` dans les adresses Web. En l'espèce, on trouve ce fait consigné dans l'article "World Wide Web" de Wikipedia en anglais : dans une section intitulée "WWW prefix in Web addresses", le texte précise en effet qu'il s'agit là d'un simple choix "for a website's subdomain name",

choix qui ne répond cependant à aucune norme technique et participe simplement d'une "long-standing practice of naming internet host (servers) according to the services they provide", en sorte que le nom d'un serveur Web commencera souvent par www, celui d'un serveur FTP par ftp, etc. Il y a là, en très peu de mots, une connaissance qui permet de dissiper certains mystères du « monde Internet » avec lequel l'utilisateur est confronté.

Hyperliens et HTML

L'organisation hypertextuelle des ressources du Web est une donnée essentielle de la RII. Dans un article intitulé "You've Got Hypertext", des spécialistes de *l'hypertext research* (Schaeffel, Carr, De Roure & Hall, 2004) adoptent une problématique qu'ils condensent dans la formule suivante : "Hypertext-as-interaction with information to build associations, and through associations to build knowledge." Nous sommes là tout près d'une problématique de la RII comme outil de construction de réponses à des questions. Or la structure hypertextuelle sur laquelle « travaille » la RII est fondée, on le sait, sur la notion d'*hyperlien* (ou *lien hypertexte*). Dans les chapitres précédents nous avons exploré très largement la phénoménologie des hyperliens du point de vue de l'utilisateur ordinaire. Nous revenons à cette notion ici ; car, si le point de vue antérieurement adopté permet d'agir, il ne permet guère de se représenter ce qui se passe réellement « au-delà de l'écran ». Quand on clique sur un « lien » – en désignant par ce mot un élément de texte ou une image « cliquable » – que se passe-t-il au juste ? Puisque, sauf exception, le navigateur affiche alors (par exemple) une nouvelle page Web, on peut penser que le « clic » fonctionne comme un ordre donné au navigateur d'afficher (après l'avoir téléchargée) une ressource déterminée, localisée à l'aide d'une adresse URL que le clic lui aura, du même coup, fait connaître. Une telle interprétation nouménologique semble, certes, compatible avec la phénoménologie du « clic ». Mais peut-elle être confirmée par la « science du Web » ? Plus précisément, par exemple, par quels mécanismes passe-t-on d'un fragment de texte "colored blue with a blue underline", comme l'écrit Moira Stephen (2003, p. 232), à l'URL de la page ainsi « commandée » au navigateur ?

Le premier problème à souligner à ce stade est, une fois de plus, celui de la relative *pénurie de réponses* pour l'usager ordinaire plongé dans son environnement ordinaire. Pour accéder à une esquisse de réponse, et comme nous l'avons fait plus haut, nous nous tournerons ici vers certains textes en ligne, et notamment vers les articles de l'encyclopédie Wikipédia, en nous situant dans la logique évoquée ci-dessus : par l'interaction avec l'information déposée sur le Web (en l'espèce, celle contenue dans les textes consultés), et à

travers la constitution d'*associations* de tels textes, parvenir à élaborer une connaissance (*knowledge*), qu'il reste toujours à soumettre à la dialectique des médias et des milieux.

Nous partons ici de l'article « Hyperlien », au demeurant fort court, de la version française de l'encyclopédie Wikipédia. Ce texte rappelle d'abord qu'un hyperlien « est une référence dans un système hypertexte permettant de passer automatiquement d'un document consulté à un document lié ». La première partie de l'exposé est intitulée « Source et destination ». On y précise qu'un hyperlien a une *source* ou *origine* et une *destination* ou *cible*, et que l'activation de « l'élément source » fait « passer automatiquement à sa destination ». On y rappelle la nature de la source – mots, phrases, images. On y distingue plusieurs cas quant à la destination : celle-ci peut être interne à la page ; elle peut « également être un autre document, voire un élément précis d'un autre document ». Enfin l'article n'omet pas de signaler les hyperliens « dits brisés, cassés ou morts ».

Rien, jusqu'ici, qui ne soit presque strictement phénoménologique ! L'article se termine par une indication fort imprécise, sinon erronée, sur ce qu'on nomme le *hotlinking*, pratique sur laquelle nous reviendrons un peu plus loin à titre d'illustration. Mais nous passerons sans plus attendre à l'article “Hyperlink” de l'encyclopédie Wikipedia en anglais. Ici, le vocabulaire paraît se faire plus précis. Le lien que « porte » un élément de texte ou une image est clairement distingué de ce « support » : on parle d'*embedded link*. Le lien est *embedded* – « encasté », « enchâssé » – dans l'élément « porteur », à la manière dont un fossile est pris dans un morceau de rocher⁸³. L'enchâssement peut avoir lieu dans une image : on retrouve alors, mais en anglais (*hot area*), ce que nous avons vu Henri Lilen appeler, en français, un « point chaud », et que d'autres auteurs nomment « zone sensible » ou « zone cliquable ». Le phénomène du point chaud correspond, nouménologiquement, à ce qu'on nomme, en HTML, une *image map*, une « carte d'image »⁸⁴. Là se place la première incursion de l'article que nous suivons jusqu'au seuil de la programmation en HTML ; on y lit en effet ceci : “a technical description of a hot area is a list of coordinates relating to a specific area on a screen created in order to hyperlink areas of the image to various destinations...” Nous reviendrons à la question des *image maps* un peu plus loin, de même que nous reviendrons sur la notion d'*inline link* abordée ensuite par l'article, qui se trouve étroitement liée au *hotlinking* déjà mentionné. Après divers développements que nous omettons ici, on en arrive alors à une section intitulée “How hyperlinks work in HTML”, sur laquelle nous nous arrêterons maintenant.

Ainsi qu'on l'a vu de façon plus ou moins explicite, un lien a deux extrémités (*ends*) et un sens (*direction*) : il va de l'une à l'autre. En HTML, ces extrémités sont appelées des « ancrs », en anglais *anchors*. Il y a ainsi l'ancre source (*source anchor*) et l'ancre de

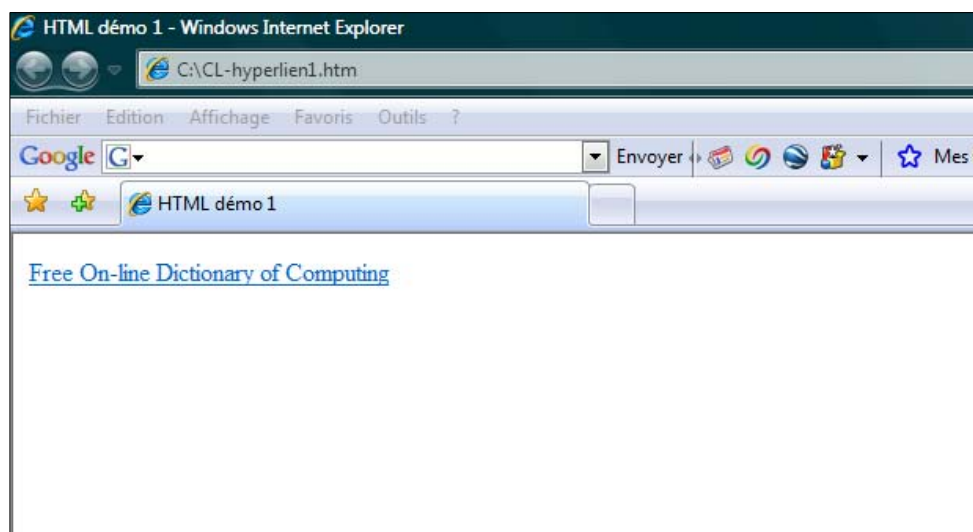
destination (*destination anchor*). Pour le profane, le texte de l'article devient alors rapidement opaque, comme l'illustre déjà le deuxième alinéa du texte que nous suivons et que nous reproduisons ici.

The most common destination anchor is a URL used in the World Wide Web. This can refer to a document, e.g. a webpage, or other resource, or to a position in a webpage. The latter is achieved by means of a HTML element with a "name" or "id" attribute at that position of the HTML document. The URL of the position is the URL of the webpage with "#attribute name" appended—this is a fragment identifier.

Pour élucider ce passage, nous nous appuyons sur un autre texte, l'article du *Free On-Line Dictionary of Computing* (Howe, 1993-2008), ouvert à l'entrée *anchor*. On y apprend d'abord que, en HTML, les ancres sont créées avec la balise ouvrante <a> (*opening tag* ou *start tag*) et la balise fermante (*closing tag* ou *end tag*). Le texte précise : "The opening 'a' tag of a source anchor has an 'href' (hypertexte reference) attribute giving the destination in the form of a URL – usually a whole node or 'page'." L'exemple suivant est alors donné :

```
<a href="http://foldoc.org/">Free On-line Dictionary of Computing</a>
```

Lorsque cette ligne figure dans le code d'une page HTML, elle suscite l'affichage sur la page du fragment de texte « Free On-line Dictionary of Computing », lequel est cliquable (voir la copie d'écran ci-après) et porte un lien dont la destination est alors l'URL <http://foldoc.org> : lorsqu'on clique sur le lien affiché, on arrive ainsi sur la page d'accueil du site FOLDOC.



Voici alors le code HTML quasi minimal qu'il convient de faire apparaître dans un éditeur HTML afin de créer le lien en question (la ligne de code « utile » est la ligne 6 ci-dessous).

```
1 <html>
2 <head>
3 <title>HTML démo 1</title>
4 </head>
5 <body>
6 <a href="http://foldoc.org/">Free On-line Dictionary of Computing</a>
7 </body>
8 </html>
```

On voit que l'ensemble est encadré par la balise ouvrante `<html>` et la balise fermante `</html>`. On trouve ensuite l'en-tête, encadré par les balises `<head>` et `</head>` ; ici cet en-tête se réduit au titre, en l'espèce « HTML démo 1 », lequel est encadré par les balises `<title>` et `</title>`. Puis vient le corps de la page (*body*), placé entre les balises `<body>` et `</body>` ; dans le cas que nous examinons, le contenu se réduit précisément à la ligne de code créant le lien enchâssé dans le texte « Free On-line Dictionary of Computing ».

L'article du dictionnaire FOLDOC que nous examinons évoque ensuite le cas d'un lien qui renvoie à une position à l'intérieur d'une page HTML, position désignée au navigateur sur cette page par la présence d'une ancre ayant la forme suivante :

```
<a name="nomdelaposition">
```

Si l'on veut alors fournir à un navigateur l'adresse de cette position, on tapera dans sa barre d'adresse l'URL de la page suivi d'un signe dièse (#), suivi lui-même de « nomdelaposition » :

```
http://www.nomdedomaine.org/page.html#nomdelaposition
```

Si on veut créer un lien renvoyant à cette position à partir d'une certaine page HTML, on insérera dans le code la ligne suivante :

```
<a href="http://www.nomdedomaine.org/page.html#nomdelaposition">Cliquez ici</a>
```

Revenons un instant à la phénoménologie, avec l'article "Hyperlink" de Wikipedia. Que se passe-t-il lorsque, sur l'écran, le curseur « passe sur un lien » – "hovers over a link" ? Voici la réponse.

When the cursor hovers over a link, depending on the browser and/or graphical user interface, some informative text about the link can be shown:

It pops up, not in a regular window, but in a special hover box, which disappears when the cursor is moved away (sometimes it disappears anyway after a few seconds, and reappears when the cursor is moved away and back). Mozilla Firefox, IE, Opera, and many other web browsers all shows the URL.

In addition, the URL is commonly shown in the status bar.

La formulation précédente n'est pas entièrement dénuée d'ambiguïté : les navigateurs explicitement nommés montrent tous l'URL, et cela non dans la "hover box" qui apparaît fugitivement au-dessus du lien, mais dans la barre d'état (*status bar*). Cette petite imprécision ne saurait pourtant troubler la quiétude des phénomènes. On va voir maintenant que l'internaute peut gagner à savoir que, derrière cette image paisible, bien des choses peuvent se produire, qu'ils ne soupçonnent pas forcément.

Le cas du hotlinking

On a parlé plus haut du *hotlinking*, appelé aussi *inline linking*, *direct linking* ou encore, de manière imagée, *offsite image grab* (on se saisit d'une image située hors de la page que l'on construit), et aussi, de façon très réaliste, de *bandwidth theft* (vol de bande passante). De quoi s'agit-il ? L'article « Hotlinking » de Wikipédia présente ainsi cette pratique, dont nous allons voir qu'elle est discutable sinon toujours répréhensible.

Le direct linking ou hotlinking consiste à utiliser l'adresse d'un fichier publié sur un site web, le plus souvent une image, pour l'afficher sur un autre site, sur un blog, dans un forum, etc. En d'autres termes, au lieu d'enregistrer l'image et de l'installer sur son propre serveur web, le hotlinqueur crée un lien direct vers le serveur d'origine.

Pourquoi ce type de lien est-il dit « chaud » ? Dans la situation examinée ici, l'image (par exemple) qui fait l'objet d'un *hotlink* n'est pas disponible sur le serveur A qui sert la page Web sur laquelle se trouve le *hotlink* vers cette image : chaque fois que l'image doit s'afficher

sur cette page, le navigateur envoie une requête au serveur B qui héberge l'image ; le serveur B répond en transférant l'image en question au navigateur de l'ordinateur client. Plusieurs problèmes en résultent solidairement : tout d'abord, le serveur B peut voir son fonctionnement ralenti si un grand nombre de *hotlinks* le sollicitent ainsi, en lui volant de fait de la bande passante, au point parfois que le site Web contenant l'image pourra être interrompu « s'il consomme trop de bande passante » ; ensuite, même quand l'hébergeur ne prend pas cette décision extrême, il y a toute chance qu'il facture au propriétaire du site B « un supplément pour dépassement du quota de données ».

« Même dans le cas où l'image est libre de droits, lit-on encore dans l'article que nous suivions, un tel lien est illégal s'il n'est pas expressément autorisé par l'éditeur du site Web concerné... » Le problème du *hotlinking* semble s'être posé plus fréquemment avec l'explosion des blogs : les systèmes de gestion de contenu utilisés pour créer et gérer un blog ne permettaient que difficilement de stocker des images ; de là le recours fréquent au *hotlinking*. Un blogueur, Antoine Boucherie, raconte avec humour sa mésaventure (Boucherie, 2007) : ayant « hotlinké » une image qu'il avait trouvée sur un autre blog (en passant par le moteur de recherches d'images de Google, lequel indique l'URL de chacune des images qu'il affiche), il s'est trouvé « puni » par l'auteur du blog d'où cette image provenait, celui-ci ayant réagi en retirant l'image en question pour la remplacer, sous le même nom et avec la même URL, par une autre, à caractère apparemment pornographique... La fabrication d'un *hotlink* sur une page HTML n'est guère compliquée. Voici d'abord le code dans lequel le lien apparaîtra.

```
<html>
<head>
<title>Hotlink 1</title>
</head>
<body>
<a href="URL de l'image">Photo de AB</a>
</body>
</html>
```

Il convient maintenant de se procurer l'URL de la « photo de AB » annoncée dans le code précédent. Celle-ci est affichée sur la page de blog dont l'adresse a été indiquée plus haut. En effectuant un clic droit sur cette photo, et en choisissant « Propriétés » dans le menu déroulant

qui apparaît alors, on fait apparaître une fenêtre dans laquelle est indiquée notamment l'adresse de la photo, à savoir ici :

`http://www.leblogdantoine.com/wp-content/themes/WP_Premium/images/moi100.jpg`

Il suffit alors d'insérer cette adresse dans le fragment de code préparé ci-dessus.

```
<html>

<head>

<title>Hotlink 1</title>

</head>

<body>

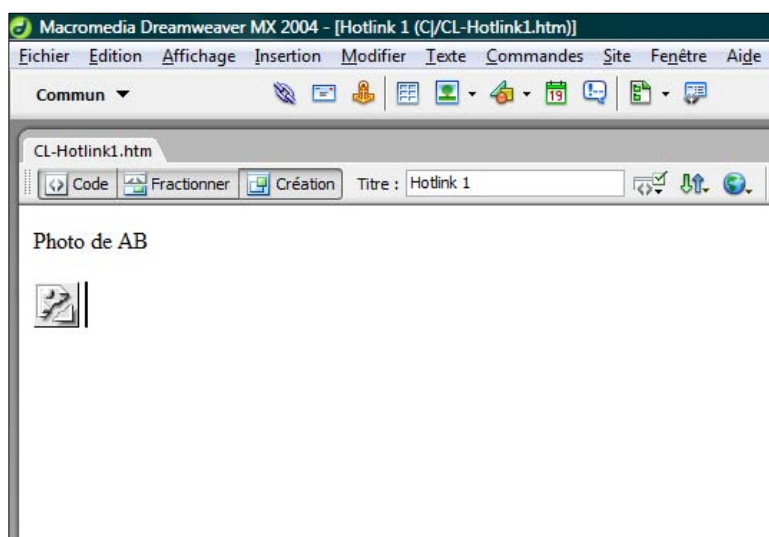
<p>Photo de AB</p>

<p></p>

</body>

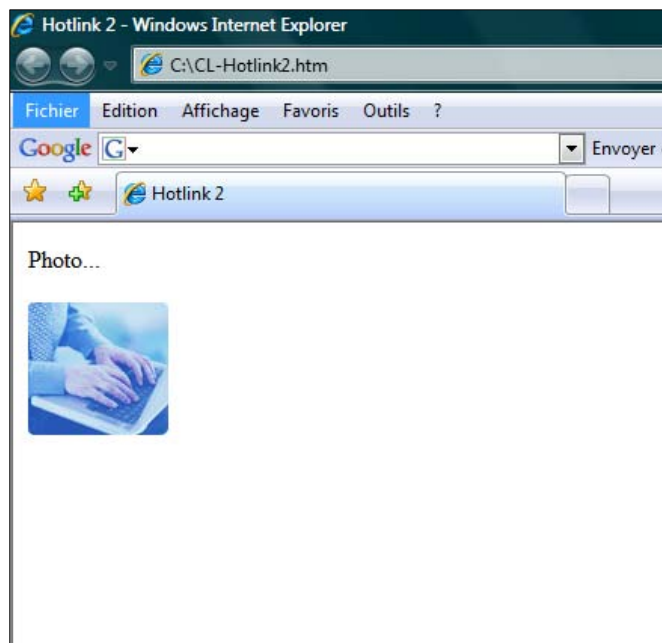
</html>
```

Voici alors une copie partielle de la page engendrée par le code précédent dans un éditeur HTML.



L'icône qu'on y aperçoit indique un lien brisé. Mais voici maintenant ce qu'on obtient en affichant cette page dans un navigateur : celui-ci va requérir l'image dont l'URL lui est indiquée dans la balise `` (voir ci-dessus). Attention ! L'adresse véritablement indiquée n'est pas ici celle de la photo du blogueur dont la mésaventure a été rappelée : pour demeurer

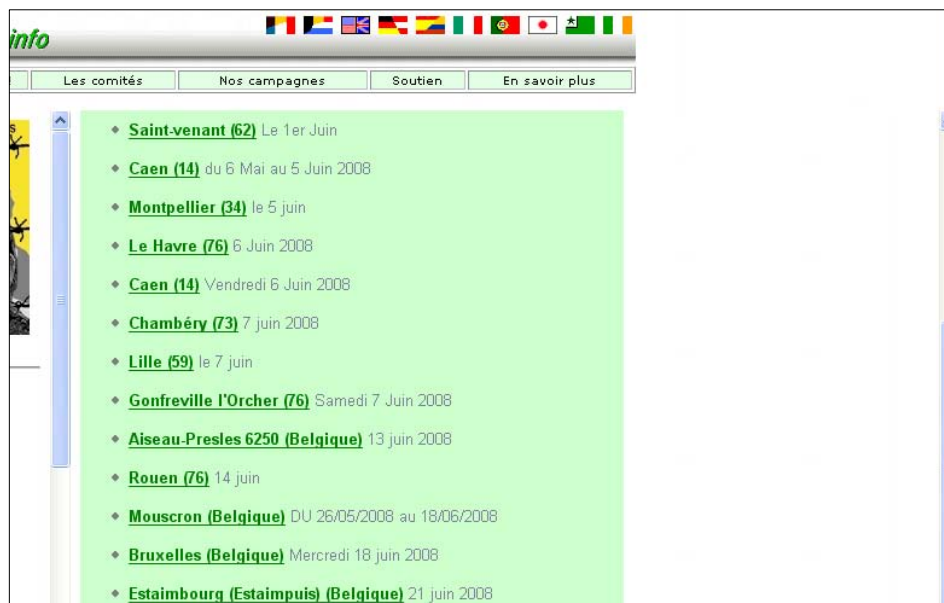
dans le périmètre de la légalité, on a choisi l'adresse d'une photo de notre site personnel. On notera en outre que, pour la cohérence, nous avons remplacé dans le code la légende « Photo de AB » par la légende « Photo... ».



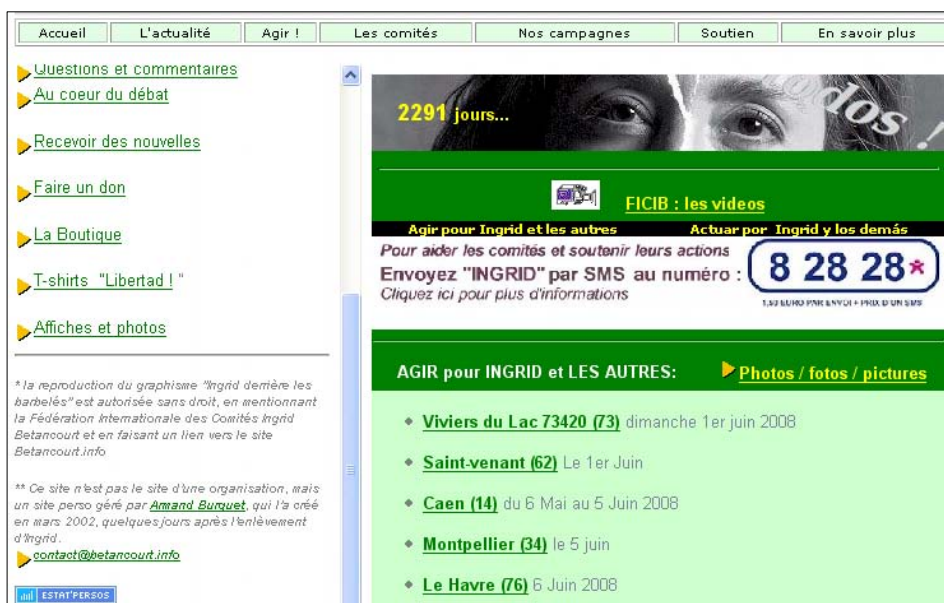
Les pages de cadres comme problème

Une pratique actuellement plus ou moins abandonnée mais dont les effets sont durablement inscrits dans le monde du Web est celle des cadres – en anglais, des *frames*. On a reproduit ci-après une page Web comportant plusieurs cadres, chaque cadre constituant un document HTML différent, qui s'affiche dans une même page à l'écran : sur la première copie d'écran on voit le cadre central en position haute ; sur la seconde copie d'écran on a utilisé la barre de défilement (*scroll bar*, à droite) pour descendre dans ce document central.





Comme on le devine ci-après, et ainsi que le précise l'article en anglais "Framing (World Wide Web)" – que nous suivrons sur le sujet des *frames* –, "headers and sidebar menus do not move when the content frame is scrolled up and down". À l'origine des *frames* se trouve le navigateur Netscape 2.0, dont l'innovation a été reprise par le W3C dans les standards du HTML 3.0. Quel est l'intérêt des pages de cadres ? Sur le site Web auquel se rapportent les copies d'écran précédentes, imaginons que le menu figurant à gauche (voir aussi la copie d'écran ci-après) apparaisse sur toutes les pages du site – lequel pourrait en comporter par exemple une centaine.



Si ce menu (comme c'est le cas ici) est affiché dans une page de cadres, c'est cette même page de cadre qui s'affiche *sur l'ensemble du site*. Grâce à cela, si l'on veut modifier ce menu (en y introduisant par exemple une nouvelle rubrique), il suffit de modifier la page de cadres contenant le menu : on réalise ainsi *une seule* modification au lieu, par exemple, d'une centaine. C'est là bien sûr un avantage évident pour les développeurs de sites.

À l'inverse, une difficulté – du point de vue de l'internaute – engendrée par l'utilisation de pages de cadres est qu'on ne connaît pas l'URL d'une telle page apparaissant dans une page Web : il s'agit en fait de la *même* URL qui s'affiche dans la barre d'adresse, alors que le contenu affiché *n'est pas le même*. (Le lecteur pourra le vérifier plus loin sur un exemple, les copies d'écran du site *Diamonds*.) L'utilisation de l'URL figurant dans la barre d'adresse conduira à l'affichage correspondant à la première copie d'écran, et non à celui de la seconde, dont l'affichage ne pourra être obtenu qu'en suivant un certain chemin à partir de la première ; il résulte de là que l'indication d'une URL est parfois trompeuse pour l'utilisateur. Il va de soi, croyons nous, que la connaissance de ce type de phénomène participe d'un savoir utile pour l'internaute, en particulier en matière de RII.

Avant de revenir sur ce type de question, nous approfondirons un peu le mécanisme examiné ci-dessus en analysant la syntaxe permettant l'utilisation de cadres. Pour créer un cadre, on remplace la balise `<body>` par une autre balise : *frameset*. L'exemple proposé par l'article *framing* est ainsi le suivant.

```
<frameset cols="85%, 15%">
  <frame src="URL OF FRAME PAGE 1 HERE">
  <frame src="URL OF FRAME PAGE 2 HERE">
</frameset>
  Text to be displayed in browsers that do not support frames
</frameset>
```

Comme on le voit, la balise ouvrante, `<frameset>`, comporte l'indication de deux colonnes : la première doit occuper 85 % de la largeur de la page Web, la seconde occupant le reste, soit 15 % (cette seconde indication est parfois remplacée par un astérisque). Dans les deux lignes suivantes sont indiqués alors ce qui va venir s'afficher dans la première colonne et dans la deuxième colonne : il faut ici connaître l'URL de la page de cadre (*framepage*) que l'on souhaite afficher. L'élément `<noframes>` est destiné aux navigateurs « qui ne supportent pas les cadres » : en ce cas, il s'affiche dans le cadre le texte prévu à cet endroit

dans le code source. Le contenu des cadres peut être hébergé sur le serveur qui héberge la page mère (*parent page*), c'est-à-dire la page organisée par le code que nous avons vu plus haut ; mais il peut aussi se trouver sur un *autre* serveur Web d'où il est téléchargé automatiquement pour être affiché dans le cadre : on parle alors de *remote loading* ou *transclusion*. Cette pratique n'est pas sans rappeler le *hotlinking* examiné plus haut. L'un de ses inconvénients pour l'internaute est que celui-ci peut croire que le contenu du cadre appartient au site qu'il visite alors que ce contenu vient d'ailleurs. Un certain nombre de sites empêchent la manœuvre en mettant en place un script à cet effet, qu'on appelle communément un *framekiller*.

Des critiques nombreuses ont été adressées à l'usage des cadres, ce qui a conduit à sa marginalisation progressive. La première objection a déjà été évoquée : l'emploi de cadres rompt le lien entre un contenu et une URL, en sorte qu'il n'est pas simple d'avoir parmi ses favoris un signet renvoyant au contenu d'un cadre. Une deuxième critique repose sur le constat que, selon le navigateur utilisé, le rendu des cadres peut varier parfois fortement ⁸⁵. Une troisième critique tient à ce que l'utilisateur novice ignore en général si le contenu d'un cadre appartient réellement au site sur lequel il se trouve, ou bien provient d'un autre site éventuellement situé sur un autre serveur : dans tous les cas, cela peut entraîner des difficultés légales liées notamment aux droits d'auteur. Une autre critique souligne qu'un résultat fourni par un moteur de recherche peut conduire à une page qui, normalement, devrait s'afficher dans un cadre, ce qui n'est alors pas le cas : la page sur laquelle on arrive en cliquant sur le lien affiché par le moteur de recherche ne possède pas de lien sortant ; elle est alors appelée pour cela, en anglais, *a blackhole page* – en français, on parle généralement de « page orpheline », qui a en quelque sorte perdu sa « page mère » (*parent page*). Autre critique encore : les cadres modifient le fonctionnement du bouton « page précédente » (*back button*), qui parfois semble ne plus fonctionner. Dernière critique enfin que nous retiendrons ici : lorsqu'on veut imprimer la page, il arrive fréquemment que les cadres soient rendus autrement que l'utilisateur ne les attendait en se référant à ce qui s'affichait à l'écran.

Parmi les inconvénients précédents, certains tiennent à la programmation des pages de cadres, d'autres au navigateur utilisé : ils s'effacent à mesure que progressent les pratiques et les outils employés. Qu'est-ce que l'utilisateur ordinaire doit savoir de tout cela ? S'adressant à un auteur d'un site Web à propos du mauvais affichage d'une page de cadres par un navigateur, un commentateur, Dave Owen, écrit : "If you think a 'click here to restore frameset' will solve this problem, you're wrong. 95% of your users will ignore such a message (most of them won't even know what it means)." La première exigence semble donc être que l'utilisateur sache ce qu'est une page de cadres (*frameset*). Le même utilisateur aura

également avantage à vérifier si le contenu qui s’affiche dans un cadre n’appartient pas à un autre site que celui qu’il visite. Pour cela, il convient d’abord qu’il subodore la présence de cadres. Il peut alors procéder à une vérification en effectuant un clic droit sur la zone « suspectée » de correspondre à un cadre. En supposant qu’on utilise le navigateur Internet Explorer, on voit alors apparaître un menu contextuel dont l’entrée « Propriétés » conduit à l’affichage de l’URL de la page de cadres. Si le nom de domaine est différent de celui qui apparaît dans la barre d’adresse, on peut conclure que le contenu du cadre provient d’un autre site, avec les problèmes éventuels y afférents.

Une situation voisine de celle évoquée précédemment a lieu, pour l’internaute, lorsque la page Web affichée à l’écran comporte ce qu’on nomme en anglais des *iframes*, contraction de *inline frames*, et qu’on appelle pour cela quelquefois, en français, des « cadres en ligne ». Considérons ces lignes de code HTML.

```
<iframe src="example.com" width="400" height="500"
      scrolling="auto" frameborder="1"></iframe>
```

Ces lignes de code vont placer un *iframe* encadré d’une bordure (ce qui correspond à l’indication `frameborder="1"` au milieu de la page par défaut), avec une largeur (*width*) et une hauteur (*height*) déterminées. On notera que, ici, l’internaute pourra faire défiler le contenu affiché dans le cadre à l’aide d’une barre de défilement, du moins si ce contenu ne tient pas tout entier dans le cadre : telle est la signification de `scrolling="auto"` (l’absence d’ascenseur, elle, s’exprimerait par `scrolling="no"`). L’utilisation d’un *iframe* permet notamment, du point de vue de la maintenance d’une page Web, de mettre à jour le contenu du cadre en ligne de manière indépendante du reste du contenu de la page. Pour l’internaute, un problème déjà évoqué à propos des *frames* peut se poser : comme on l’imagine à la lecture du fragment de code ci-dessus, l’URL de la page Web affichée à l’écran ne fournit pas l’URL de la page dont le contenu est inséré dans le *iframe*. Si l’on veut alors se référer à ce contenu de façon indépendante, on peut examiner le code source de la page Web, y repérer la balise *iframe* pertinente, et identifier, dans le code, l’URL – ou l’adresse *relative* de la page si celle-ci appartient au même site.

Une URL absolue indique l’adresse complète de la ressource : on l’utilise obligatoirement dans un lien pointant vers un objet situé sur un autre site (lien absolu). L’article « Le langage HTML/Liens » de Wikibooks (en français) propose l’exemple suivant.

```
<a href="http://fr.wikipedia.org">Une encyclopédie vraiment géniale</a>
```

En revanche, une URL relative ne contient ni protocole (http://), ni nom de domaine (fr.wikipedia): un lien relatif indique l'adresse de la ressource par rapport à la page actuelle. Si le fichier appartient au même répertoire que la page courante, on indique le nom du fichier avec son extension, comme ci-après.

```
<a href="exemple.html">Lien vers le fichier « exemple.html » se trouvant dans le même répertoire que la page actuelle</a>
```

Si le fichier appartient à un sous-répertoire du répertoire contenant la page courante, il convient d'indiquer le nom du répertoire, que l'on fait suivre alors d'une barre oblique / et du nom du fichier, comme ci-après.

```
<a href="repertoire_exemple/exemple.html">Lien vers le fichier « exemple.html » se trouvant dans le répertoire « repertoire_exemple » qui se trouve dans le même répertoire que la page actuelle</a>
```

Si le fichier se trouve dans le répertoire parent du répertoire actuel, ou dans un sous-répertoire du répertoire parent, il convient de « remonter » dans l'arborescence à l'aide de deux points (voir ci-après).

```
<a href="../exemple.html">Lien vers le fichier « exemple.html » se trouvant dans le répertoire parent du répertoire dans lequel se trouve la page actuelle</a>
```

L'usage des URL relatives permet de copier l'ensemble d'un site Web sur un autre serveur sans changer les URL. En revanche, si la page actuelle change de répertoire sans que l'objet cible d'un lien sortant de cette page « suive », ce lien est brisé : l'adresse relative devient erronée.

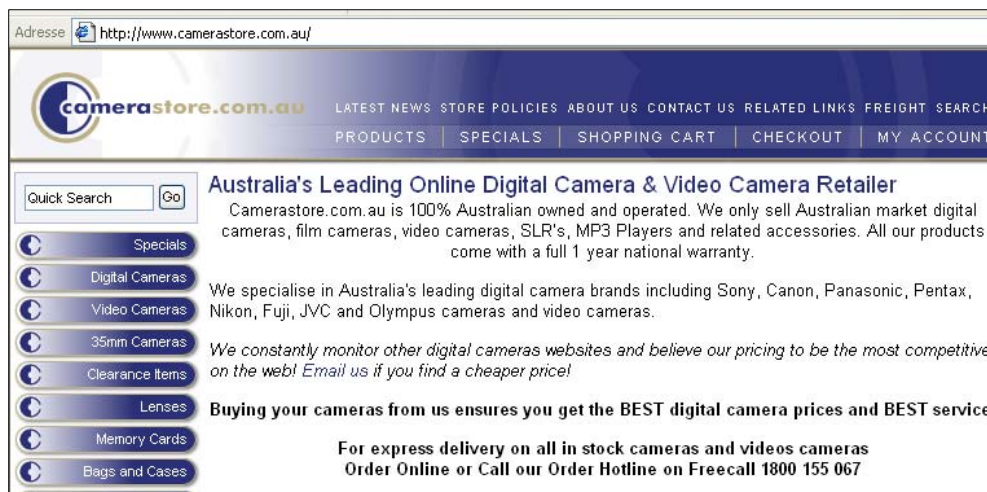
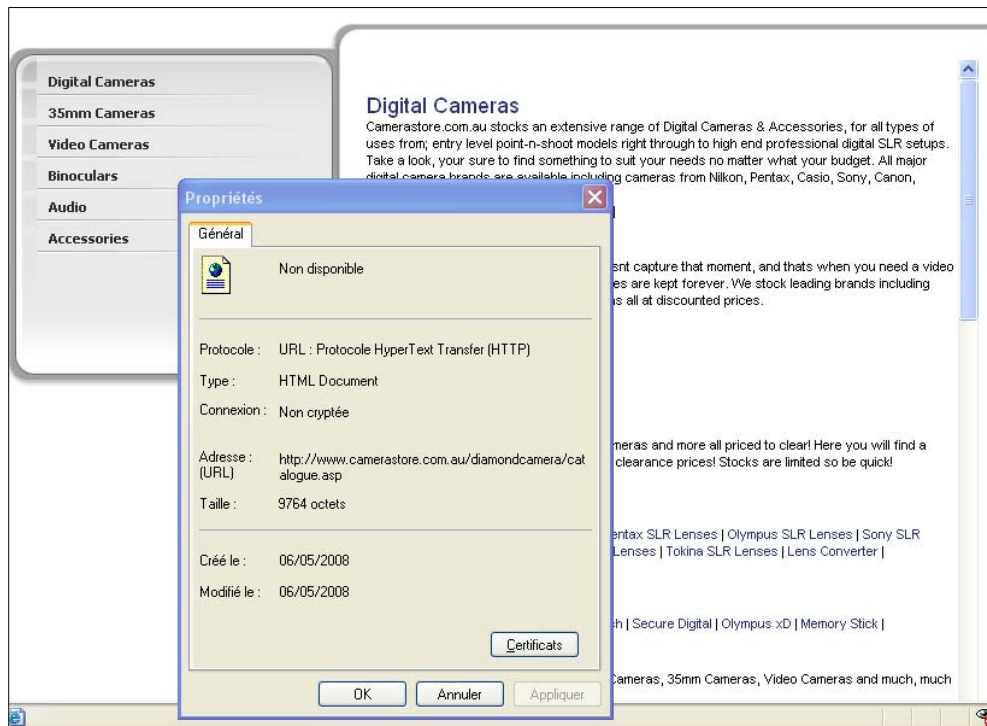
Les copies d'écran ci-après illustrent la manœuvre précisée ci-dessus : sans être véritablement complexe, celle-ci suppose une certaine culture de la part de l'internaute. À l'écran (ci-après) se trouve affichée la page d'accueil d'un site australien, *Diamonds*, dont l'URL est <http://diamondscamera.com.au/httpdocs/index.php>. Lorsque, dans le cadre figurant à gauche, on clique sur "Accessories", on voit apparaître, dans le cadre de droite, un contenu que l'on peut parcourir à l'aide de la barre de défilement apparue à droite.



Par une manœuvre déjà explicitée, on fait apparaître alors ce que montre la première copie d'écran ci-après : l'URL de la page où se trouvent les contenus affichés dans le cadre central est, cette fois, <http://www.camerastore.com.au/diamondcamera/catalogue.asp>. Ces contenus proviennent donc d'un autre site, dont on a reproduit dans la seconde copie d'écran ci-après une partie de la page d'accueil. De son côté, l'examen du code source de la page <http://diamondscamera.com.au/httpdocs/index.php> révèle les lignes ci-après, que nous abandonnons à la sagacité du lecteur.

```
<iframe id="prodframe" name="prodframe"
src="http://www.camerastore.com.au/diamondcamera/productlist.asp"
```

```
scrolling="auto" marginwidth="0" marginheight="0" frameborder="0"
vspace="0" hspace="0" height="550px" style="overflow:visible;
width:100%; "></iframe>
```



Quels contenus ? Quels pièges ?

Le repérage que l'on a opéré jusqu'ici s'agissant des praxéologies du Web n'a encore cerné qu'une partie des questions dont l'utilisateur pourrait avoir avantage à être instruit : les précisions apportées ci-dessus, *après coup*, à propos des notions d'adresse absolue et d'adresse relative montrent, sur un point au reste assez classique, qu'aucun « chapitre » n'est jamais clos

définitivement. Si l'on s'en tient au grand problème de la programmation de pages Web par exemple, ce que nous avons recensé jusqu'ici ne constitue au mieux qu'une « base d'orientation » praxéologique. Or, sans jamais supposer que l'utilisateur adopte la position d'un *Web designer*, mais qu'il doit simplement connaître quelque peu les coulisses de la programmation Web pour en déjouer les pièges et en exploiter les particularités dans ses recherches d'information, il conviendrait de s'interroger sur la masse des outils de programmation aujourd'hui disponible (et qui ne cesse de s'accroître, dans le même temps où certains de ces outils tombent en désuétude). Le HTML de base s'est ainsi enrichi dans sa version 4 des *Cascading Style Sheets* (CSS), les « feuilles de style en cascade », qui permettent une présentation uniforme sur tout un site et autorisent donc des changements qui valent également pour toutes les pages d'un site, lesquelles deviennent aussi, toutes choses égales par ailleurs, davantage lisibles, en même temps que leur chargement se fait plus rapide. Du point de vue qui nous intéresse plus particulièrement ici, la séparation des indications de mise en page et de présentation, d'un côté, et du texte, de l'autre, en allégeant dans le code source ce qui relève des premières, dégage le texte et offre ainsi à l'indexation par les moteurs de recherche des lignes de code moins hérissées de balises de toutes sortes. Sur un patron voisin, le langage de balisage XML défini par le W3C et destiné à remplacer le HTML, peut servir de base à une page Web, tandis qu'un fichier XSL (*eXtended Stylesheet Language*), c'est-à-dire une feuille de style XML, mettra en forme les données du fichier XML.

L'un des problèmes clés posés par le HTML est que les pages créées sur cette base sont totalement statiques, sans possibilité d'interaction avec le visiteur : seules deux balises HTML (<marquee> et <blink>) et quelques fonctionnalités des CSS donnent un peu de dynamisme à la page (en permettant par exemple des changements de couleur lors d'un survol avec la souris). Pour rendre une page Web plus dynamique, une possibilité est d'y insérer des « objets », au premier rang desquels il faut citer les *images* contenues dans des fichiers au format standard pour le Web (.gif, .jpg ou .jpeg et .png). Par contraste les fichiers contenant des documents *multimédia* de type son ou vidéo ne sont pas en général lus par les navigateurs : leur lecture supposera l'installation d'un *plugin* adéquat. Au-delà, on doit mentionner l'utilisation possible (et aujourd'hui fréquente) d'une *animation Flash* contenue dans un fichier ayant l'extension .swf (acronyme de *Shockwave Flash* qui se prononce en anglais *swif*) : une page Web peut contenir une animation Flash (par exemple pour dérouler un menu) ou encore peut être entièrement en Flash, mais toujours avec un squelette HTML. Le programme Flash (créé par Macromedia, société rachetée par Adobe) permet de créer des animations de qualité ayant en outre une certaine interactivité grâce au langage de

programmation *ActionScript*, langage de script basé sur ECMAScript, version standardisée de *JavaScript*.

Du point de vue qui est le nôtre ici, nous parvenons, avec l'emploi du programme Flash, à une limite : la création d'une animation Flash est souvent, en effet, une tâche d'une grande complexité. Si donc on laisse de côté l'idée de maîtriser la programmation d'une animation Flash, mais en nous situant toujours du côté de la RII, il n'en reste pas moins qu'il peut être utile à l'internaute voulant utiliser toute l'information du Web de pouvoir *se référer* à une animation Flash et, dans certains cas, de pouvoir *recupérer* certains de ses contenus (texte, image, son), ce qui n'est pas immédiat ni même toujours possible ⁸⁶.

Parmi les enrichissements possibles d'une page Web « standard », on trouve encore les *applets Java* appelés parfois en français, les *appliquettes* : il s'agit de petits programmes comportant une interface programmée à l'aide du langage Java, et complètement autonome (quoique éventuellement paramétrable). Intégrée dans une page Web, ils sont téléchargés avec la page et exécutés par l'ordinateur client. Les *ActiveX*, créés par Microsoft pour les plates-formes Windows, ont des principes de fonctionnement très semblables à ceux des applets. L'inventaire ainsi ébauché pourrait être encore longtemps poursuivi. À titre de repère, nous nous arrêterons un instant sur les problèmes que peuvent poser à l'internaute l'utilisation, par les créateurs de sites, de certains *scripts*.

Comme on le verra dans ce qui suit, nous nous référons ici au langage JavaScript. Apparue en 1995 avec le navigateur Netscape 2, développé par Netscape et Sun Microsystems, ce langage avait pour but de permettre d'intégrer des effets graphiques et d'autres fonctionnalités dans les pages Web : apparition de textes ou d'images quand la souris passe sur certaines zones (mouseovers ou rollovers), vérification de la validité d'une saisie, calcul de totaux dans des formulaires, etc. JavaScript est un « petit » langage de programmation, relativement simple, qui est interprété directement par le navigateur et reste visible (et copiable) dans le code source d'une page HTML. Les notions clés en sont la notion d'objet – on y manipule par exemple l'objet window (fenêtre), l'objet document, l'objet form (formulaire), etc. –, les notions de fonction et de méthode – les méthodes sont des fonctions prédéfinies, telles alert() ou confirm() – et les événements – tels onClick (au clic de la souris sur un élément HTML), onLoad (au chargement de la page par le navigateur), ou onMouseOver (au survol d'un élément HTML par la souris). Pour plus d'information, on pourra se reporter par exemple à l'ouvrage de Luc Van Lancker (2008).

Ainsi n'est-il pas rare, que, une certaine page étant affichée par le navigateur, le bouton de retour en arrière, supposé activé, ne réagisse pourtant pas lorsqu'on clique dessus : le retour en arrière – à la page précédente – semble impossible ! Très généralement, la chose a été voulue par le *Web designer*, parfois pour des raisons tout à fait justifiées – si chaque page présente quelques questions d'un questionnaire qui se déploie sur une suite de pages, il s'agira d'interdire à l'enquêté de revenir en arrière pour modifier ses réponses en tenant compte des questions venant *après*. Bien souvent, pourtant, il semble que la raison soit autre : il s'agit par exemple de faire que l'internaute reste sur la page ou suive l'un des liens qui lui sont proposés. Cette « interdiction » est réalisée à travers un script tel le suivant.

```
<html>
<head>
<title>Page A</title>
<script language='javascript'>
window.history.forward(1);
</script> </head>
```

Expliquons d'abord le fragment de code ci-dessus, ou plus exactement le sous-fragment contenu entre les balises script ouvrante et fermante :

```
<script language='javascript'> window.history.forward(1); </script>
```

La mention `language='javascript'` indique au navigateur que ce qui suit est écrit en JavaScript. Ce qui suit, en l'espèce, se réfère à l'objet `window` et à « l'objet fils » nommé `history` (« historique »), lequel « permet de naviguer dans l'historique des pages consultées, en arrière et en avant » (Hondermarck, 2007, p. 226). Ici intervient la « méthode » `forward()`, avec un « détournement » opportuniste qui consiste formellement à introduire le chiffre 1 entre les parenthèses : `forward(1)`. Lorsqu'il figure sur une page A, sur laquelle on peut suivre un lien vers une page B, ce petit script va interdire de revenir de B en A en utilisant le bouton de retour (*back button*) : lorsque, en effet, on clique sur le bouton de retour, le navigateur « revient » sur le code de la page A qu'il interprète normalement, code qui *lui indique alors d'avancer d'une page dans l'historique* – tel est le sens de `forward(1)` –, en sorte qu'on « reste » sur la page B. (Bien entendu, on peut toujours revenir à la page A en introduisant son URL – si on la connaît – dans la barre d'adresse du navigateur, ou encore en suivant un lien qui irait de la page B à la page A.)

Ce qu'il faut surtout souligner alors – dans la perspective que nous explorerons dans la quatrième partie de notre travail –, c'est que le script présenté ci-dessus est d'un accès malaisé dans l'offre praxéologique en matière de scripts : pour notre part, nous l'avons emprunté à un forum de discussion qui a le mérite de montrer les débats que suscite ce type de pratiques, comme on le verra ci-après ⁸⁷.

Bharadwajrv

May 14th, 2002, 06:10 AM

Hi,

can u tell me how to disable back button.. (i.e., user should not be able to view the previous page from back button...but they can go through hyper link provided in the web page...)

thanks in adv.

Venu

bharadwajrv@yahoo.com

"Never look back, b'caz u will loose things in front"

chavansu

May 17th, 2002, 04:52 AM

Page A calls Page B

User should not be allowed to go to Page A from Page B.

Put the code below in Page A.

<Script>

window.history.forward(1);

</Script>

Sam Hobbs

May 18th, 2002, 10:52 AM

Oooooooooohhhhhhhhhhh!!!!!!!!!!!!!!

I really, really, really, really, really do not like web pages that do not allow me to go back. If I can't go back then I will close that browser window as soon as I can. Disabling the back button is probably a symptom of bad design.

proxima centaur

May 18th, 2002, 11:15 AM

I agree.

Coercitive programming should be outlawed! Behaviour of software you own should not be changed by scripts or gizmos!

:p

Kaytrim

May 21st, 2002, 04:24 PM

Situation: you are accessing a page that uses a form to post information to a database.

If you use the browser back button the information that you entered on the form will be lost. This is why us programmers want to be able to disable the back button on the browser.

That way we don't get complaints from users saying that they entered the data but it 'disappeared'.:confused:

proxima centaur

May 21st, 2002, 04:39 PM

But it is annoying when you want to go back to two or three pages before this one when the browser won't let you.

Don't solve a problem by creating another ;-)

Sam Hobbs

May 21st, 2002, 05:31 PM

The problem of the data being lost happens to messages in this forum also. It does not happen everytime; only about 20% of the time. It happens using this software and it happened in the previous CodeGuru forum software. It happens in many other forms, also; probably all forms. It is a problem I would solve in a more elegant manner if I had the time. I agree with proxima centaur; disabling the back button is not elegant.

I think this is an excellent example of a bad choice for solving a problem. Since this is common behavior (can be expected) you can simply explain to your users that it is a problem with the browser and they should expect it to happen with most forms.

I am not really experienced with cookies but it seems to me it would be easy enough to save the form data in cookies and then retrieve it if it exists. Isn't there a type of cookie that is temporary for the session or something such as that?

When creating messages for this forum, I am in the habit of doing a Ctrl-A ("Select All") then a Ctrl-Inst (Copy) before clicking on the preview button because sometimes the data is gone when I browse back to it. It is a problem I wish they would fix for this software too, but disabling the back button

would make the preview button useless. Well, actually, I guess this software makes it unnecessary to go back after a preview but I wish it would be fixed so that the data is not lost.

proxima centaur

May 22nd, 2002, 09:03 AM

Informing the user that this page is not available anymore is enough for me. It is to the point and is not intrusive. If you absolutely do not want the user to see the page, and I don't know if this is possible at all, why don't you redirect the browser go to a valid page?

just my two cents again ;)

ts_kar

February 21st, 2005, 04:49 AM

[QUOTE=bharadwajrv]Hi,

can u tell me how to disable back button.. (i.e., user should not be able to view the previous page from back button...but they can go through hyper link provided in the web page...)

thanks in adv.

Venu

Can you please tell me the result in mail..

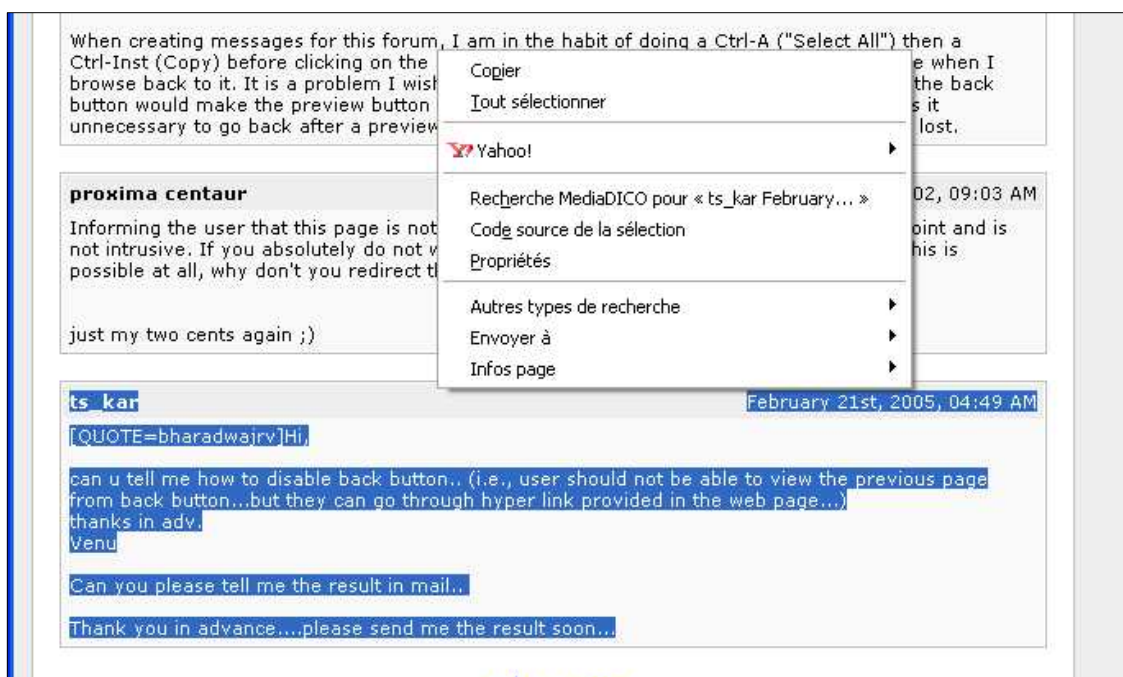
Thank you in advance....please send me the result soon...

Voici maintenant un script qui empêche de faire un clic droit sur la page (Hondermarck, 2007, p. 277).

```
<script type="text/javascript">
function noClicDroit()
{
    alert('Clic droit interdit');
    return(false);
}
document.oncontextmenu = noClicDroit;
</script>
```

La première ligne indique au navigateur qu'il s'agit d'un script écrit en JavaScript. La deuxième ligne déclare une fonction, à savoir la méthode `noClicDroit()`. Le code entre

accolades (qui est techniquement facultatif, mais constitue une marque de politesse à l'endroit de l'utilisateur) indique au navigateur d'afficher un message d'alerte dont le contenu est tout simplement « Clic droit interdit » chaque fois que l'internaute tente un clic droit avec sa souris. L'avant-dernière ligne a trait à l'événement consistant à interdire d'accéder au menu contextuel qui apparaît normalement lorsqu'on fait un clic droit sur une zone cliquable. La copie d'écran de la page ci-après donne un exemple de ce que cela *empêche* l'utilisateur d'obtenir. On voit ainsi, par exemple, que si l'on souhaitait *copier* le passage sélectionné, cela n'est maintenant plus possible, puisque le menu contextuel ne s'affiche plus – ce qui s'affiche étant le message d'alerte « Clic droit interdit ».



Même quand il se veut innocent, l'emploi de scripts peut avoir des effets gênants pour l'utilisateur. Un auteur déjà mentionné, présentant ce que permettent de faire les scripts, propose à titre illustratif la liste d'actions suivantes (Van Lancker, 2008, p. 7).

- animer du texte ou des images,
- réagir à l'action de la souris,
- détecter le type et la version du navigateur,
- vérifier la saisie dans les formulaires,
- effectuer des calculs simples,
- demander une confirmation,
- afficher la date et l'heure,
- gérer les menus de navigation,

- rediriger le visiteur vers une autre page
- etc.

Toutes ces possibilités semblent *a priori* bénéfiques. L’antépénultième item de la liste, « gérer les menus de navigation », correspond ainsi à ce fait connu de tout internaute : l’apparition d’un menu déroulant lorsque la souris survole un lien de la barre de navigation, ce dont on trouvera une illustration ci-après.



Nous nous arrêterons plus longuement sur l’action suivante : « rediriger le visiteur vers une autre page ». Pour saisir le problème que cela peut poser à l’internaute, examinons le script ci-après⁸⁸, qui nous est présenté en ces termes (nous avons conservé l’à-peu-près du français de son auteur) : « Ce script redirige l’internaute vers un page de sa langue. Il fonctionne avec tout les navigateurs et vous pouvez mettre les langues que vous voulez et le nombre que vous voulez. Si une langue n’est pas dans la liste de choix le navigateur prend la dernière (là index_UK.htm). »

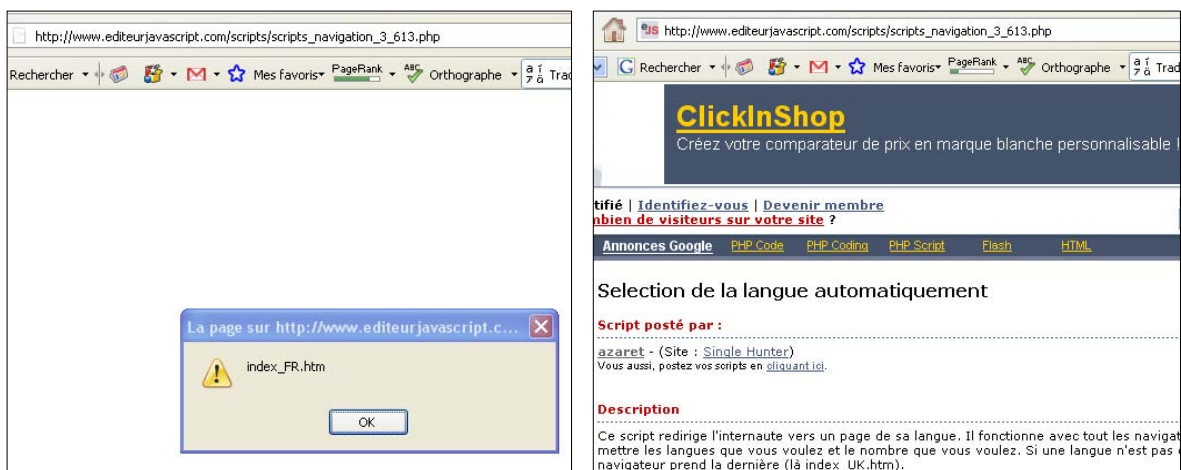
```
<SCRIPT LANGUAGE="JavaScript" TYPE="text/javascript"><!--
//
if (navigator.appName == 'Netscape')
    var language = navigator.language;
else
    var language = navigator.browserLanguage;
var code = language.substring(0,2);
if ((code == 'fr') || (code == 'de') || (code == 'it') || (code ==
'sp'))
```

```

{
if (code == 'fr') url ="index_FR.htm";
if (code == 'de') url ="index_DE.htm";
if (code == 'sp') url ="index_SP.htm";
if (code == 'it') url ="index_IT.htm";
}
else {url ="index_UK.htm";}
// Decommentez cette ligne :
// document.location.href = url
alert(url);
//-->
</SCRIPT>

```

Le script déclare d'abord une variable, `language`, qui a pour valeur la langue du navigateur. Une seconde variable est alors déclarée, `code`, dont la valeur est le code de la langue (`fr`, `de`, `it`, `sp`), extrait de l'objet `language.substring`. Le script contient alors un test conditionnel : si la variable `code` a pour valeur `fr`, alors le navigateur doit afficher la page dont l'URL relative est (ici) `index_FR.htm`, etc. ; sinon, le navigateur affichera la page dont l'URL relative est `index_UK.htm`. Les dernières lignes du code font afficher la boîte de dialogue (alerte) reproduite ci-après : cette boîte bloque l'exécution du script tant que l'utilisateur n'a pas cliqué sur OK pour la refermer.



En règle générale, la boîte de dialogue que le script examiné ici fait apparaître (ci-dessus, à gauche) n'existe pas : l'utilisateur est redirigé vers la version française du site par exemple (ci-dessus à droite) sans que cela lui soit indiqué explicitement, et donc en quelque sorte à son insu. Or l'internaute pourrait vouloir accéder – parce qu'il recherche des informations

linguistiques par exemple – à la version en anglais ou en espagnol dudit site. Dans un certain nombre de cas, il est vrai, le webmaster aura heureusement prévu que l'utilisateur sélectionne lui-même, parmi un choix proposé, la langue qu'il souhaite (en effectuant par exemple un clic sur un petit drapeau correspondant à cette langue). Dans d'autres cas, ce sont d'autres procédures qui devront être appliquées ; plus largement, la frontière entre script « légitime » et script peu honnête peut fluctuer au gré de l'opinion des *Web designers*. C'est ainsi que, précisément, il semble être considéré comme acceptable de rediriger automatiquement un internaute dont le navigateur a été reconnu comme « parlant français » vers une page en version française (ce qui ne peut que conforter les frontières linguistiques existantes...). En revanche, d'aucuns trouveront peu acceptables les scripts empêchant le clic droit. Ainsi en va-t-il d'un certain « SuperDevy », qui, dans un forum, proteste énergiquement contre l'usage déloyal qui pourrait en être fait ⁸⁹.

Je déconseille à quiconque de mettre ce genre de code sur sa page web.

1/ Ça fait "je me la pète avec mes super scripts"

2/ C'est inutile puisqu'il suffit de faire CTRL+C pour copier le code.

3/ Il est toujours possible d'éditer la page pour prendre ce dont on a besoin.

4/ C'est extrêmement désagréable pour le visiteur.

5/ C'est ridicule : on ne mets pas un contenu sur internet pour empêcher les visiteurs de l'utiliser convenablement.

SuperDevy

D'une manière générale, les rédacteurs de scripts tendent souvent, nous semble-t-il, à prendre des décisions à la place de l'internaute qui visite la page : raison pour laquelle le praticien de la RII doit être informé de leur existence, aux conséquences parfois déplaisantes – même si, bien entendu, rien n'oblige personne à utiliser des scripts de cette façon.

Les cookies et la liberté de l'internaute

Nous avons rencontré en passant, dans la première partie de ce travail, puis croisé à nouveau dans la présente partie, la notion de *cookie* – de « témoin de connexion ». On sait que le nom de *cookie* désigne un petit gâteau : le mot appartient à l'anglais américain où il apparaît au début du XVIII^e siècle, inspiré du néerlandais *koekje* « biscuit, petite brioche ». Le *Onelook Dictionary* le définit, au sens où nous le prenons ici, de la façon suivante : "A short line of text that a web site puts on your computer's hard drive when you access the web site."

Précisons cela à l'aide de l'article "HTTP cookie" de Wikipedia (en anglais). Un navigateur demande une page à un serveur Web en lui adressant une requête HTTP qui peut être la suivante (si la page demandée a pour URL <http://www.example.org/index.html>) :

```
GET /index.html HTTP/1.1
Host: www.example.org
```

Le serveur envoie alors une réponse HTTP qui peut contenir, comme ci-après, une demande adressée au navigateur de stocker des cookies :

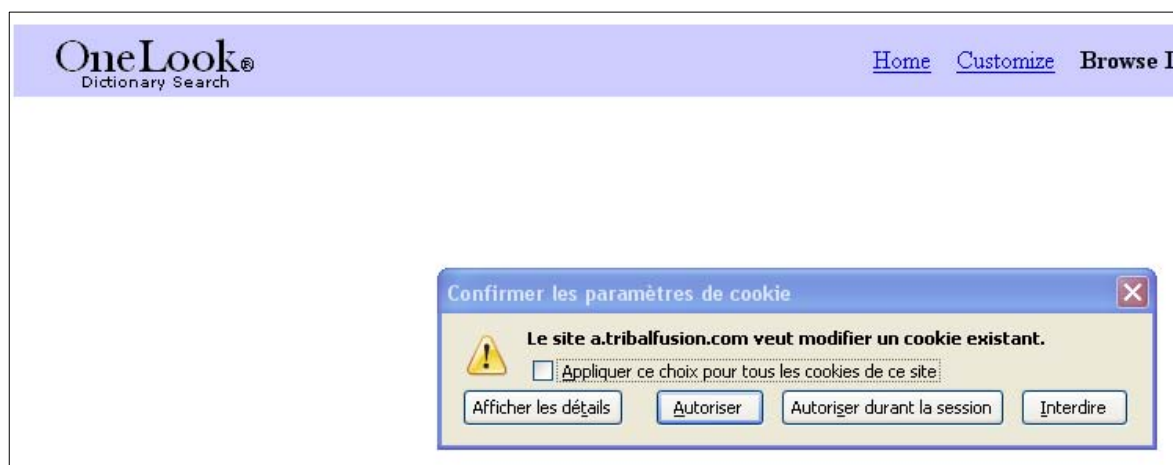
```
HTTP/1.1 200 OK
Content-type: text/html
Set-Cookie: name=value
(contenu de la page)
```

La ligne `Set-cookie` n'est adressée au navigateur que si le serveur souhaite qu'il stocke un cookie. En l'espèce, le navigateur reçoit la demande de stocker la chaîne `name=value` (qui peut être par exemple `temperature=20`). Si le navigateur est autorisé à stocker le cookie, chaque requête HTTP adressée ultérieurement au serveur contiendra le cookie ; si par exemple le navigateur demande ensuite la page <http://www.example.org/spec.html>, il adressera au serveur la requête suivante :

```
GET /spec.html HTTP/1.1
Host: www.example.org
Cookie: name=value
Accept: */*
```

Le serveur renvoie alors la page requise, en demandant éventuellement au navigateur – auquel il adresse pour cela une nouvelle ligne `Set-Cookie: name=newvalue` – de changer la valeur du cookie. Pour qui ne l'a jamais observé, le « ballet » des cookies est impressionnant. En voici un exemple dans lequel, la *toolbar* de Google étant installée, et l'option « me demander chaque fois » de la rubrique *Cookies* (à laquelle on arrive par **Outils ► Options ► Vie privée**) du navigateur Mozilla Firefox étant activée, on appelle la page d'accueil du *OneLook Dictionary* : successivement trois demandes sont d'abord présentées (ci-après) ;


ayant chaque fois choisi l'option « Autoriser durant la session », on voit enfin s'afficher la page d'accueil dudit dictionnaire.



On interroge alors le dictionnaire à propos du mot anglais *profile*⁹⁰. Avant que l'affichage des résultats n'apparaisse, un cookie doit être encore modifié. Puis les résultats apparaissent ; mais aussitôt un autre cookie doit être modifié. On a alors accès aux divers dictionnaires consultés. Supposons qu'on veuille examiner l'entrée *profile* du premier figurant dans la liste, l'*Encarta World English Dictionary* : on clique pour cela sur le lien porté par [profile](#).

OneLook®
Dictionary Search
[Home](#) [Customize](#) [Browse](#)

Confirmer les paramètres de cookie


Le site a.tribalfusion.com veut modifier un cookie existant.

☐ Appliquer ce choix pour tous les cookies de ce site

Afficher les détails
Autoriser
Autoriser durant la session
Interdire

OneLook®
Dictionary Search
[Home](#) [Customize](#) [Browse D](#)



[Regardez les émissions TV en direct sur votre PC](#)

Clic '**Prochain**' pour commencer.

Confirmer les paramètres de cookie


Le site toolbarqueries.google.fr veut modifier un cookie existant.

☐ Appliquer ce choix pour tous les cookies de ce site

Afficher les détails
Autoriser
Autoriser durant la session
Interdire

We found 67 dictionaries that include the word *profile*:

➔ **General** (33 matching dictionaries)

1. [profile](#) : Encarta® World English Dictionary, North American Edition [[home](#), [info](#)]
2. [profile](#) : Compact Oxford English Dictionary [[home](#), [info](#)]

OneLook®
Dictionary Search
[Home](#) [Customize](#) [Browse D](#)



[Regardez les émissions TV en direct sur votre PC](#)

Clic '**Prochain**' pour commencer.

Confirmer les paramètres de cookie


Le site encarta.msn.com veut définir un cookie.

☐ Appliquer ce choix pour tous les cookies de ce site

Afficher les détails
Autoriser
Autoriser durant la session
Interdire

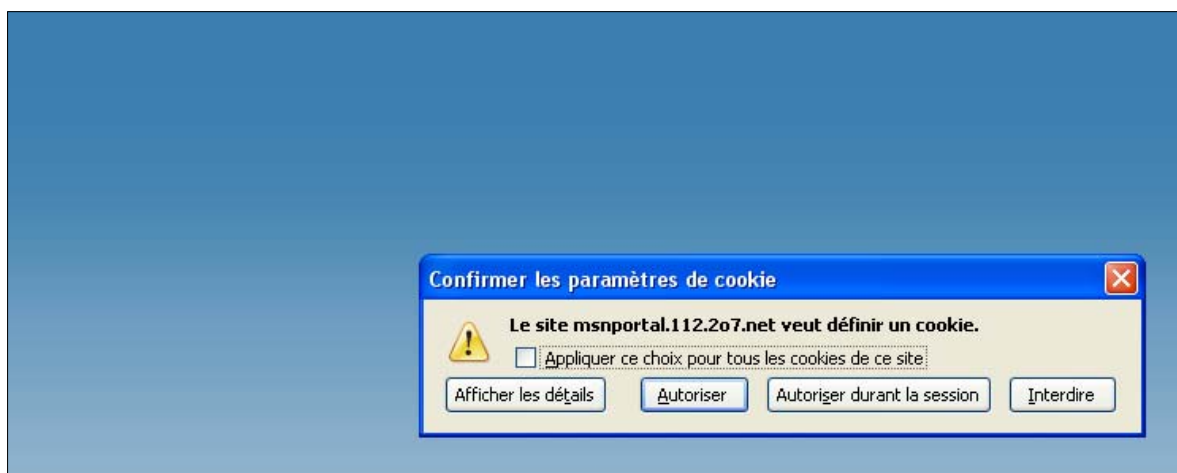
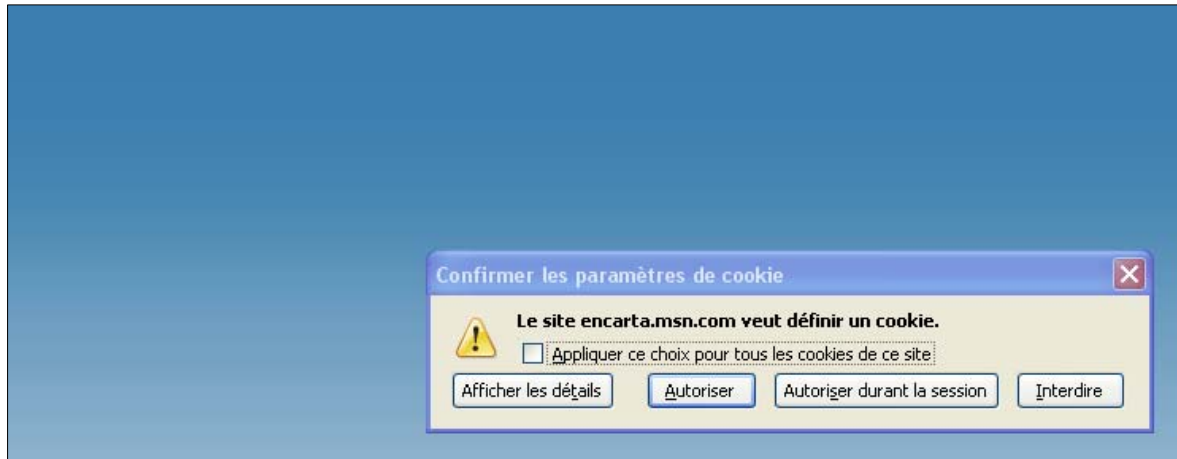
We found 67 dictionaries that include the word *profile*:

➔ **General** (33 matching dictionaries)

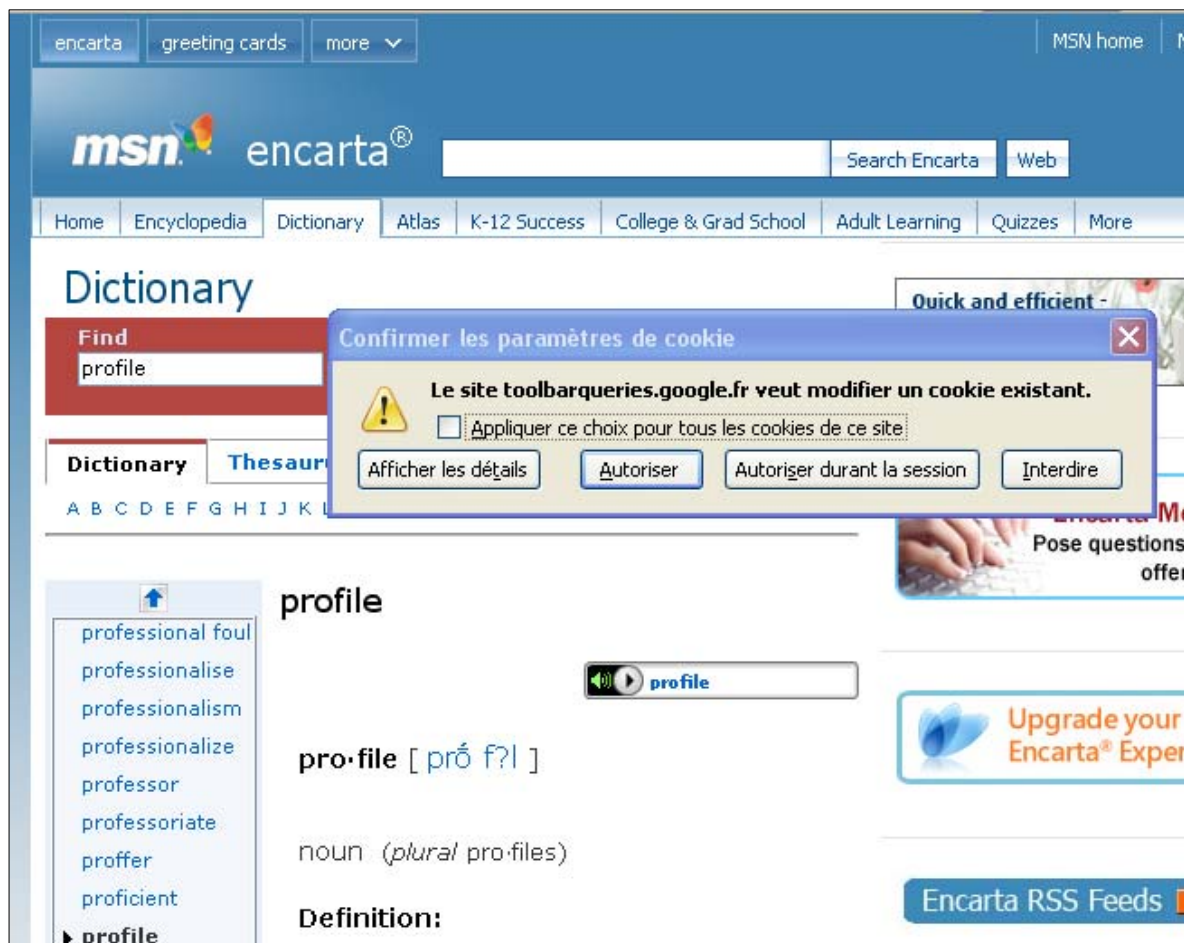
1. [profile](#) : Encarta® World English Dictionary, North American Edition [[home](#), [info](#)]
2. [profile](#) : Compact Oxford English Dictionary [[home](#), [info](#)]

425

Le cookie étant accepté, la page attendue commence à s'afficher, mais une nouvelle demande interrompt l'opération, puis une encore, puis une autre encore.



L'accès à la page donnant le texte attendu est bloqué par une nouvelle et dernière demande, que voici.



Bien entendu, si l'on revient en arrière par le *back button*, on rencontrera des demandes de modification des cookies existants : le premier clic vers l'arrière par exemple conduit à l'affichage ci-après ; etc.



Tout cela rapidement explicité⁹¹, pourquoi donc les serveurs Web envoient-ils des cookies ? Rien n'a été dit jusqu'ici, en effet, de la *fonction* de ces petits fichiers de texte (d'au plus 4 Ko) qu'on trouvera stockés sur le disque dur de son ordinateur (à un endroit qui dépend notamment du navigateur utilisé). L'*Encarta World English Dictionary* précise les fonctions des cookies par cette définition concise : "A computer file containing information about a user that is sent to the central computer with each request. The server uses this information to customize data sent back to the user and to log the user's requests." Le *Merriam-Webster's Online Dictionary* confirme une telle formulation : "A small file or part of a file stored on a World Wide Web user's computer, created and subsequently read by a Web site server, and containing personal information (as a user identification code, customized preferences, or a record of pages visited)." L'article « Cookie (informatique) » de l'encyclopédie Wikipédia fournit de même les précisions suivantes.

Les cookies sont de petits fichiers textes stockés par le navigateur web sur le disque dur du visiteur d'un site web et qui servent (entre autres) à enregistrer des informations sur le visiteur ou encore sur son parcours dans le site. Le webmaster peut ainsi reconnaître les habitudes d'un visiteur et personnaliser la présentation de son site pour chaque visiteur...

On voit que, ce qui est mis en avant ici, c'est notamment la *personnalisation* – la *customization* – de la réponse du serveur (en fonction de l'utilisateur), laquelle est généralement présentée comme une bonne chose pour l'usager. C'est ainsi par exemple que l'article « Cookie (informatique) » déjà cité avance que « les cookies sont largement utilisés pour simplifier la vie des visiteurs et leur présenter des informations plus pertinentes ». L'article "HTTP cookie" retient d'abord trois usages regardés manifestement comme désirables. Le premier est lié à l'invention même des cookies par Lou Mulnelli en juin 1994, alors qu'il travaillait pour Netscape Communications :

HTTP cookies were introduced to provide a way for realizing a "shopping cart" (or "shopping basket"), a virtual device into which the user can "place" items to purchase, so that users can navigate a site where items are shown, adding or removing items from the shopping basket at any time.

Un deuxième usage concerne l'authentification du visiteur d'un site, qu'un cookie bienveillant assure automatiquement après la première visite.

Allowing users to log in to a website is another use of cookies. Users typically log in by inserting their credentials into a login page; cookies allow the server to know that the user is already authenticated, and therefore is allowed to access services or perform operations that are restricted to logged-in users.

La troisième fonction « positive », la personnalisation, a déjà été soulignée.

Many websites also use cookies for personalization based on users' preferences. Sites that require authentication often use this feature, although it is also present on sites not requiring authentication. Personalization includes presentation and functionality. For example, the Wikipedia Web site allows authenticated users to choose the webpage skin they like best; the Google search engine allows users (even non-registered ones) to decide how many search results per page they want to see.

Mais il est surtout un quatrième usage, que la plupart des auteurs s'accordent à regarder comme négatif : les atteintes possibles, et diverses, à la « confidentialité » et à la vie privée.

Cookies are also used to track users across a website. Third-party cookies and Web bugs, explained below, also allow for tracking across multiple sites. Tracking within a site is typically done with the aim of producing usage statistics, while tracking across sites is typically used by advertising companies to produce anonymous user profiles, which are then used to target advertising (deciding which advertising image to show) based on the user profile.

La traque par des « tierces parties », souvent avec une visée publicitaire, est particulièrement condamnée. Pour information, voici le principe de ce « pistage » tel que le présente l'article « Cookie (informatique) » déjà cité.

1. Le visiteur consulte une page web P_1 d'un site web A.
2. La page P_1 contient une image I_1 provenant d'un site web S spécialisé dans le suivi.
3. Le site S enregistre que l'image I_1 sert à la page P_1 du site A.
4. Le site S envoie l'image au visiteur avec un cookie C.
5. Le visiteur consulte une page web P_2 d'un site web B, qui n'a *a priori* rien à voir avec le site A.
6. La page P_2 contient une image I_2 du même site de suivi S.
7. Comme le navigateur du visiteur envoie automatiquement le cookie C avec la requête pour charger l'image I_2 , le site S peut observer que le même visiteur a visité les sites A et B.
8. Le site S peut suivre la trace du visiteur sur tous les sites comportant une de ses images.

9. Si en outre le visiteur s'est identifié auprès d'un des sites A ou B qui a transmis ces informations au site S, alors le suivi n'est plus anonyme.

Dans tous les cas, il s'agit de dresser un « profil » de l'utilisateur, par exemple pour devancer ses demandes futures ou ses attentes supposées. Certains auteurs se dressent toutefois contre les idées parfois alarmantes à propos des cookies en rappelant que ceux-ci se bornent à donner une mémoire, limitée qualitativement et quantitativement, à un système, le protocole HTTP, qui est en lui-même *sans mémoire*, puisque chaque requête HTTP envoyée au serveur est indépendante des requêtes précédentes. Un auteur brosse ainsi le tranquille bilan que voici (Lemainque, 2007, p. 449).

Un cookie peut servir à stocker des informations saisies par l'utilisateur dans un formulaire Web ou tout autre périphérique d'entrée en ligne. Il peut aussi mémoriser des comportements, comme les liens hypertextes suivis ou les images affichées. Cela permet au responsable du site de cibler sa publicité, d'évaluer l'efficacité des différents dispositifs de navigation et d'identifier les sujets qui retiennent l'attention des visiteurs.

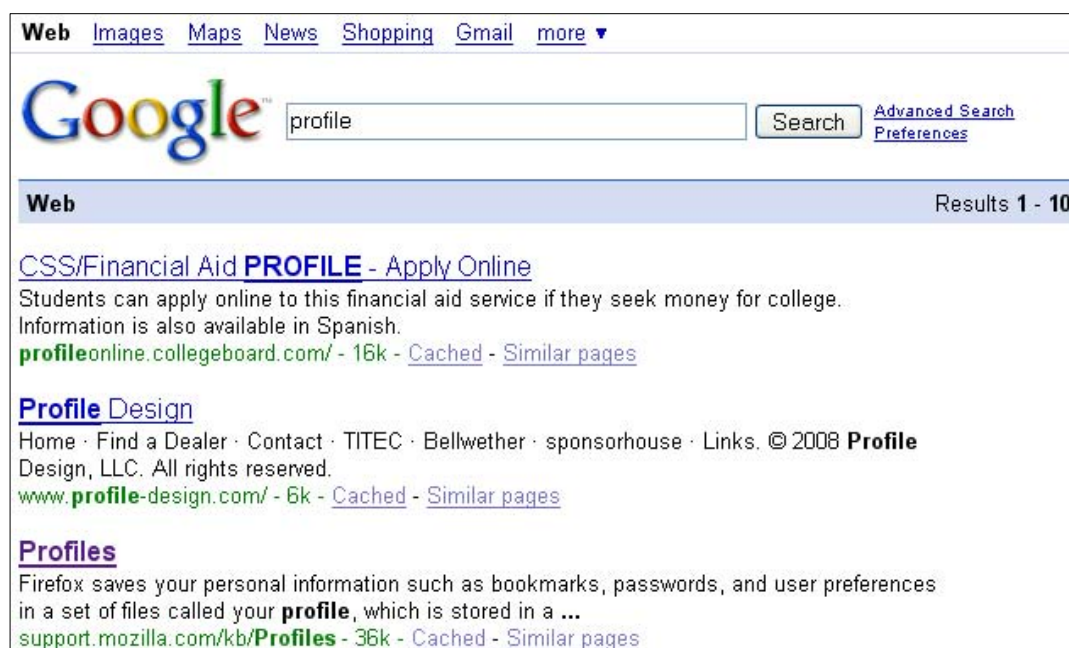
Malgré leur potentiel et en dépit de certaines idées reçues, les cookies restent limités. Ils ne peuvent exécuter un programme situé sur votre disque dur, télécharger ou installer un virus, imposer à votre ordinateur d'envoyer un message électronique à votre insu, lire un élément non volontairement fourni par l'intermédiaire d'un formulaire Web ou, d'une façon plus générale, provoquer une quelconque modification de votre système. Un cookie stocke une information élémentaire dans un endroit précis de votre disque dur [...]. C'est tout.

Les cookies interfèrent-ils cependant avec la RII – qui n'a pas en elle-même à se soucier de questions de « vie privée » ? Il est à cet égard notable que la distinction faite plus haut entre le « pistage » d'un utilisateur à l'intérieur d'un site et sa traque à travers différents sites par une tierce partie, surveillance dont les résultats vont dans tous les cas peser sur ce qui sera présenté ultérieurement à l'internaute, n'a pas, du strict point de vue de la RII, de pertinence *a priori*. La question cruciale, à cet égard, reste la suivante : se peut-il que l'utilisation de cookies aide ou au contraire gêne la recherche d'informations dans le but d'élaborer une réponse *R* à une certaine question *Q* ? Dans quelques cas, la réponse peut raisonnablement être regardée comme positive. Les deux copies d'écran ci-après montrent des pages du site Amazon.com où sont proposés divers ouvrages censés intéresser l'utilisatrice en fonction de son « profil », lié notamment à ses acquisitions antérieures sur ce site.

changes this number to 20 in the preference page, the server modifies the cookie with NR=20. Every time the user queries the search engine, the cookie is sent to the server along with the query. This way, the server knows how many hits should be shown in each page.

À l'inverse, toutefois, la « personnalisation » de la réponse affichée par certains sites (en se fondant sur le « profil » de l'utilisateur) peut *gêner* la RII, et notamment la communication de résultats de recherche sur le Web, en cela que, en fonction de son « profil », un autre internaute, se fiant aux références proposées, pourra pourtant ne pas retrouver à l'identique les résultats annoncés⁹². D'une manière plus générale, les sites Web tendent à prendre des décisions qui entravent le libre accès à l'information et la libre circulation de l'internaute.

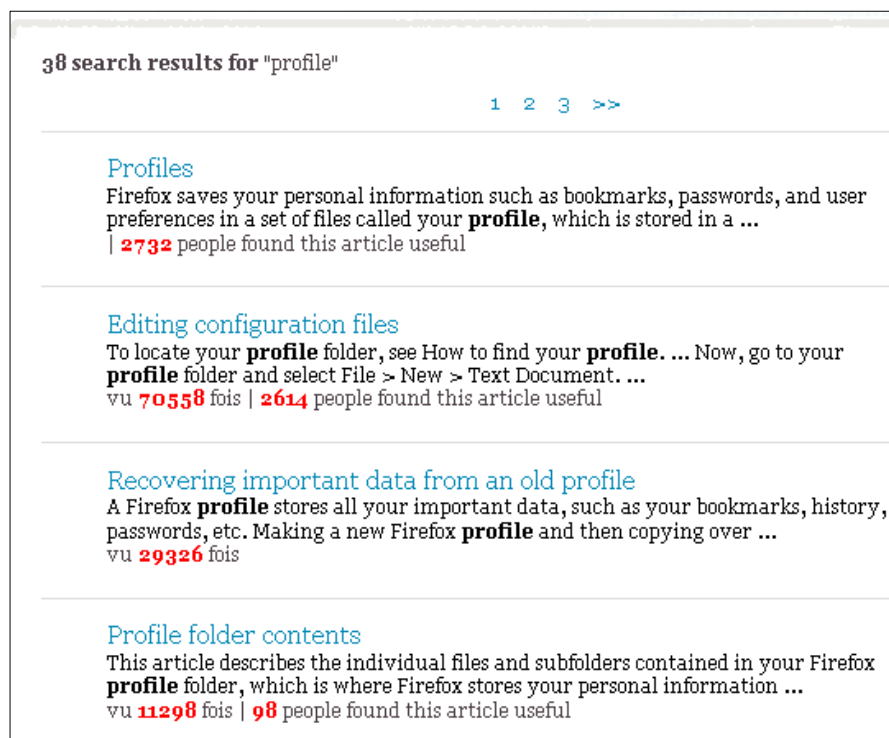
Un dernier exemple illustrera ce phénomène relativement massif. Localisée en France, je soumetts à "Google.com in English" la requête **Profile**. La première page de résultats est partiellement reproduite ci-après.



Je décide d'examiner la page dont l'annonce – en anglais – apparaît en troisième position, à propos du « profil utilisateur » défini par le navigateur Firefox, page dont je ne connais à ce stade que l'extrait suivant : "Firefox saves your personal information such as bookmarks, passwords, and user preferences in a set of files called your profile, which is stored in a ..." Or si je clique sur le lien [Profiles](#), j'arrive sur une page *en français* (ci-après). Alors que la page attendue a pour adresse <http://support.mozilla.com/kb/Profiles>, celle qui m'est proposée ici a pour adresse <http://support.mozilla.com/fr/kb/Profiles> !



Si j'essaie d'introduire la bonne adresse dans la barre du navigateur Firefox (ou du navigateur Internet Explorer), je suis invariablement ramenée à l'adresse parasite... Si j'utilise le moteur de recherche dans la *Knowledge Base* de Firefox (voir ci-dessus), j'obtiens 38 résultats en anglais, dont le premier n'est autre que celui cherché en vain jusqu'ici.



Si je clique sur le premier des 38 résultats annoncés, je suis à nouveau ramenée à la page en français déjà affichée ! La recherche d'information bute ici sur une stratégie qui, tout en se voulant bien intentionnée (on parle en français aux Français, etc.), restreint sensiblement la liberté de l'internaute.

Conclusion

Un équipement praxéologique adéquat en matière de RII suppose à l'évidence une connaissance non négligeable du monde Informatique & Internet, saisi dans la diversité de ses dimensions – techniques, technologiques, théoriques, historiques, politiques, juridiques, culturelles. À l'instar de Platon interdisant – selon une légende tardive, il est vrai – l'entrée de son académie à qui ne serait pas géomètre, car il faut être géomètre pour bien philosopher, nul n'entre véritablement dans le royaume de la RII s'il n'est, dans une certaine mesure, un « géomètre » de l'informatique et de l'Internet. Cela admis, il est plus délicat de déterminer quelles praxéologies du monde Informatique & Internet se révéleront précieuses, en certaines circonstances au moins, pour la RII.

Les treize domaines que nous avons retenus (treize, en comptant séparément les mathématiques, la physique et les technologies numériques) ont été choisis sur la base des explorations et autres rencontres réalisées dans les deux premières parties de ce travail. Nous l'avons déjà souligné : nous ne prétendons pas à l'exhaustivité. Dans la problématique où nous nous situons, celle de l'étude des conditions et contraintes de la définition d'une culture praxéologique de la RII créatrice d'intelligibilité et d'opérationnalité, l'enquête amorcée ici se veut seulement une ébauche, à poursuivre nécessairement en d'autres contextes de recherche⁹³. Dans l'état actuel, soulignons-le en passant, l'un de ses mérites devrait être, croyons-nous, de concourir à déconstruire le déni de problématique dont la RII fait trop souvent l'objet. Cela dit, l'identification des grands domaines praxéologiques pertinents pour une culture qui reste largement à construire va rarement de soi, et notre liste est donc provisoire et ouverte. Nous nous permettons ici une comparaison que seule la disponibilité d'un témoignage prestigieux autorise : en 1926, Sigmund Freud (1856-1939) écrit, sous la pression des circonstances⁹⁴, un petit livre polémique intitulé *Die Frage der Laienanalyse* (« La question de l'analyse profane »), où il aborde la question des contenus de formation des jeunes psychanalystes, en ferraillant contre ceux qui souhaitent leur donner surtout une formation de médecin. Soucieux de rectifier sur ce point la *doxa* de son temps, Freud note sans détour (1926/2003, p. 241) : « La formation analytique recoupe, certes, le champ de la préparation médicale, mais ne l'inclut pas et n'y est pas incluse. » Dans une école supérieure de psychanalyse, indique-t-il alors, il faudrait enseigner « bien des choses qu'enseigne

également la Faculté de médecine » (biologie, tableaux cliniques en psychiatrie, etc.) ; mais il faudrait englober aussi « des spécialités qui sont étrangères au médecin et qu'il ne rencontre pas dans son activité professionnelle ». Freud cite à ce titre l'histoire de la civilisation, la mythologie, la psychologie des religions et la littérature. Dans une postface de 1927, il ajoutera plus largement les « sciences de l'esprit », la sociologie, l'histoire de l'évolution (Freud, 1926/2003, p. 263) : la liste n'est pas close. Il commente alors l'énumération faite en ces termes : « Sans une bonne orientation en ces domaines, l'analyste reste sans comprendre une grande partie du matériel qui s'offre à lui. » Toutes proportions gardées, on retrouve avec la perspective d'une RII « pour tous » la situation décrite ici pour la psychanalyse, si du moins d'aucuns prétendaient faire de la RII un simple rejeton de la *computer science*, d'un côté, ou de la *library and information science*, d'un autre. Mais il nous importe davantage de noter un fait au reste banal dans la vie des savoirs : la nécessité, pour constituer une culture adéquate de la RII, de recourir à des champs praxéologiques *multiples*, que, jusqu'ici, la vie institutionnelle avait socialement séparés et, parfois, culturellement opposés. Ajoutons : d'y recourir d'une façon *spécifique*. Sur ce point, Freud est insistant. La connaissance des os du tarse comme la connaissance de certaines découvertes récentes de la médecine (par exemple « tout ce que la médecine a mis au jour concernant les agents pathogènes bacillaires et les moyens de les combattre ») n'ont pas la moindre importance pour l'analyste, écrit-il (p. 243), car elles « ne l'aident pas directement à comprendre une névrose et à la guérir, et ne contribuent pas davantage à aiguïser les facultés intellectuelles qu'exige au plus haut point son activité ». C'est là le thème des savoirs *utiles* – utiles à un certain projet praxéologique, pour nous la « science » de la RII.

Qu'est-ce qui permet alors, non plus de déterminer les disciplines éventuellement pertinentes, mais, au sein de chacune d'elle, d'identifier les connaissances utiles, qui pourraient bien « importer » (à la psychanalyse, à la RII) ? La réponse nous semble être, en règle générale, celle-ci : *la pratique* (pour la psychanalyse spécifiquement, la pratique freudienne, et celle de quelques autres psychanalystes autour de Freud). C'est en tout cas cette réponse qui s'impose, très modestement, dans notre propre cas : c'est notre pratique concrète et notre pratique « théorique » de la RII, du Web, de l'Internet, au cours des années passées d'abord, puis tout au long des travaux dont nous rendons compte ici, qui a guidé notre parcours à travers les « territoires » praxéologiques visités dans cette partie.

La TAD a dégagé, en matière d'enquête codisciplinaire (ou, de façon équivalente, de parcours d'étude et de recherche), six « dialectiques » utiles à l'enquêteur⁹⁵. Celle d'entre elles qui s'est trouvée d'emblée au premier plan de notre travail est sans doute la dialectique *du sujet et du hors-sujet*, praxéologie répondant à ce fait que, « dans une recherche non

aménagée par avance » – par un professeur bienveillant, par exemple –, « on doit à chaque instant prendre le risque du hors-sujet », risque qui inclut certainement le « risque d’aller examiner telle œuvre O_i en espérant y découvrir une réponse R^\diamond qu’on n’y trouvera pas » et, plus encore peut-être, le « risque de s’engager en tel chemin ouvert par telle question Q_j , qui ne mènera nulle part » (Chevallard, 2007d, p. 743). Mais il est une autre dialectique que nous avons eu à manier constamment : la dialectique *du parachutiste et du truffier*. Yves Chevallard a signalé en plusieurs occasions que le rapprochement de ces mots un peu inattendus était dû à l’historien Emmanuel Le Roy Ladurie. Il n’est pas inintéressant de revenir ici à la formulation originale de l’auteur cité, que voici (Le Roy Ladurie, 1982, pp. 207-208).

...selon l’expression qu’employait Van Regemorter, l’un de mes camarades de la rue d’Ulm devenu connaisseur de l’histoire russe, l’historien se devait d’être à la fois parachutiste et truffier. Les parachutistes ratissent un territoire (en fonction d’une métaphore regrettamment empruntée à la guerre d’Algérie). Ils s’efforcent d’en prendre la mesure et la connaissance la plus totale. Les truffiers, eux, repèrent tel trésor solitaire, longtemps enterré et caché.

Dans les vastes étendues que l’on s’efforce de « ratisser » (en s’appuyant pour cela sur des exposés que l’on espère appropriés), quels « trésors » attirent l’œil ? Tout d’abord, sans doute, il y a tout ce qui permet de restituer à ce geste vécu souvent comme strictement individuel et *privé* qu’est une *Web search*, sa plénitude de geste *social*, inscrit dans une histoire, participant *volens nolens* d’une politique et d’une économie, n’échappant pas aux règles du droit, non plus qu’à l’état de développement des technologies du numérique ou à l’activité tant de la population du globe presque toute entière aujourd’hui que de personnages singuliers et d’organismes spécialisés sans lesquels l’Internet ni le Web n’existeraient. Quand on ignore ce que sont les treize « serveurs de la racine », peut-on se demander ce qui se passerait si, brutalement, leur activité se trouvait totalement interrompue ? De même, lorsqu’on « oublie » l’énergie électrique nécessaire à la vie du Web, on rompt mentalement le lien inéluctable entre son propre *clickstream* et les problèmes écologiques de la planète ⁹⁶.

Une recherche d’information sur Internet est en ce sens ce que Marcel Mauss a appelé un *fait social total*, où « s’expriment à la fois et d’un coup toutes sortes d’institutions » (Mauss, 1923-1924/1950a, p. 147), notion que le sociologue Camille Tarot (1999) a pu commenter en ces termes.

Le fait social total n'est sûrement pas la totalité exhibée d'une société, car qui a jamais pu saisir pareille totalité, même dans la meilleure des monographies ? Il est cette propriété unique des faits qui sont l'objet même de la science de l'homme, d'être signifiants, c'est-à-dire partiels, contingents, arbitraires et cependant reliés, rattachés, dépendants toujours de quelque chose qui est en eux, à la fois manifesté et caché par eux. (p. 663)

L'exploration de la totalité qui se dissimule en un geste souvent éphémère fait apparaître ce qui se passe, comme l'on dit en anglais, *behind the scenes* ; ou, selon une opposition empruntée à Bachelard, nous plonge dans la nouménologie de la *Web search*.

S'il est possible d'avancer un principe heuristique simple, on peut dire que, au lieu de traiter le monde comme le fruit de conventions arbitraires que l'on prend à leur valeur faciale, on doit idéalement s'arrêter à chaque *mot* – et cela d'autant plus qu'il est opaque – et s'attacher à élucider la « chose » – l'élément praxéologique – à laquelle il est réputé renvoyer : l'étude porte ainsi d'un même mouvement sur les mots et les choses, et sur leurs raisons d'être institutionnelles. Bien entendu, l'attention aux mots n'est prônée ici, on l'aura compris, que pour autant qu'elle ouvre une voie d'accès vers les choses – les praxéologies –, ou plutôt en tant qu'elle est une partie du travail d'appropriation des choses : là où Diderot (cité par Durkheim, 1938/1990, p. 332) faisait valoir en son temps « le besoin de substituer à l'étude des mots l'étude des choses », nous considérons l'étude des mots comme une part certes limitée, mais souvent cruciale, de l'étude des choses – dont il faut bien que l'on puisse parler si on veut les étudier.

L'heuristique mise en œuvre conduit alors naturellement à pousser l'interrogation sans faire grand cas des frontières praxéologiques toutes tracées : à qui veut comprendre et agir, rien de ce qui est utile ne doit être étranger, au prétexte par exemple que telle connaissance serait l'affaire de l'informaticien, ou du juriste, ou du programmeur, ou du mathématicien, etc. La « pulsion praxéologique » que détermine un projet humain pleinement assumé a ainsi nécessairement partie liée avec *l'esprit encyclopédique*, comme le soulignait autrefois Durkheim citant la *Grande Didactique (Magna Didactica)* de Comenius (1592-1670).

« Tous ceux, dit Comenius, qui sortent de l'école pour aborder l'univers non comme de simples spectateurs, mais comme acteurs, *quicumque in mundum, non solum ut spectatores, sed etiam ut actores futuri immittuntur*, doivent posséder des notions de toutes les choses essentielles qui sont ou se font, *omnium principalium qua sunt et fiunt fundamenta*. » (Durkheim, 1938/1990, p. 329)

Et Durkheim, démarquant Comenius, concluait : « La culture doit être encyclopédique, parce qu'elle doit préparer l'homme à toutes les formes possibles d'action. »

Bien entendu, ce qui s'oppose pratiquement à l'encyclopédisme fonctionnel de l'acteur (qui n'est pas l'encyclopédisme formel du simple spectateur), ce sont les frontières érigées sur la carte des savoirs. Parmi les nombreux points d'arrêt que nous avons rencontrés dans cette troisième partie, l'un d'eux est peut-être plus emblématique que d'autres : il s'agit de cette ligne de démarcation que dessine la connaissance de la *programmation Web*. Le leitmotiv, ici comme ailleurs, est connu : *faut-il aller jusque-là !* Or pourquoi cette question est-elle soulevée comme légitime ? La réponse nous semble être toujours la même : parce qu'il existe une ou des corporations établies – ici, disons, celle des créateurs de sites Web, celle des programmeurs – qui possèdent, en quelque sorte par définition, cette connaissance. On voit le sophisme : cette connaissance serait l'apanage de corporations existantes ; l'acquérir reviendrait à prétendre y entrer ! Bien entendu, il n'en est rien : la connaissance de la programmation Web pour qui veut simplement être un *Web searcher instruit* n'est pas celle du programmeur de métier ; et de même pour le droit, l'informatique, les mathématiques, la physique, etc. Mais en même temps, acquérir l'essentiel des praxéologies utiles relevant de tel domaine, ainsi que nous l'avons souligné à propos du droit, c'est, selon le cas, apprendre ou réapprendre de l'informatique, du droit, de la physique, de la programmation en HTML, des mathématiques, sans qu'on ait jamais à s'excuser pour cela, mais au contraire en nous abandonnant au postulat, posé par méthode dans l'*Introduction* de notre deuxième partie, selon lequel personnes et institutions « font preuve d'une intelligence "intrinsèque" toujours suffisante pour apprendre (en un temps suffisant), quel que soit le contenu à apprendre »...

Loin que cette partie close l'inventaire des praxéologies utiles, elle en est seulement une plate-forme d'orientation, à partir de laquelle tout commence. Ose savoir ! On retrouve ici, ou on voudrait retrouver, le *Sapere aude !* kantien⁹⁷. Dans la partie suivante de notre travail, on découvrira pourtant que, même si l'on ose s'affranchir sans façon des frontières praxéologiques de toute nature et si l'on est bien décidé, le cas échéant, à aller « voir de l'autre côté », des obstacles d'une autre sorte vont surgir.

ANNEXE 3

Sur la notion d'éléments d'une science

Dans l'article « Éléments des sciences » de l'*Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers* (1751-1772), évoquant une science quelconque – ce serait pour nous la science de la RII –, d'Alembert écrit :

... supposons que cette science soit entièrement traitée dans un ouvrage, en sorte que l'on ait de suite et sous les yeux les propositions, tant générales que particulières, qui forment l'ensemble de la science et que ces propositions soient disposées dans l'ordre le plus naturel et le plus rigoureux qu'il soit possible... (1751/1986b)

Nous allons voir ici la fonction de cette supposition dans le schéma que d'Alembert déploie en cet article essentiel en matière d'organisation et de diffusion praxéologiques ⁹⁸.

Un premier mot doit être présenté et commenté sans attendre : celui d'*élément*, ou plutôt d'*éléments* (au pluriel). Émile Littré, dans son *Dictionnaire de la langue française* (1872-1877) en donne d'abord cette définition illustrée : « Notions premières. Les éléments de la grammaire. N'avoir que les premiers éléments d'une science, en être aux éléments. N'avoir pas les premiers éléments d'une science, l'ignorer complètement. » Telle est bien l'acception commune, courante, que chacun connaît. Littré ajoute un peu plus loin : « Éléments est le titre de certains ouvrages qui contiennent les premières notions d'un enseignement. Éléments de grammaire latine. » Derrière cet exemple se cache en vérité un autre sens, relativement effacé dans la langue française contemporaine, mais qui transparaît encore dans l'exemple – emprunté à Charles Rollin (1661-1741) – par lequel Littré prétend illustrer l'assertion précédente : « Ses Éléments [d'Euclide] contiennent une suite de propositions qui sont la base et le fondement de toutes les autres parties des mathématiques. » C'est que, contrairement au point de vue commun, des éléments ne sont pas des *rudiments* mais « la base et le fondement » d'une science.

Le concept remonte à l'antiquité, ambiguïté sémantique incluse. Dans un commentaire des *Éléments* d'Euclide (vers 300 av. J.-C.), se référant à Proclus (412-485), l'historien des mathématiques grecques Maurice Caveing (1990, p. 85) rappelle à cet égard l'opposition subtile autant que décisive entre *éléments* proprement dits et propositions *élémentaires*.

Sont des « éléments » les théorèmes grâce auxquels la théorie progresse vers la connaissance d'autres propositions et qui fournissent la solution des difficultés que celles-ci recèlent. Ils sont en cela comparables aux lettres de l'alphabet qui sont les éléments simples, premiers et indivisibles dont l'arrangement constitue tout mot et tout discours écrits. On retrouve donc ici le sens étymologique qui renvoie à *στοίχος* : « rang », « rangée », « ordre (de bataille) ». Des objets A sont des *στοιχεῖα* de l'objet B seulement si leur arrangement produit B. De même, pour l'ensemble de la Géométrie, poursuit Proclus, il est certains théorèmes directeurs, entretenant avec ceux qui les suivent un rapport de principes à conséquences, qu'on retrouve partout, et qui procurent les démonstrations de nombreux cas particuliers. Par contre, sont élémentaires toutes les propositions qui, bien qu'intervenant dans un certain nombre d'autres, même un assez grand nombre, et ne manquant ni de simplicité ni d'élégance, n'ont pourtant pas valeur d'éléments du fait que leur considération n'est pas commune à la science en toutes ses parties, comme, par exemple, que les hauteurs d'un triangle sont concourantes.

De fait, cette dernière propriété mathématique (les hauteurs d'un triangle sont concourantes), élémentaire au sens de Proclus comme au sens courant du terme, *n'apparaît pas* dans les *Éléments* d'Euclide : elle n'est pas un élément des mathématiques. Un livre d'éléments n'a pas vocation à organiser *toutes* les connaissances disponibles : on voit que la définition proposée est, dans son principe, rigoureuse. À sa lumière, le fait qu'un élément apparaisse parfois, pour l'usager ordinaire d'une science, comme un « simple rudiment » est chose secondaire, quoique ce double régime – être un élément que d'aucuns jugent « rudimentaire » – ait le démérite de masquer la puissance générative propre qui est celle de tout élément. Par contraste, une connaissance simplement rudimentaire peut être inapte à engendrer d'autres connaissances et tout rudiment n'a donc pas à figurer parmi les éléments d'une science, contrairement à ce qu'une tradition paresseuse a pu laisser croire. C'est ce que note encore M. Caveing dans ce passage.

La confusion entre la notion d'« élément » et celle de « proposition élémentaire » est à l'origine de l'assimilation du traité d'Euclide à un manuel scolaire. Sans doute Proclus indique le rôle pédagogique des *Éléments* [...], en fait plutôt propédeutique, en ce que les mathématiques doivent, selon la doctrine des Platoniciens, élever l'esprit jusqu'aux idéalités. L'interprétation purement scolaire des *Éléments* comme un traité « élémentaire » a conduit, au cours des âges, et spécialement en Europe depuis la Renaissance, à l'édition [...] de nombreuses traductions amputées, se limitant aux quatre, ou aux six premiers Livres [sur les treize que comptent les *Éléments*], et à de nombreuses adaptations, avec rajouts et omissions, qui ont défiguré l'œuvre d'Euclide.

Le texte de d'Alembert s'inscrit sans ambages dans la haute tradition euclidienne des *Éléments*. L'auteur note d'abord, en conformité avec cette tradition, ceci.

... il est facile de distinguer les propositions ou vérités générales qui servent de base aux autres, & dans lesquelles celles-ci sont implicitement renfermées. Ces propositions réunies en un corps, formeront, à proprement parler, les *éléments* de la science, puisque ces *éléments* seront comme un germe qu'il suffiroit de développer pour connoître les objets de la science fort en détail.

Il n'oublie pas le second sens du mot, qui n'est cependant que secondaire, parce que « moins philosophique » que le premier, précisément.

... on peut encore considérer les *éléments* d'une science sous un autre point de vûe : en effet, dans la suite des propositions on peut distinguer celles qui, soit dans elles-mêmes, soit dans leurs conséquences, considerent cet objet de la maniere la plus simple ; & ces propositions étant détachées du tout, en y joignant même les conséquences détaillées qui en dérivent immédiatement, on aura des *éléments* pris dans un second sens plus vulgaire & plus en usage, mais moins philosophique que le premier.

D'Alembert rapproche ensuite les deux acceptions pour les comparer, avant de déclarer nettement que c'est au *premier sens* qu'il s'attachera dans la suite de son analyse.

Les *éléments* pris dans le premier sens, considerent pour ainsi dire en gros toutes les parties principales de l'objet : les *éléments* pris dans le second sens, considerent en détail les parties de l'objet les plus grossieres. Ainsi des *éléments* de Géométrie qui contiendroient non-seulement les principes de la mesure & des propriétés des figures planes, mais ceux de l'application de l'Algebre à la Géométrie, & du calcul différentiel & intégral appliqués aux courbes, seroient des *éléments* de Géométrie dans le premier sens, parce qu'ils renferméroient les principes de la Géométrie prise dans toute son étendue ; mais ce qu'on appelle des *éléments de Géométrie ordinaire*, qui ne roulent que sur les propriétés générales des figures planes & du cercle, ne sont que des *éléments* pris dans le second sens, parce qu'ils n'embrassent que la partie la plus simple de leur objet, soit qu'ils l'embrassent avec plus ou moins de détail. Nous nous attacherons ici aux *éléments* pris dans le premier sens ; ce que nous en dirons pourra facilement s'appliquer ensuite aux *éléments* pris dans le second.

La grande ambition, donc, c'est de parvenir à dégager et à organiser des *éléments* d'une science donnée. Le problème est immense : écrire des *Éléments* (d'algèbre, de philosophie, etc.) est autre chose que d'écrire des *Leçons* (d'algèbre, de philosophie, etc.), par exemple. Un second mot doit alors être avancé, un verbe : *élémenter*. Et le substantif qui lui correspond : *élémentation*. Du verbe, le *Dictionnaire historique de la langue française* (Rey et al., 1993) a conservé la trace, tout en signant son avis de décès – « “simplifier de façon à donner les principes, les éléments” (XVIII^e s.) n'est plus en usage », y lit-on en effet. Le mot apparaît sous la plume de Lakanal lorsqu'il commente, en octobre 1794, l'échec du concours ouvert par un décret de la Convention daté du 9 pluviôse an II (28 janvier 1794) touchant la rédaction de livres *élémentaires* pour l'école primaire (Julia, 1981, p. 239) ; il écrit en effet ceci.

Les citoyens qui ont travaillé pour ce concours ont généralement confondu deux objets très différents, des *élémentaires* avec des *abrégés*. Resserrer, coarcter [= rendre plus étroit] un long ouvrage, c'est l'abrégé ; présenter les premiers germes et en quelque sorte la matrice d'une science, c'est l'élémenter.

Notons que le critère d'une élémentation réussie est ainsi d'abord *épistémologique*, bien avant que d'être « pédagogique » : l'organisation praxéologique prime l'organisation didactique, qu'elle inspire et détermine. Mais il nous faut surtout, ici, insister encore sur la distinction des rudiments et des éléments : car ce sont des rudiments plutôt que des éléments que, en dépit de leurs titres trompeurs, on trouve fréquemment dans les ouvrages dits « élémentaires » : ceux-ci, lorsqu'ils sont rédigés en anglais, proposent ainsi de prétendus *basics* et autres *fundamentals* qui, de par leur dénomination, seraient censés contenir les “principles from which other truths can be derived”, comme l'indique tel le *OneLook Dictionary*, ce qui est souvent loin d'être le cas. Les exposés sur la RII n'échappent évidemment pas à cette règle, qui semble n'avoir que de rares exceptions. La matière qui, souvent, y est rassemblée n'est en effet nullement organisée « selon l'ordre des raisons » : quand elle n'est pas le pur produit d'un rassemblement sans principes directeurs nets, elle apparaît comme la résultante de multiples motivations opportunistes – *musts* du moment, exigences d'éditeur, marottes d'auteur, visées publicitaires, volonté de séduire un certain public, etc.

Il semble cependant que la période actuelle voit l'effacement de désignations traditionnelles – celle d'*abrégé* ou celle de *précis* (ouvrage dont Littré dit qu'il « se distingue par sa rigueur », ne comporte « rien de superflu » et se veut « substantiel » lorsque l'abrégé est simplement « court ») – et la floraison corrélative des *Fondamentaux de...* et autres *Éléments de...* – tels des « Éléments de programmation » proposés en ligne pour le B2i niveau

École⁹⁹. Malgré l'ambiguïté sémantique déjà notée, il est possible qu'on puisse entendre dans ce déplacement lexical un obscur désir, collectivement ressenti, d'élémentation des savoirs. Sans doute le concept exact d'éléments est-il plus qu'à demi perdu ; mais ce serait alors sa *raison d'être* dans le fonctionnement de la connaissance qui serait ainsi spontanément, quoique confusément, retrouvée. Quoiqu'il en soit, la reconquête de ce concept suppose, nous semble-t-il, un approfondissement de la rupture avec la notion de rudiments, qui permette de passer de l'idée commune de b.a.-ba à celle, déjà plus avancée, d'abc, et, au-delà encore, à la notion même d'éléments. Or cet approfondissement réserve quelques surprises à la *doxa* de notre temps.

Sans doute d'Alembert s'inspire-t-il en toute rigueur de Descartes écrivant dans les *Secondes réponses* (Descartes, 1647/1979, p. 174) : « L'ordre consiste en cela seulement que les choses qui sont supposées les premières doivent être connues sans l'aide des suivantes, et que les suivantes doivent après être disposées de telle façon qu'elles soient démontrées par les seules choses qui les précèdent. » Mais ce précepte seul ne suffit pas à faire des éléments. Une courte analyse permet à « l'encyclopède » d'Alembert, comme l'appelait Voltaire, de parvenir à cet égard à la conclusion que voici.

... les conditions nécessaires pour qu'une proposition entre dans les *éléments* d'une science pris dans le premier sens, sont que ces propositions soient assez distinguées les unes des autres, pour qu'on ne puisse pas en former une chaîne immédiate ; que ces propositions soient elles-mêmes la source de plusieurs autres, qui n'en seront plus regardées que comme des conséquences ; & qu'enfin si quelqu'une des propositions est comprise dans les précédentes, elle n'y soit comprise qu'implicitement, ou de manière qu'on ne puisse en appercevoir la dépendance que par un raisonnement développé.

Le principe, on le voit, ne livre pas un algorithme implacable : il faut parfois l'interpréter, et savoir prendre avec lui les accommodements qu'appelle notre incapacité à percevoir d'un coup des liens de consécution logique parfois profondément enfouis. Mais un autre point d'importance doit être pris en charge dans un processus d'élémentation bien conçu, ainsi que le précise ici l'encyclopédiste.

N'oublions pas de dire qu'il faut insérer dans les *éléments* les propositions isolées, s'il en est quelqu'une qui ne tienne ni comme principe ni comme conséquence, à aucune autre ; car les éléments d'une science doivent contenir au moins le germe de toutes les vérités qui font l'objet de cette science : par conséquent l'omission d'une seule vérité isolée, rendrait les *éléments* imparfaits.

Loin donc qu'un ouvrage d'éléments soit un « abrégé », rognant parfois sur des connaissances utiles, il doit contenir *tout* ce qui est nécessaire pour *engendrer la totalité utile d'une science donnée*. La notion d'abrégé (ou de digest, de condensé, d'aperçu, d'épitomé, de compendium, etc.) se réfère à la lettre, au *texte* ; la notion d'éléments touche à l'épistémologie, au *savoir*.

Ce dernier principe a une conséquence essentielle, quoique difficile à entendre aujourd'hui : contre un certain sens commun, la notion d'éléments a en effet partie liée avec *l'idéal encyclopédique* ; disons le mot : avec *l'encyclopédisme* au sens où Littré entend encore ce terme – celui de « système des encyclopédistes », dont le principe directeur, qui seul nous importe ici, a pu être énoncé dans les termes suivants (Kintzler, 1996, p. 50).

... l'*Encyclopédie* est fondée sur la thèse selon laquelle toute connaissance, pour être exposée encyclopédiquement et faire l'objet d'un article, peut être reconstruite dans un ordre éclairant, dans un ordre qui la rend intelligible et appropriable. Cet ordre gouverne l'idée directrice qui inspire chaque article de *l'Encyclopédie*, rédigé de telle sorte que le lecteur puisse partir de ce qu'il y a de plus élémentaire pour parvenir à ce qu'il y a de plus complexe.

Cette *obligation encyclopédique* est évidemment liée à l'exigence définitoire que des éléments d'une science forment une « grammaire générative » (pour user ici d'un anachronisme) capable d'engendrer *la totalité* des connaissances d'une science, dans son irréductible diversité, ou du moins la totalité de celles qu'on pense utiles à un certain projet de connaissance et d'action. Selon la métaphore cartographique employée par d'Alembert dans le *Discours préliminaire*, on aura en ce cas un système de « cartes » (les articles) constituant ensemble la « mappemonde » (l'encyclopédie elle-même). Bien entendu, cette organisation « selon l'ordre des intelligibilités » possède *pour cela même* une vertu didactique. Le grec *enkuklios paideia*, adapté en latin de la Renaissance sous la forme *encyclopedia*, ne désigne-t-il pas, au reste, une instruction (*paideia*) qui embrasse le cercle (*enkuklios* dérive de *kuklos*, « cercle ») des connaissances d'un certain domaine ? Mais l'organisation praxéologique à construire doit proposer pour cela un *texte* du savoir – le texte des éléments – qui soit respectueux du *savoir* élémenté lui-même.

La confusion des deux acceptions du nom d'éléments est à l'origine de la péjoration courante, actuelle, de « l'encyclopédisme », comme le rappelle ci-après un commentateur attentif (Pena-Ruiz, 2005, p. 98).

La critique de l'idéal encyclopédique de l'instruction relève d'une confusion devenue fréquente entre cercle raisonné du savoir bien compris en ses principes essentiels et empilement de « connaissances » assimilées à des « contenus », à des « informations » qu'il s'agirait d'emmagasiner. Cette caricature ouvre la voie à l'acception péjorative de ce qui est appelé « encyclopédisme ». L'idéal encyclopédique bien compris invite tout au contraire à une connaissance maîtrisée de la différenciation des savoirs en leurs principes élémentaires, non à une érudition exhaustive.

Quelque « science » que l'on considère, ou, dans le vocabulaire de la TAD, quelque système praxéologique qu'on ait en vue, il convient, ne serait-ce que pour *préciser ce dont on parle*, d'en envisager une élémentation – encyclopédique par définition – appropriée à la problématique praxéologique que cette élémentation doit servir : on n'élémente pas la « science informatique » de la même façon selon qu'on vise par là à élémenter le domaine de l'algorithmique ou celui de la RII. Ce point est évidemment décisif : s'il est vrai qu'on ne saurait élémenter une science *une fois pour toutes*, ce n'est pas *seulement* parce que cette science vit, bouge, se développe en diachronie ; c'est aussi parce que toute élémentation, par son principe même, sert en synchronie un projet de connaissance et d'action déterminé, auquel elle doit par définition se rendre adéquate : elle « représente » (au sens quasi diplomatique du terme) cette science auprès de ce projet et de ses besoins praxéologiques propres. Une élémentation de l'algèbre matricielle en vue de son emploi en *économie* diffèrera sans doute d'une élémentation du même domaine mathématique au profit de l'utilisateur de la *théorie des graphes* en sciences sociales, par exemple. Semblablement, il est douteux que le didacticien des mathématiques ait des besoins en matière de linguistique que, par l'effet d'une harmonie préétablie, puissent spontanément satisfaire les deux premières années d'un cursus universitaire classique de linguistique : ce qui est « éléments » de linguistique pour le didacticien sera tantôt jugé *rudimentaire*, et tantôt *non* élémentaire (mais au contraire « avancé ») du point de vue de l'organisation canonique de l'allocation des connaissances dans un cursus classique de linguistique, c'est-à-dire dans une formation adressée idéalement à de futurs linguistes *qui ne seraient guère que cela*. Or l'un des grands dangers de la gestion praxéologique de nos sociétés (et l'un des pièges essentiels de l'épistémologie spontanée qui s'y exprime) réside précisément dans le fait d'identifier une « science » à *telle élémentation* fixée (et souvent figée) de cette science, dès lors que cette élémentation est devenue traditionnelle, voire quasi naturalisée, du fait notamment de la canonisation scolaire ou universitaire du texte du savoir qui l'exprime (en relation étroite avec la cartographie des « disciplines » établies), à l'exclusion de toutes les autres élémentations possibles. Cette

quasi-sacralisation d'une certaine version d'une « science » se produit parfois alors même que ladite version a peut-être cessé d'élémenter cette science de façon authentique et surtout de manière adéquate aux besoins praxéologiques actuels qu'elle peut avoir à satisfaire : il faudra alors *réélémenter* la science en question, et cela peut-être plusieurs fois plutôt qu'une, en fonction de la diversité des projets à servir.

¹ L'un d'eux apparaît dans le dernier chapitre de la deuxième partie, sur une copie d'écran : il s'agit de l'ouvrage de Greg R. Notess (2006) intitulé *Teaching Web Search Skills*, sur lequel nous reviendrons.

² Le lecteur qui l'aurait oublié pourra retrouver ce « bruit reconnaissable par n'importe qui » (d'après l'article « Modem » de l'encyclopédie Wikipédia) à l'adresse suivante : <http://www.lazylaces.com/56Kmodem/>. Dans un article mis en ligne le 5 janvier 2008 sur le site du quotidien belge *Le Soir*, le journaliste Alain Jennotte note, sous le titre « Fin janvier, Belgacom débranchera ses modems », que les connexions dites « dial-up » ne touchent plus, aujourd'hui que 6 à 7 % des internautes, surtout présents dans les zones rurales.

³ Sur MCI, que l'auteur mentionne ci-après, voir l'article "MCI Communications" de l'encyclopédie Wikipedia.

⁴ Steve Ballmer, mentionné ci-après, est le PDG de Microsoft.

⁵ Sur ce sujet, voir Haski, 2008.

⁶ On ne peut que renvoyer, ici, à la première partie de ce travail.

⁷ Voir l'article "Library Science" de l'encyclopédie Wikipedia.

⁸ Voir <http://savoircdi.cndp.fr/pedago/Refbibliogr/presentbibliolyc.htm>.

⁹ Pour une illustration touchant un thème examiné ci-après, voir <http://www.jetaide.com/read/f228/i48t48/droit-image.htm>.

¹⁰ Le B2i est le « Brevet informatique et internet ». La dénomination « B2i – Brevet informatique et internet » a été déposée comme marque, pour le compte de l'Éducation nationale, le 5 décembre 2000 à l'Institut National de la Propriété industrielle. Le niveau « école » du B2i est intégré aux programmes de l'école primaire depuis la rentrée 2002 : l'objectif est que, à terme, tous les élèves sortant de l'école primaire soient titulaires de l'attestation « B2i école ». Le B2i niveau « collège » est obligatoire pour l'obtention du DNB (diplôme national du brevet) à partir de la session 2008. Dans la perspective des niveaux 1 (école) et 2 (collège), la création d'un niveau 3 (lycée) du B2i est à l'étude (voir <http://www2.educnet.education.fr/formation/certification/b2i/> et <ftp://trf.education.gouv.fr/pub/edutel/bo/2006/42/encart42.pdf>).

¹¹ Voir <http://www.education.gouv.fr/bo/2006/29/MENE0601490A.htm>.

¹² En réponse à la question « Pourquoi le domaine "attitude responsable" est-il particulier ? », on lit ainsi sur le site de l'académie d'Aix-Marseille (<http://b2i.ac-aix-marseille.fr/strategie.htm>) : « Ces ensembles d'items (4 pour l'école, et 7 pour les collèges et lycées) sont plus des attitudes que des compétences. Leur présence va

permettre aux professeurs d'orienter les élèves vers une prise de conscience de la prudence, du respect des autres, de l'aide que l'on peut apporter grâce ou avec ces technologies. »

¹³ Pour plus d'information, voir <http://www.education.gouv.fr/bo/2005/15/MENT0500561C.htm> (circulaire du 7 avril 2005) et aussi <http://www2.c2i.education.fr/sections/c2i1/referentiel/refa>.

¹⁴ Voir Iteanu, 2007, ainsi par exemple que le blog « Identité numérique » (<http://www.identite-numerique.fr/>).

¹⁵ Voir <http://www.commentcamarche.net/droits/droit-image.php3>.

¹⁶ Le CRDP est le Centre Régional de Documentation Pédagogique (voir <http://www.crdp.ac-grenoble.fr/>).

¹⁷ Le document du site Eduscol (<http://eduscol.education.fr/D0053/accueil.htm>) souligne que le B2i « s'inscrit en cohérence avec le certificat informatique et internet (C2i) mis en place simultanément dans l'enseignement supérieur ». Voir aussi <http://www.education.gouv.fr/bo/2000/42/encart.htm>.

¹⁸ Voir <http://conventions.coe.int/treaty/Commun/QueVoulezVous.asp?NT=005&CL=FRE>.

¹⁹ Voir <http://www.assemblee-nationale.fr/12/propositions/pion1029.asp>.

²⁰ La loi DADVSI est la loi n° 2006-961 du 1^{er} août 2006 « relative au droit d'auteur et aux droits voisins dans la société de l'information ». Voir par exemple <http://fr.wikipedia.org/wiki/DADVSI>.

²¹ Voir http://www.wipo.int/treaties/fr/ip/berne/trtdocs_wo001.html#P148_30082.

²² Voir http://www.celog.fr/cpi/lv1_tt2.htm#3.

²³ Pour plus de précision, voir <http://www.zdnet.fr/actualites/imprimer/0.50000200.2118898.00.htm>.

²⁴ Voir <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000266350&dateTexte=>.

²⁵ Pour plus d'informations pratiques, une personne intéressée peut se reporter par exemple aux pages *Légamédia* du site Educnet (<http://www2.educnet.education.fr/legamedia>) ; ou encore aux pages appropriées de l'ouvrage de référence dû à Yann Buttner et André Maurin, *Le droit de la vie scolaire* (2007).

²⁶ Voir Larrieu, 2005, pp. 52-53.

²⁷ Ainsi que l'auteur cité le rappelle dans une note de bas de page, « méga = 10⁶, giga = 10⁹, téra = 10¹², péta = 10¹⁵ ».

²⁸ Une autre question, qui constitue un horizon de cette recherche, porte sur les conditions et contraintes sous lesquelles la diffusion de telle ou telle connaissance apparaîtrait possible à un coût raisonnable : nous en différons l'examen, pour nous concentrer ici sur la première question.

²⁹ Voir http://www.louisneel-centenaire.inpg.fr/IMG/pdf/13-Fiche_alternateur.pdf.

³⁰ Voir <http://www.education.gouv.fr/bo/1999/hs4/physique.htm>.

³¹ Dans le passage ci-après, comme dans le suivant, c'est nous qui soulignons.

³² Voir <http://www.cs.utexas.edu/users/EWD/ewd08xx/EWD898.PDF>. Dijkstra avait l'habitude de rédiger à la main des notes, connues sous le sigle d'EWD (les initiales de son nom), adressées à de multiples correspondants. La dernière, datée de 2002, l'année de sa disparition, est l'EWD 1318 : Dijkstra y démontre avec une extrême économie de moyens une égalité algébrique utilisée par H. S. M. Coxeter dans son *Introduction to Geometry*

(1961) : si l'on pose $s = (a + b + c)/2$, alors on a l'identité $abc = s(s - b)(s - c) + s(s - c)(s - a) + s(s - a)(s - b) - (s - a)(s - b)(s - c)$. (Voir <http://www.cs.utexas.edu/~EWD/welcome.html>.)

³³ Voir http://fr.wikipedia.org/wiki/Edsger_Dijkstra.

³⁴ Le nom de Bill Gates est ainsi lié à la diffusion de BASIC.

³⁵ On observera ici une petite imprécision, due sans doute au souci de l'auteur de faire court : le disque dur, que l'auteur a présenté dans la série des différentes sortes de mémoire, n'est pas fixé sur la carte mère – il s'y trouve seulement connecté. La chose serait évidente si, comme nous le verrons chez d'autres auteurs, une représentation figurée était proposée.

³⁶ Dans cette problématique, l'étude de la *genèse* des organisations praxéologiques observées ne saurait avoir qu'une fonction de *moyen* de notre travail, et non un statut de *fin*.

³⁷ La vue proposée ne mentionne pas de graveur de DVD, sans doute pour faire court (pour la même raison, il est question de graveur de CD, non de lecteur de CD...). Le texte lui-même consacrera des développements spécifiques respectivement au lecteur/graveur de CD et au lecteur/graveur de DVD.

³⁸ Voir <http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?refid=1861590666>.

³⁹ Voir <http://www.commentcamarche.net/pc/rom.php3>.

⁴⁰ Voir <http://www.vulgarisation-informatique.com/bios.php>.

⁴¹ Voir <http://www.dicodunet.com/annuaire/def-453-bios.htm>.

⁴² Voir <http://linuxfr.org/2007/08/22/23027.html>.

⁴³ Voir <http://franceterme.culture.fr/FranceTerme/recherche.html>.

⁴⁴ Voir <http://elearning.algonquincollege.com/coursemat/pincka/dat2343/lectures.f03/16-LMC-Bootstrap.htm>.

⁴⁵ Son emploi décline après la Première Guerre mondiale. Selon l'article « Quaternion » de l'encyclopédie Wikipédia, les systèmes de nombres hypercomplexes « sont maintenant considérés comme des exemples d'algèbres semi-simples ». (Voir cependant Weisstein, 1999-2008.)

⁴⁶ À la requête "**chercher sur le Web**" le moteur de recherche Google associe 36 800 résultats, tandis qu'il n'en renvoie que 4910 en réponse à la requête "**fouiller le Web**" (observé le 13 mars 2008 à 11 h 30). Par ailleurs, « fouille » est maintenant usité en français par rendre *mining* dans *data mining* : on parle de « fouille des données » (ou « dans les données ») et l'on a proposé récemment de parler de « fouille du Web » pour rendre l'anglais *Web mining* (voir Ertzscheid, 2005).

⁴⁷ On trouve sur Internet (voir <http://www.cp.asso.fr/depall/europe/edint03a.htm>) l'annonce d'un colloque qui s'est tenu en 2003 au Lycée François d'Estaing de Rodez ; sur le programme en français, la matinée du vendredi est présentée comme « consacrée à une recherche documentaire (soit sur Internet, soit au C.D.I.) pour produire une page sur un thème concernant les insectes ». Dans la version anglaise du même programme, on lit que cette matinée "will be dedicated to a research work (on the Internet or at the school library) to write a page on a

subject about insects”. Sans doute aurait-il été plus conforme à l’usage de parler de *search* en lieu et place de *research*.

⁴⁸ En catalan, toutefois, la distinction est entre *cerca* et *recerca* : elle est donc analogue à celle faite en anglais.

⁴⁹ Nous revenons plus loin sur les différents types de couches (*layers*).

⁵⁰ C’est ainsi que, dans l’édition du 26 octobre 1993 du *New York Times*, on pouvait lire ceci : “One of the technologies Vice President Al Gore is pushing is the information superhighway, which will link everyone at home or office to everything else—movies and television shows, shopping services, electronic mail and huge collections of data.”

⁵¹ Voir l’entretien avec Vinton Cerf paru dans *Le Monde* du 5 avril 2008 (Cerf avec Alberganti, 2008).

⁵² Dans *Le nouvel esprit scientifique* (1934), Gaston Bachelard (1884-1962) écrit ainsi : « La véritable phénoménologie scientifique est donc bien une phénoménotechnique. Elle renforce ce qui transparaît derrière ce qui apparaît. Elle s’instruit par ce qu’elle construit. » Dans *La formation de l’esprit scientifique* (1938), il écrira encore : « La phénoménotechnique *étend* la phénoménologie. Un concept est devenu scientifique dans la proportion où il est devenu technique, où il est accompagné d’une technique de réalisation. »

⁵³ Virieux, 2004, pp. 11-12.

⁵⁴ Le lecteur pourra vérifier que les deux écritures « fonctionnent » également.

⁵⁵ Sur le premier des logiciels mentionnés ci-après, voir l’article « KaZaA » de Wikipédia. Le logiciel eDonkey2000 n’est plus officiellement développé ; c’est aujourd’hui le logiciel libre eMule, lancé en 2002, qui semble le plus largement utilisé. Notons par ailleurs que l’expression « paire à paire » employée par l’auteure semble le fruit d’un lapsus (sous l’influence peut-être de la terminologie du câblage des réseaux) : le *Journal officiel* du 13 mai 2005 propose pour traduction de *peer-to-peer* en premier lieu « poste à poste », en second lieu « pair à pair ».

⁵⁶ Le protocole TCP est officiellement décrit dans le RFC 793 daté de septembre 1981. (Nous revenons plus loin sur la notion de RFC.)

⁵⁷ Les sigles POP et IMAP correspondent respectivement aux dénominations *Post Office Protocol* et *Internet Message Access Protocol*.

⁵⁸ L’auteure précise encore : « Quand le transfert des données est terminé, la connexion est fermée afin de libérer les ressources qui avaient été réservées. »

⁵⁹ Voir par exemple l’article « Somme de contrôle » de l’encyclopédie Wikipédia.

⁶⁰ Les adresses Internet sont, on le sait, codées sur 32 bits, alors que les adresses Ethernet le sont sur 48 bits (coder une adresse Ethernet revient en effet à coder 6 entiers compris entre 0 et 255 ; pour chacun d’eux, il faut pour cela 8 bits ; il faut donc $6 \times 8 = 48$ bits pour coder l’adresse). On a $256^6 = 281\,474\,976\,710\,656$, soit plus de 281 000 milliards.

⁶¹ Voir ainsi l’article de “Bit error ratio” de l’encyclopédie Wikipedia.

⁶² Voir l'article "Communications protocol" de Wikipedia.

⁶³ Pour un poste de travail équipé de Windows Vista, on pourra voir ainsi l'aide spécifique apportée par la firme Microsoft (voir <http://windowshelp.microsoft.com/Windows/fr-FR/default.msp>) ou encore tel ouvrage traitant de l'Internet (par exemple Lilen, 2008, pp. 22-125).

⁶⁴ Toute publicité mise à part, on pourra à titre d'exemple voir l'offre de service figurant à l'adresse suivante : <http://www.connectassistance.fr/ca/index.php?r=services&p=install&gclid=COST-fOuw5ICFQYUuwodBQ42Xw>.

⁶⁵ Observons toutefois que, sans pour autant vider le cache, on peut aussi vouloir solliciter à nouveau le serveur, parce qu'on souhaite réactualiser la page, c'est-à-dire en obtenir le dernier état, dans la mesure où cette page a pu être modifiée depuis sa mise en cache sur l'ordinateur de l'utilisateur.

⁶⁶ Nous nous référons ici à la version accessible en ligne le 29 avril 2008. Cette version indique que la dernière modification a eu lieu le 28/04/2008 à 19 h 41.

⁶⁷ L'article "Hit (internet)" de l'encyclopédie Wikipedia précise : "A **hit** is a request to a web server for a file (web page, image, JavaScript, Cascading Style Sheet, etc.). When a web page is uploaded from a server the number of "hits" or "page hits" is equal to the number of files requested, therefore one page load does not always equal one hit because often pages are made up of other images and other files which stack up the number of hits counted. Because one page load does not equal one hit it is an inaccurate measure of a websites popularity or web traffic. A more accurate measure of web traffic is how many page views a web site has." Le nombre de *hits* permet toutefois d'évaluer les besoins de puissance du serveur hébergeant le site. Sur tel serveur dont nous avons pu consulter les statistiques pour le mois de juin 2008, on observe ainsi les résultats suivants : 322 visiteurs différents y ont effectué 535 visites, qui se sont traduites par 1603 visites de pages (*page views*) et 4306 « hits ».

⁶⁸ Voir <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>.

⁶⁹ Voir <http://groups.google.com/group/alt.hypertext/msg/395f282a67a1916c>.

⁷⁰ Voir <http://tenyears-www.web.cern.ch/tenyears-www/Declaration/Page2.html>.

⁷¹ Le *Dictionnaire bilingue...* présente Gopher comme « un système d'information distribuée créé par l'université du Minnesota en 1991, aujourd'hui sérieusement concurrencé par la Toile, et qui fonctionne en mode client-serveur sur l'Internet » (Beneson & Juanals, 2004, p. 176).

⁷² Le développement de *Mosaic*, commencé fin 1992, sera arrêté le 7 janvier 1997 ; *Netscape Navigator* va alors lui succéder.

⁷³ Voir <http://blog.netscape.com/2007/12/28/end-of-support-for-netscape-web-browsers/>.

⁷⁴ HyperCard, précise le *Dictionnaire bilingue* de Benenson et Juanals, fut « le premier logiciel (en 1987) qui permettait à des utilisateurs non informaticiens de concevoir de façon simple un programme hypermédia ».

⁷⁵ Voir <http://www.rfc-editor.org/rfc/rfc3986.txt>.

⁷⁶ Le *Journal Officiel* du 8 décembre 2002 propose *arobase* comme synonyme d'*arrobe*, mais c'est ce dernier vocable que la Commission générale de terminologie et de néologie semble préférer. (Voir le site FranceTerme : <http://franceterme.culture.fr/FranceTerme/recherche.html>.)

⁷⁷ En fait, la technique qui sera mise en œuvre dans la troisième partie de notre travail se réfère à des corpus d'exposés déterminés à l'avance selon des critères indépendants des questions qui leur seront posées, alors que, ici, on doit seulement trouver des exposés, s'il en existe, apportant une réponse à la question posée.

⁷⁸ Voir <http://www.commentcamarche.net/internet/url.php3>.

⁷⁹ Contrairement au choix fait dans ce tableau pour schématiser la structure d'une URL, la barre oblique qui précède le chemin d'accès ne fait en principe pas partie de ce chemin. Comme le précise le RFC 1738 déjà cité, "the '/' between the host or port and the URL-path is NOT part of the URL-path" (p. 5).

⁸⁰ Voir http://fr.wikipedia.org/wiki/Adresse_web.

⁸¹ Voir http://fr.wikipedia.org/wiki/Discuter:Adresse_Web.

⁸² Voir <http://www.learnthenet.com/english/html/16addrss.htm>. Le lecteur attentif aura d'ailleurs noté que l'expression "Web address" figure dans un passage de l'article (en anglais) de Wikipedia intitulé "Uniform Resource Locator" que nous avons cité dans le deuxième chapitre de cette partie.

⁸³ D'après le *Online Etymology Dictionary* (Harper, 2001), le verbe *to embed* (dont la formation est évidente) apparaît d'abord, en anglais, comme "a geological term in reference to fossils in rock" (1778). Le sens figuré n'apparaît qu'à partir de 1835 environ.

⁸⁴ Pour le domaine de la télédétection spatiale, le site FranceTerme propose, pour *image map*, la traduction « iconocarte » (*JO* du 22 septembre 2000).

⁸⁵ Il s'agit là d'un état des choses daté.

⁸⁶ Nous n'irons pas plus avant, ici, sur ces questions.

⁸⁷ Voir <http://www.codeguru.com/forum/archive/index.php/t-179262.html>. Soulignons encore que la question n'est pas de *concevoir* un tel script, ce qui suppose une certaine agilité sans être un exploit de la programmation en JavaScript, mais bien de trouver un tel script tout fait *dans la littérature existante*.

⁸⁸ Voir http://www.editeurjavascript.com/scripts/scripts_navigation_3_613.php.

⁸⁹ Voir http://www.editeurjavascript.com/scripts/scripts_autres_2_138.php.

⁹⁰ On verra ci-après l'importance de cette notion. La traduction en français par « profil » de l'anglais *profile* est, dans l'acception visée, relativement récente. Le *Dictionnaire culturel en langue française* (2005) indique la date de 1967 pour l'apparition de *profil* comme calque de l'anglais *profile* dans le sens de « esquisse psychologique (d'une personne), notamment quant aux aptitudes professionnelles », et date de 1998 celle de *profilage* comme calque de l'anglais *profiling*, « fait de dresser le profil psychologique d'un criminel, d'un tueur récidiviste inconnu, pour chercher à l'identifier ». Le nom de *profileur*, calque de l'anglais *profiler*, serait – en ce sens – de 1999.

⁹¹ Pour l'examen détaillé d'un exemple de cookie, voir par exemple Lemainque, 2007, pp. 453-457.

⁹² Cette situation ne se rencontre que trop fréquemment dans les formations relatives à l'Internet, où les formés disposent en règle générale d'ordinateurs dont la « personnalisation » de fait ne permet guère au formateur d'espérer que chacun retrouvera sur son écran ce que le plan de marche de la formation prévoit d'y faire apparaître...

⁹³ Nous laisserons le lecteur découvrir par lui-même si, dans la quatrième partie de ce travail, d'autres domaines pourraient se rajouter aux treize déjà identifiés.

⁹⁴ Dans la préface qu'il a donnée à l'édition de *Die Frage der Laienanalyse* que nous utilisons, Jean-Bertrand Pontalis résume la chose en quelques lignes (Freud, 1926/2003, p. 7) : « Au cours du printemps 1926, Theodor Reik, membre de la Société psychanalytique de Vienne, est l'objet d'une plainte pénale pour exercice illégal de la médecine. Quand s'engage l'instruction du procès, Freud entreprend de rédiger son opusculé. Commencé en juin, celui-ci est chez l'imprimeur en juillet et en librairie dès septembre. On le voit, les choses ne traînent pas. Peut-être grâce à Freud, sans doute aussi par défaut de preuves (la plainte d'un malade, tenu pour peu digne de foi, était à l'origine de l'affaire), le procureur met fin à l'action judiciaire après l'enquête préalable. »

⁹⁵ Ce sont les dialectiques *du parachutiste et du truffier, du sujet et du hors-sujet, des boîtes noires et des boîtes claires, de l'« excription » textuelle et de l'inscription textuelle, de la conjecture et de la preuve* (ou *des médias et des milieux*), enfin *de la production et de la réception* (voir Chevallard, 2007d).

⁹⁶ Dans un article récent (Strand, 2008) intitulé "Google's addiction to cheap electricity", à propos du projet d'installation d'une ferme de serveurs à The Dalles (Oregon), afin de tirer profit du barrage installé non loin de là, l'auteur écrit d'abord : "Every time someone clicks the 'Google Search' button, thousands of servers [...] reel into action." Il précise encore : "A query for *American Idol*, the top search on Google News in 2007, trolls through petabytes of data, using tens of billions of CPU cycles. Velcroed together, stacked in racks, and lined up in back-to-back rows, the servers require a half-watt in cooling for every watt they use in processing..." Conclusion : "Based on a projected industry standard of 500 watts per square foot in 2011, the Dalles plant can be expected to demand about 103 megawatts of electricity—enough to power 82,000 homes, or a city the size of Tacoma, Washington."

⁹⁷ Empruntée au poète latin Horace, la formule a été popularisée par Emmanuel Kant dans son opusculé de 1794, *Réponse à la question : qu'est-ce que les Lumières ? (Beantwortung der Frage: Was ist Aufklärung?)*. Voici un ultime exemple de ce refus des frontières : on n'a mentionné qu'en passant, plus haut, le format JPEG (ainsi que les formats GIF et PNG) ; or on peut légitimement désirer en savoir plus à ce sujet. On pourra se reporter alors, par exemple, à l'étude récente d'Élise Janvresse et Thierry de la Rue (2008), étude où entre un peu de mathématiques, que chacun devrait maîtriser moyennant un effort *limité et spécifique* – qui ne suppose nullement, et de très loin, que l'on soit un « mathématicien ». Le lecteur y comprendra, notamment, pourquoi il vaut mieux ne pas « recadrer plusieurs fois ses photographies » (p. 69).

⁹⁸ Sur la théorie de l'élémentation, nous empruntons l'essentiel à Yves Chevallard, 2001b.

⁹⁹ Voir http://www.ac-nancy-metz.fr/ia54/cgi-bin/fichesb2i/programmation_v2007.asp.

QUATRIÈME PARTIE

**DES SAVOIRS SPÉCIFIQUES DE LA RII ?
BESOINS ET OFFRES PRAXÉOLOGIQUES**

Introduction

*Ne pouvant penser n'importe quoi à tout moment,
nous ne pensons que dans les frontières du discours du moment.
Tout ce que nous croyons savoir est limité à notre insu,
nous n'en voyons pas les limites et ignorons même qu'il y en ait.*

Paul Veyne, 2008

La question de l'émergence et de la diffusion sociales d'une culture de la RII satisfaisant certaines conditions appelle l'examen de la notion de *culture*, telle du moins qu'on peut l'entendre dans le cadre de la TAD. Avant cela, rappelons que, parmi les conditions évoquées ici, figure celle de reposer sur une « science de la RII » dûment « élémentée » et participant elle-même d'une « science du Web », à ce jour encore *in statu nascendi*, qui tire une partie de sa substance de nombreuses sciences affines – telle la science de la RI, que nous invoquerons plus loin. Cela noté, il n'est pas inutile de se livrer à une courte enquête historique sur le mot de culture lui-même. Dans son étude classique et minutieuse du mot, Raymond Williams (1983, p. 88) notait ceci ¹ : “In French, until C18, **culture** was always accompanied by a grammatical form indicating the matter being cultivated, as in the English usage already noted. Its occasional use as an independant noun dates from mC18, rather later than similar occasional uses in English.” Dans son *Dictionnaire de la langue française* (1872-1877), Émile Littré donne ainsi pour exemple « la culture des lettres, des sciences, des beaux-arts » et, citant des auteurs illustres, note ces formulations : « Le peu de connaissance que j'ai, je le dois à la culture des bonnes lettres », « Quoique livré presque uniquement à des études et à des ouvrages ecclésiastiques, il n'avait pas entièrement abandonné la culture des lettres » (D'Alembert), « Si la culture des sciences est nuisible aux qualités guerrières, elle l'est encore plus aux qualités morales » (J.-J. Rousseau). En tout cela, « culture » désigne le fait de *cultiver* (les lettres, les sciences, etc.) : le mot désigne ainsi un *processus*. Mais il désigne aussi le *produit* de ce processus, ce qui résulte du fait que l'on a *cultivé* tel ou tel domaine, que l'on *est cultivé*, que l'on *a une culture* en ce domaine.

Ce qui précède, on l'aura observé, se situe du point de vue de *l'individu*, qui cultive tel ou tel domaine et par cela se cultive en ce domaine. On passe à un point de vue collectif dans l'usage de la notion de culture avec des expressions telles que « culture philosophique », « culture littéraire », « culture scientifique ». En vérité, il faut ici considérer non pas un terme, mais bien deux : il y a le *domaine* que l'on cultive et il y a *l'instance* qui cultive ce domaine – et dont le produit de ce processus constitue *la culture* en ce domaine (littérature, sciences, etc.). En ne retenant que ce second terme, on parlera par exemple de « culture populaire » ou de « culture bourgeoise », etc., sans préciser davantage ; et, plus complètement, et plus clairement aussi sans doute, de « culture populaire en matière de loisirs », de « culture bourgeoise en matière alimentaire », etc., expressions où les *deux* déterminants interviennent à la fois.

Comment définir alors la notion de culture – processus *et* produit – dans le cadre de la TAD ? Le *processus*, on l'a rappelé, est premier ; en ce sens, on dira qu'une culture est un processus dynamique qui, à propos d'un domaine d'activité sociale donné (loisirs, alimentation, formation, RII, etc.), engendre un système de questions *et* de réponses relatives à ce domaine – réponses qui sont, rappelons-le, des praxéologies ou des fragments de praxéologies. Le *produit* d'un tel processus à un instant donné, soit une certaine culture du domaine d'activité considéré, est ainsi un système plus ou moins intégré, plus ou moins organisé de questions posées, de réponses apportées ou pendantes, de praxéologies (ou de fragments de praxéologies), avec leurs techniques parfois incorporées en habitus, leurs technologies et, plus largement, leur *logos* parfois naturalisé en « évidences » de « la » culture, c'est-à-dire de *cette* culture-là. Le *processus* culturel, c'est-à-dire le processus de formation ou de réformation d'une culture, a un « siège », qui est un système plus ou moins intégré d'*institutions* et constitue l'instance qui, à travers l'action plus spécifique de certaines de ces institutions (ou encore par le truchement de créations institutionnelles *ad hoc*), cultive le domaine d'activité en question.

Un tel processus suppose que des questions soient posées et des réponses cherchées et proposées. Reprenant des notions plus ou moins anciennes, mises en œuvre dans divers travaux inspirés par la TAD, nous appellerons *forum* (de questions) relatif au domaine considéré tout dispositif institutionnel (résultant parfois du « bricolage » d'une institution existante) dont l'une des raisons d'être est d'offrir un lieu où, dans ce domaine, des questions puissent être énoncées, entendues, enregistrées. Nous nommerons de façon également générique *atelier* (de fabrication de réponses) tout dispositif institutionnel où l'on travaille de manière méthodique à élaborer des réponses à des questions relatives à un certain domaine d'activité. Enfin, nous désignerons sous le nom de *boutique* (de réponses) dans le domaine

considéré tout dispositif institutionnel qui diffuse auprès de certains publics des réponses à des questions relatives au domaine considéré.

La notion de forum (de questions, dans un certain domaine) a été notamment mise en œuvre dans la formation des élèves professeurs de mathématiques à l'IUFM d'Aix-Marseille : sur ce sujet, voir Cirade, 2006. Pour être adéquatement utilisé ici, le mot de boutique doit être allégé de ses surcharges sémantiques « modernes » : issu du grec apothêkê, « lieu de dépôt », « magasin (de vivres, d'approvisionnements) », ce vocable désigne un lieu où l'on peut s'approvisionner en tel ou tel « bien » – où l'on peut par exemple tenter de mettre son équipement praxéologique en meilleure adéquation avec ses besoins propres. En ce sens ou en un sens voisin, le mot est employé dans l'expression « boutique des sciences » : on pourra se reporter à cet égard à l'article « Les Boutiques de sciences » (s. d.) de l'Espace des citoyens / santé et bien-être du Québec. Dans le cadre de la TAD, le mot a été notamment utilisé à propos d'un dispositif – la « boutique de mathématiques » – dont on trouvera une courte description dans Chevallard, Bosch et Gascón, 1997, p. 51. Selon Y. Chevallard (communication personnelle, 16 mars 2007), le mot d'ouvroir a pu être employé quelquefois (avec une nuance d'ironie) pour désigner un atelier.

Même si son acception est plus large – il désigne une *fonction* plutôt que tel type particulier de dispositifs – le mot de *forum* a, dans le cadre de la TAD, un sens qui s'accorde spontanément avec l'emploi usuel qui en est fait aujourd'hui dans le monde Internet². Un forum est un lieu où une question se transmue, localement au moins, en un *problème*. Un problème, c'est, étymologiquement, une difficulté « lancée devant » une communauté afin que celle-ci tente de lui apporter une solution (le grec *problêma* est formé de *pro*, « devant », et *ballein* « jeter, lancer »). Par le fait d'être « exposée » devant un collectif – les participants au forum –, une question, qui exprime au départ une difficulté éprouvée subjectivement par une personne, s'objective en ce qui devient alors, pour le collectif, un problème. D'une manière générale, l'étude de la dynamique d'une culture suppose que l'on recherche ce que sont les forums de questions où sont énoncées et enregistrées les questions Q à étudier comme autant de problèmes, les ateliers où cette étude sera conduite et où, pour cela, seront apportées et « exploitées » des œuvres diverses, et notamment des réponses R^\diamond , afin d'élaborer des réponses R^\heartsuit , enfin les boutiques de réponses où ces réponses R^\heartsuit seront proposées, illustrées, défendues³.

Comment préciser dans le cadre ainsi tracé la problématique de la « culture de la RII » qui est au centre de notre travail ? En premier lieu, le domaine d'activité sociale à laquelle une telle culture se rapporte est donc celui de la RII, la recherche d'information sur Internet. En second lieu, il conviendrait de préciser l'instance qui est le siège de la dynamique culturelle considérée. Ici, de multiples réponses semblent *a priori* possibles. La culture de la RII qui constitue l'horizon de ce travail doit avoir pour siège, si l'on peut dire, la société tout entière, et concerner en puissance chacun de ses membres – ce que l'on peut appeler, pour faire court, une culture *citoyenne* de la RII. Dans ce but, nous avons commencé de dessiner, dans cette quatrième partie de notre travail, les contours de ce qui serait une culture *mère de toutes les autres*, en ce sens que toute culture de la RII pourrait y être engendrée. Pour cela, nous nous sommes tournée vers diverses institutions porteuses de cultures de la RII, et tout d'abord vers ces institutions qui se préoccupent de la culture adressée à *l'utilisateur ordinaire*, culture que nous ne saurions perdre de vue sans dénaturer le projet dans lequel s'inscrit la présente recherche, dont la diffusion est assurée en grande partie par ce qu'on pourrait appeler une *autodidaxie partagée*, laquelle se concrétise notamment dans ce fait que d'aucuns, qui ne sont pas des spécialistes et moins encore des « gourous » de la RII, s'il en existe, n'hésitent pas à « ouvrir boutique », en proposant livres ou sites Web portant (en totalité ou en partie) sur la RII. Cette culture-là sera présente notamment dans le premier des trois chapitres qui composent cette partie de notre travail. Mais pour des raisons tenant aux exigences de notre méthodologie de recherche, nous ferons ici une place plus importante d'une part aux institutions porteuses d'une culture de la RII formée au voisinage immédiat de la *computer science* et de l'intelligence artificielle (Witten, Gori & Numerico, 2006 ; Boughanem & Savoy, 2008), d'autre part aux institutions porteuses de cette culture de la RII qui s'est développée dans le monde de la documentation et des bibliothèques (Mesguich & Thomas, 2006a ; Notess, 2006). Chez les premières, nous apercevrons ce qui pourrait constituer le *fondement* des cultures de la RII ; chez les secondes, nous trouverons des variantes multiples d'un texte du savoir en formation visant, au-delà de « l'utilisateur ordinaire », l'utilisateur « professionnel », qu'il appartienne à la catégorie des documentalistes, à celle des journalistes, ou des universitaires, etc. (Lenormand, 2007).

Comment connaître une culture de la RII donnée ? Comment éprouver l'*offre praxéologique* qu'elle renferme ? Filons ici la métaphore à laquelle invite le mot de boutique (de réponses). Imaginons que l'on ait identifié une boutique où s'exposent et se proposent – de manière non nécessairement exclusive – un choix de praxéologies appartenant à cette culture ; pénétrons-y. Une première stratégie consiste à examiner, comme il en irait dans un inventaire, l'ensemble des produits praxéologiques qui s'y trouvent, à en admirer le cas

échéant la diversité, la profusion, le caractère précieux – ou au contraire à constater la pauvreté de l'offre. Il s'agit là d'une stratégie de base, indispensable pour se faire une idée assez poussée de ce que la culture correspondante recèle. C'est la stratégie que nous avons mise en œuvre jusqu'ici. Nous l'utiliserons à nouveau, brièvement, dans le premier chapitre de cette quatrième partie de notre travail pour montrer son rendement sur un exposé qui, participant de la culture de l'utilisateur « ordinaire », et en dépit de sa concision assumée, offre à son lecteur un vaste panorama praxéologique, au point, on le verra, de lui faire rencontrer de nouvelles frontières, que notre analyse des exposés déjà examinés n'avait pas identifiées. Mais ce type de lecture (ou de visite), que nous nommerons (au prix d'un quasi-néologisme) lecture *inventoriante* d'un exposé (ou visite inventoriante d'une boutique) peut être fortement trompeuse quand il s'agit d'apprécier la richesse praxéologique de cet exposé et, plus largement, à travers elle, l'offre praxéologique portée par la culture dont cet exposé participe. Pour savoir ce qu'un livre contient, ainsi, il ne suffit pas de le lire d'une façon linéairement, attentivement énumérative – « inventoriante ». Il faut *aussi* le questionner, le livrer à une lecture *questionnante*. Soit la question que voici : « Comment obtenir qu'un moteur de recherche retourne des pages contenant un certain mot – que l'on saisit dans la boîte de requête – ou, aussi bien, des *synonymes* de ce mot ? » L'exposé que l'on a là devant soi apporte-t-il une réponse à cette question ? Ou bien sa richesse apparente cache-t-elle une réponse vide sur ce point ? (Notons que l'inverse peut se produire : un exposé resserré peut receler des ressources insoupçonnées.) Tel est le critère essentiel.

La difficulté à mettre en œuvre ce critère tient sans doute à deux facteurs qui unissent leurs effets. En premier lieu, il semble qu'existe indéniablement un habitus qui nous porte à nous soumettre à l'ordre explicite d'un texte, en suivant sagement le parcours qui s'y trouve, si l'on peut dire, tracé à l'avance. À l'instar d'Augustin d'Hippone (354-430), tout se passe pour nous comme si nous entendions une voix nous dire irrésistiblement : *Tolle, legge !* – prends et lis⁴. Mais un autre obstacle surtout fait barrage à une lecture questionnante, qui tient tout simplement à ce fait que, ordinairement, *nous n'avons pas de questions à poser au texte que nous lisons*, en sorte que, même si l'on en suit scrupuleusement le propos, on ne pourra guère, ensuite, l'interroger explicitement.

Le problème qui se pose à nous est donc celui *de l'absence ou de la pénurie de questions* de RII. Pour aller au-delà de lectures guidées, organisées entièrement par le texte que l'on lit, il faut poser à ce texte des questions. De telles questions, il est vrai, doivent satisfaire divers réquisits. Si l'on considère par exemple une personne qui débute en matière de RII, l'offre praxéologique qui l'intéresse est d'abord celle qui répondra aux questions qui vont se poser *à elle* assez fort pour qu'elle finisse bientôt par *se les poser* et par chercher à

leur apporter des réponses en explorant la culture de la RII où elle se trouve pour un temps cantonnée. Bien entendu, il n'y a aucune raison pour ne pas généraliser cette situation : l'exploration de l'offre praxéologique, notamment par la lecture questionnante d'exposés, doit pouvoir concerner tout type d'utilisateurs et toute culture de la RII. Où trouver de telles questions pour interroger l'offre praxéologique en la matière ? La réponse est en un sens fort simple. Tout texte porte en lui ces questions auxquelles les assertions ou les présuppositions qu'il contient peuvent être regardées comme constituant des (fragments de) réponses – sans pour cela offrir toujours des réponses pleinement développées. De cette remarque, que nous expliciterons plus avant dans le deuxième chapitre de cette partie, nous avons tiré de longues listes de questions qui donneront déjà une idée de ce que à quoi une culture-mère de la RII devrait pouvoir apporter réponse : une culture ne se décrit pas seulement par un inventaire praxéologique, mais aussi par le tableau des questions auxquelles elle se soucie de répondre.

La technique de lecture questionnante d'un exposé doit ainsi être « armée » de questions. Elle permet alors de voir si tel exposé « résiste » au questionnement qu'on lui impose. Plus généralement, on peut tenter de déterminer si une culture de la RII – qu'on peut, avec une bonne approximation, identifier à un certain corpus d'exposés – résiste ou non à tel ou tel questionnement. La problématique qui sera la nôtre ici est pourtant légèrement différente : elle vise à apprécier si telle culture de la RII permet à qui se pose tel ensemble de questions Q de trouver aisément des réponses R^\diamond à ces questions, ou encore d'y trouver telles réponses R^\diamond déterminées, sans parler encore de trouver de « bonnes » réponses R^\diamond . Si, par exemple, on se pose la question « des synonymes » évoquée plus haut, lui trouvera-t-on une réponse dans tel corpus documentaire déterminé – fait d'exposés imprimés ou en ligne – que telle catégorie d'utilisateurs a pris l'habitude de fréquenter ? En l'espèce, la réponse consistant à utiliser, avec le moteur de recherche Google, l'opérateur tilde (~), réponse que David Vise et Mark Malseed (2006) présentent comme un « truc » simple et utile (ainsi que nous l'avons vu dans notre *Introduction générale*), ne se rencontre par exemple ni dans l'ouvrage de Patrick Lenormand (2007), qui, sous un titre avantageux – *Google : les secrets d'un champion* –, consacre tout de même à Google quelque quarante pages (pp. 30-70), ni dans le livre de Véronique Mesguich et Armelle Thomas (2006a), qui offre pourtant à son lecteur, sous le titre « Et les synonymes ? » (pp. 74-75), un riche développement sur la question posée. À cet égard, on distinguera deux notions proches mais non superposables : celle de l'accessibilité *culturelle* d'une réponse, et celle de son accessibilité *technique*. Tout semble indiquer par exemple que, dans le cadre français, la réponse mentionnant l'opérateur tilde a une accessibilité culturelle faible, sans doute parce qu'elle ne fonctionne guère qu'en anglais, cette frontière. En revanche, son accessibilité technique ne présente guère de difficulté : en tapant

dans la boîte de recherche de Google la requête **Google synonyms**, on obtient parmi les tout premiers résultats un document qui fournit la réponse cherchée⁵. Dans quelques contextes choisis, notre dernier chapitre aura pour objet d'explorer l'un et l'autre type d'accessibilité praxéologique.

CHAPITRE 15

La recherche d'information sur Internet : un exposé encore

Dans la deuxième partie de ce travail, nous avons vu ce que peuvent être le contenu et son organisation dans des exposés « ordinaires » sur la RII, exposés que bornent ou traversent de multiples frontières, quand ils ne se situent pas très en deçà de certaines d'entre elles. Nous nous arrêterons ici sur un ultime exemple, celui d'un ouvrage plusieurs fois mentionné dans la partie précédente : *Les bons usages d'Internet* d'Isabelle Bréda et Carole Hourt (2007). Son chapitre II, intitulé « Se documenter, s'informer », appartient typiquement à l'ensemble Σ des exposés sur la RII. Une introduction – en italique dans le texte – présente d'abord la notion de *Web*, nous l'avons vu, en attirant l'attention, selon une opposition classique mais contestable, entre ce qui pourrait être « une gigantesque bibliothèque universelle », dotée d'un catalogue exhaustif fabriqué par des humains, où le bibliothécaire serait un médiateur « entre un lecteur et les documents dont il a besoin », et la réalité du Web, où s'observent « un foisonnement d'informations hétérogènes et diverses par leur qualité », en sorte que le Web ne saurait être regardé comme « une bibliothèque au sens classique du terme », bien que il soit « un formidable gisement d'informations » (p. 38).

Un exposé tout simple

Le sommaire du chapitre permet de se faire une idée de la matière qui, d'une façon toujours concise mais attentive, s'y trouve abordée ; le voici.

Faire une recherche efficace : moteurs et annuaires

ANNUAIRE, MOTEUR, MÉTAMOTEUR : QUEL OUTIL UTILISER ?

POURQUOI CES OUTILS SONT-ILS GRATUITS ?

FAUT-IL SE MÉFIER DE GOOGLE ?

COMMENT INTERROGER UN ANNUAIRE ?

COMMENT INTERROGER UN MOTEUR ?

QUAND FAUT-IL UTILISER LA RECHERCHE AVANCÉE ?

QUE FAIRE EN L'ABSENCE DE RÉSULTATS PERTINENTS ?

Sélectionner, valider et utiliser l'information : les bons réflexes

L'ORDRE D'APPARITION DES RÉSULTATS EST-IL IMPORTANT ?

QUELLE EST LA STRATÉGIE DES ÉDITEURS DE SITES POUR ÊTRE BIEN RÉFÉRENCÉS ?

COMMENT DÉCRYPTER UNE ADRESSE URL ?

COMMENT S'ASSURER DE LA FIABILITÉ D'UNE INFORMATION ?

FAUT-IL SE PASSER DU COPIER-COLLER ?

Se tenir informé : les médias en ligne

DU « TRADITIONNEL » AU « CITOYEN », QUELS SONT LES MÉDIAS EN LIGNE ?

QU'EST-CE QU'UNE NEWSLETTER ?

COMMENT CONSTRUIRE SON ACTUALITÉ SUR MESURE AVEC LES FILS RSS ?

QU'EST-CE QUE LE PODCASTING ?

DEVIENDRONS-NOUS TOUS JOURNALISTES ?

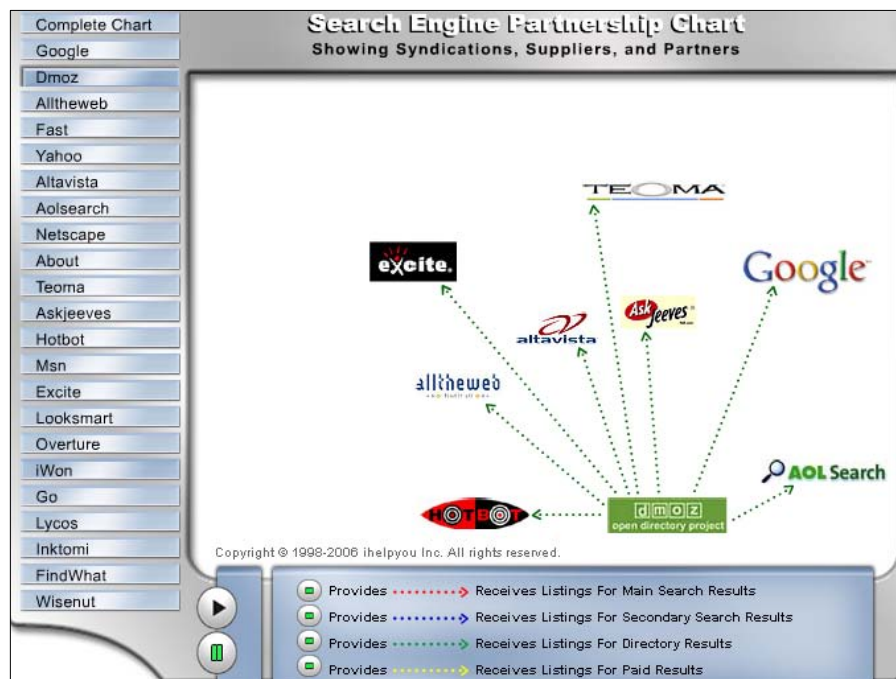
Nous ne ferons qu'un petit nombre de remarques sur l'ensemble du parcours d'étude ainsi organisé.

Annuaire, moteurs, métamoteurs

À propos des annuaires, d'abord, il est indiqué que, « pour chaque site retenu, les “cyberdocumentalistes” rédigent un bref descriptif de son contenu, et le classent dans une ou plusieurs catégories ». Même assortie de guillemets, la catégorie des « cyberdocumentalistes » évoquée ici rapproche un peu inexactement deux histoires solidaires sans doute mais distinctes. Sur le Web, la saga des annuaires généralistes et publics semble aujourd'hui en voie de se terminer. Le plus célèbre d'entre eux, DMOZ, qui alimente plusieurs outils de recherche connus – on se reportera là-dessus à la copie d'écran ci-après ⁶ –, et qui, consulté le 9 juillet 2008, indique répertorier 4 600 733 sites à l'aide de plus 590 000 catégories, est bâti par des *editors* dont le nombre, le même jour, s'élevait à 80 366. Or ces *editors*, choisis avec soin, le sont pourtant sans relation formelle avec ce qu'on appellerait en France le monde de la documentation. À la question “Are there requirements for being an editor?”, le site de l'Open Directory Project (ODP) répond en ces termes ⁷.

Everyone is welcome to join the ODP. All you need is an interest or passion and a computer. While there are no specific pre-requisites, we seek people who have a genuine interest in building a directory that is free of commercial interests and favoritism. Fairness and objectivity prevail here.

Those who still believe the Web should be free and accessible to all, without bias and unnecessary noise, will most enjoy the ODP experience. Potential editors should demonstrate a keen eye for spotting quality and useful sites, attention to detail, and possess good grammar, spelling and communication skills.



D'un autre côté le métier de cyberdocumentaliste est apparu récemment ; la définition suivante a pu en être proposée ⁸.

On retrouve des cyberdocumentalistes au sein de différents types de sociétés. Chez celles qui développent des moteurs de recherche, pour commencer : là, ils ont pour mission de faciliter le surf des internautes et de leur faire gagner du temps. Avant tout, il s'agit pour eux de sélectionner les sites dont le contenu est le plus intéressant dans des catégories données, le logement, l'emploi, le sport, la musique... Mais on trouve également des cyberdocumentalistes dans des cabinets spécialisés dans la veille. Dans ce cadre, ils travaillent pour des clients et sont donc chargés de réaliser des recherches ou une veille sur des sujets précis, en rapport avec un secteur d'activité, une technologie... Après l'activité de recherche, un cyberdocumentaliste doit hiérarchiser les informations et organiser les données pour en faciliter la consultation. L'organisation avant tout. Évidemment, de la rigueur, d'excellentes facultés d'organisation et de classement s'imposent. Mais une bonne dose de débrouillardise, de la curiosité et de bonnes capacités rédactionnelles sont également requises.

Les auteures ont donc utilisé un raccourci qui ramène les annuaires existants sur le Web dans le domaine de la documentation – où d’ailleurs les cyberdocumentalistes d’entreprise ne se situent que partiellement.

Ayant présenté rapidement la notion de moteur, évoqué l’indexation du contenu du Web à l’aide d’un robot qui voyage de lien en lien, « mémorisant au passage chaque page visitée », et indiqué (de façon un peu inexacte, nous le verrons) que, lorsqu’on formule une requête sur un moteur, « il fournit les liens vers toutes les pages contenant les mots demandés », les auteures concluent que de nombreux outils seraient désormais « mixtes, à la fois moteurs et annuaires », ce qui sur-valorise, nous semble-t-il, la place et le rôle actuels des annuaires. Au demeurant, l’opposition entre annuaire et moteur, qui présente le moteur de recherche comme la pure création d’un robot logiciel, est partielle et peut-être partielle. Nombre d’administrateurs de sites, par exemple, se signalent expressément à Google afin que leur site soit répertorié plus sûrement dans l’index de ce moteur de recherche, qui propose d’ailleurs une procédure officielle explicitée sur la copie d’écran ci-après ; et il en va de même d’un certain nombre d’autres moteurs de recherche tel Exalead⁹.

Ajouter l'URL de votre site à Google

Indiquez-nous l'adresse de votre site Web.

Chaque fois que nous explorons le Web, nous ajoutons et mettons à jour de nouveaux sites dans notre index et nous vous invitons à utiliser cette page pour nous soumettre l'URL de votre site. Nous n'ajoutons pas dans notre index toutes les URL qui nous parviennent et nous ne pouvons ni prévoir, ni garantir le moment ou l'éventualité de leur inclusion.


Entrez l'adresse URL complète de votre site, y compris le préfixe `http://`. Exemple : `http://www.google.fr/`. Vous pouvez également ajouter des commentaires ou des mots clés décrivant le contenu de votre site. Ils sont uniquement utilisés pour information et n'ont pas d'incidence sur l'indexation ou l'utilisation de votre site par Google.

Remarque : Seule la page de premier niveau d'un hôte étant nécessaire, vous n'avez pas besoin de soumettre chaque page une à une. Notre robot d'exploration, Googlebot, se chargera de trouver le reste. Google mettant régulièrement à jour son index, il n'est pas nécessaire d'indiquer les liens mis à jour ou obsolètes. Les liens désactivés disparaissent de notre index au cours de l'exploration suivante, lorsque nous mettons à jour l'ensemble de l'index.

URL :

Commentaires :

Facultatif : Pour nous aider à distinguer les URL indiquées manuellement de celles soumises automatiquement, entrez le mot tel qu'il apparaît dans la zone ci-dessous.



Après avoir défini la notion de métamoteur et souligné sans l'expliciter le lieu commun selon lequel un métamoteur « ne permet pas de formuler des requêtes très fines »¹⁰, le texte renvoie à un « panorama complet (*sic*) des outils de recherche » en ligne sur le site de l'ADBS, l'association des professionnels de l'information et de la documentation. Notons ici que sous la rubrique « Les annuaires généralistes », ce panorama propose¹¹, outre l'Open Directory, seulement deux autres « annuaires », Looksmart et Snap, que nous laisserons le lecteur découvrir par lui-même, s'il le souhaite¹². Quant aux auteures, elles citent comme annuaires généralistes Yahoo! et Nomade ; comme moteurs Google, MSN, Exalead ; comme métamoteur KartOO, qui semble être goûté pour sa particularité de « présenter ses résultats sous la forme de cartes ».

L'article "Yahoo!" de l'encyclopédie Wikipedia précise : "On February 18, 2004, Yahoo! dropped Google-powered results and returned to using its own technology to provide search results." L'article "Yahoo! Directory" de la même encyclopédie décrit ainsi la situation qui en est résultée : "Yahoo! provides both a search engine and a directory service, and the directory is searchable separately from the rest of their search engine results." (Sur l'histoire du moteur de recherche de Yahoo!, on pourra voir l'article "Yahoo! Search" de Wikipedia.) Il est ainsi curieux que, écrivant en 2006, les auteures ne présentent ici Yahoo! que comme un annuaire. Rendant compte avec retard de la situation nouvellement créée, un article paru le 14 mai 2004 dans le quotidien Le Monde sous le titre « La bataille entre Google, Microsoft et Yahoo ! est lancée » (Leser, 2004) voyait alors en Yahoo! nouvelle manière « la menace la plus immédiate » pour Google – prédiction que les chiffres n'ont pas cessé de démentir depuis. Si, en effet, on suit le baromètre Xiti des moteurs de recherche (<http://barometre.secrets2moteurs.com/>), en août 2006 (période pendant laquelle les auteures ont relevé les adresses URL et les copies d'écran de page Web figurant dans leur ouvrage), les cinq moteurs de recherche les plus utilisés étaient Google (85,82 %), Yahoo! (4,63 %), Voila (3,8 %), Live Search [anciennement MSN Search] (2,88 %), AOL (0,98 %). Pour le mois de juin 2008, la même source indique le classement suivant des dix premiers moteurs de recherche les plus utilisés : Google (90,61 %), Yahoo! (2,85 %), Live Search (2,35 %), AOL (1,54 %), Orange (1,35 %), Free (0,43%), Alice (0,32 %), Ask (0,14 %), Altavista (0,06 %), Exalead (0,05 %). On voit donc, en passant, que le choix de moteurs proposé par les auteures est sans doute inspiré par d'autres critères que celui de la « popularité ».

La sous-section intitulée « Pourquoi ces outils sont-ils gratuits ? » est brève : elle évoque en quelques lignes le modèle économique fondé sur la publicité. Ce modèle est expliqué sommairement, sans que soit mentionnée la notion de lien sponsorisé (qui apparaîtra plus loin dans l'exposé) ni *a fortiori* le système du *pay per click*. Les auteures évoquent la vente d'espaces publicitaires par les outils de recherche « aux entreprises soucieuses de leur visibilité sur le Web », ce qui est un peu restrictif (un parti politique, qui n'est pas une entreprise, peut fort bien acheter de tels espaces). Elles insistent sur le fait que l'interface de certains outils de recherche est en fait un *portail*, « site “repère”, à partir duquel on va pouvoir se rendre sur d'autres sites », le but stratégique étant que cette interface enrichie « devienne la page d'accueil du plus grand nombre possible d'internautes, afin d'attirer des annonceurs publicitaires et, en conséquence, des revenus ». On notera ici que l'exception à cette loi que constitue Google, dont la page d'accueil est, depuis l'origine, connue pour son minimalisme, n'est pas mentionnée (ce qui est consonant sans doute avec le silence sur les liens sponsorisés).

Google, toujours lui !

C'est pourtant à Google que s'intéresse la sous-section qui suit, intitulée « Faut-il se méfier de Google ? ». Sans doute cette question est-elle inévitable, étant donné les craintes exprimées à l'envi, ici ou là, nous le savons, touchant l'existence et le fonctionnement de Google. Mais cette atmosphère anti-Google dominante ne rend pas facile l'élaboration d'une réponse, soumise à la pression d'éléments polémiques toujours les mêmes, et que les auteures vont reprendre sans façon. Le premier élément de la polémique présente Google comme une entreprise gloutonne, diversifiant à l'infini ses activités – en absorbant au passage une kyrielle d'entreprises jusque-là indépendantes. Un deuxième argument est celui de la propension qu'aurait Google à ne pas respecter le droit d'auteur : sur ce point, le développement porte essentiellement sur YouTube, racheté par Google en 2006. À cet égard, notent les auteures, Google n'est pas seul en cause, car, notent-elles, « tous les sites de partage de contenus sont concernés par la question de la propriété intellectuelle ». Le troisième argument est celui de l'atteinte supposée à la *privacy* dont Google se rendrait coupable en collectant des « données personnelles » concernant ses utilisateurs.

La rhétorique du texte examiné balance entre affirmations génériques et allégations visant spécifiquement Google (de façon directe ou allusive), comme dans l'assertion suivante : « Ainsi, comme tous les moteurs de recherche, Google stocke sur ses – très

puissants – serveurs l’ensemble des requêtes effectuées. » Ici c’est seulement l’allusion à la (mystérieuse) surpuissance des serveurs de Google qui peut évoquer pour le lecteur peu informé quelque monstruosité menaçante. Ce balancement cède pourtant dans la phrase suivante, qui indique que, lorsqu’un internaute sollicite le moteur de recherche Google, il active du même coup « des cookies qui permettent à l’outil de suivre ses déplacements sur la Toile » – ce qui, pour ce qui est des cookies, n’est certes pas une exclusivité de Google, et, s’agissant des déplacements sur la Toile de l’internaute, doit évidemment être restreint à ceux dont Google a à connaître, par exemple le déplacement vers un site en suivant un lien proposé par Google comme réponse à une certaine requête. La suite de cette sous-section est de la même eau : on y trouve encore mentionné le fait que Gmail – le service de messagerie de Google – conserve des copies des courriels qu’il transmet. Comme on le voit en d’autres exposés, un privilège est ainsi accordé aux seuls arguments anti-Google. Il semble que, peut-être pour des raisons éditoriales, l’instruction annoncée – Faut-il se méfier de Google ? – ne puisse être conduite à charge *et à décharge*. On est loin alors d’un exposé complet des faits, notamment parce qu’on semble faire litière des possibilités offertes à l’internaute par Google, en l’espèce.

Rechercher une information

La sous-section « Comment interroger un annuaire ? » est brève : elle préconise l’exploration en allant du plus large au plus précis dans la nomenclature des catégories. Bien que celles-ci varient d’un annuaire à l’autre, y lit-on, « on en devient rapidement familier ». En outre, nombre d’annuaires comportent un moteur de recherche qui « permet de chercher des mots-clés dans les fiches descriptives rédigées par les cyberdocumentalistes ». La question « Comment interroger un moteur ? », qui suit, donne lieu à un développement plus long, d’une page environ. Un premier paragraphe commence par évoquer la notion de recherche « en texte intégral » et mentionne alors le caractère asémantique de cette recherche, qui peut dès lors être mitée par des synonymies, homonymies, et polysémies « qui nuisent à la pertinence des résultats ». Pour s’en défaire, les auteures préconisent d’utiliser plusieurs mots clés combinés à l’aide d’opérateurs logiques (l’adjectif « booléen » n’apparaît pas) : sont cités les opérateurs ET, OU, SAUF. Il est précisé que le premier est implicite, tandis que les deux suivants s’écrivent sur plusieurs moteurs OR pour l’un et sous la forme d’un simple signe *moins* (« - ») pour l’autre. Ces indications sont complétées par une mise en garde concernant l’accentuation et l’emploi des minuscules ou majuscules, qui peuvent être soumises à des règles variables selon le moteur de recherche considéré.

Tout cela est en quelque sorte repris autrement dans la sous-section suivante, « Quand faut-il utiliser la recherche avancée ? ». Grâce aux options qu'il offre, le « formulaire » de recherche avancée (distingué du formulaire de recherche « standard ») permet de contourner l'écriture symbolique des opérateurs logiques, qui varie de moteur à moteur. À cette avantage formel s'ajoutent des avantages fonctionnels, que les auteurs présentent, dans leur généralité, comme « utiles pour limiter les résultats », formulation dans laquelle on peut entendre un thème classique – celui de la *pléthore* de résultats proposés par les moteurs. Cette « limitation », qui est en fait une sélection dans l'ensemble des résultats, peut se faire selon divers critères : la combinaison de mots clés constituant la requête peut ainsi être recherchée uniquement dans les fichiers au format pdf, ou dans les pages des sites du gouvernement français (il faut pour cela imposer à l'URL de contenir la chaîne de caractères .gouv.fr), ou encore dans des documents datés de moins de six mois par exemple ¹³.

La sous-section qui suit – et qui clôt la section intitulée *Faire une recherche efficace : moteurs et annuaires* – aborde une question clé : « Que faire en l'absence de résultats pertinents ? » La stratégie conseillée comporte plusieurs composantes, telle la pratique persévérante consistant, si l'on n'aboutit pas du premier coup, à « procéder par tâtonnements », tout en n'oubliant pas de contrôler l'orthographe de la requête. Mais la tactique essentielle semble tenir dans l'exploitation habile d'une opposition binaire des plus ordinaires : il faudrait tantôt « élargir la requête », tantôt « restreindre la requête ». Dans le premier cas, ainsi, on pourra introduire dans la requête, à l'aide de l'opérateur de disjonction, des synonymes des mots clés déjà choisis. Dans le second cas, on pourra par exemple ajouter des mots clés, et, au-delà de l'utilisation des fonctionnalités de la recherche avancée déjà mentionnées, utiliser les guillemets pour requérir telle ou telle « expression exacte ». Au-delà de ces manœuvres, l'internaute est invité à changer de « famille d'outils », car aucun outil n'indexe la totalité du Web. L'épilogue nous ramène au « monde d'avant » : il arrive, y lit-on, que les informations que l'on ne parvient pas à dénicher sur le Web se trouvent « dans un livre, une revue, un film... ».

Au-delà des apparences : ordre des résultats et référencement

La deuxième section de l'exposé est intitulée, on l'a vu, *Sélectionner, valider et utiliser l'information : les bons réflexes*. La première question abordée, rappelons le aussi, est la suivante : « L'ordre d'apparition des résultats est-il important ? » La réponse apportée indique que, en principe, les pages affichées parmi les premières sont celles où le mot clé apparaît « le plus grand nombre de fois », à moins qu'il ne figure dans « le titre ». Dans le cas où la requête

est composée de plusieurs mots clés, les pages privilégiées sont, de même, celles « où ces mots sont proches ». À nouveau, Google a droit à un traitement à part, toujours un peu négatif, comme on le verra en ce passage.

Google a par ailleurs établi un critère de popularité des sites qui a largement contribué à son succès : plus un site est pointé par d'autres avec un lien hypertexte, mieux il est répertorié par le moteur.

C'est la prime à la popularité, mais pas forcément à la pertinence !

La référence précédente rapproche *pertinence* et *popularité*, ce qui n'est pas la même chose ¹⁴.

La question qui donne son titre à la sous-section suivante, n'est évidemment pas sans lien avec ce qui précède : « Quelle est la stratégie des éditeurs de sites pour être bien référencés ? » La réponse apportée, là encore, montre à l'endroit des pratiques du Web, une attitude de suspicion – dont le mobile est peut-être seulement « pédagogique » – qui s'exprime en autant de mises en garde à l'adresse du lecteur. Ce dernier est ainsi informé de ce que certains éditeurs de sites tentent de figurer en bonne place dans les résultats fournis par les moteurs de recherche à propos de certains mots clés en multipliant artificiellement les occurrences de ces mots, non dans le texte que peut lire l'internaute qui visite la page, mais dans ce qu'on appelle le code source. En ce point, l'exposé rencontre une frontière dont nous avons déjà souligné l'importance dans la partie précédente de ce travail : celle de la programmation Web. Les auteures, en effet, éprouvent très normalement le besoin de préciser ce qu'est le code source. Elles ne le feront pas dans le corps du texte mais dans une note de bas de page, que nous reproduisons ci-après ¹⁵.

Code source : suite d'instructions écrites par un informaticien pour créer un programme. Le langage de programmation, compréhensible des seuls spécialistes, est interprété par l'ordinateur, pour afficher une interface lisible par l'utilisateur final. À l'origine d'une page Web, il y a un code que l'on peut voir en sélectionnant la fonction « Source » du menu « Affichage » de son navigateur.

Bien entendu, cette note ne fait guère qu'informer le lecteur profane de l'existence d'une frontière dont il pourra découvrir les marques concrètes en procédant à la manœuvre indiquée – même s'il risque bien sûr de ne rien entendre au code source que le navigateur affichera alors. Il n'en reste pas moins que l'on a là, dans un exposé des plus « resserrés », une indication qu'on ne trouve pas fréquemment dans les présentations ordinaires de la RII telle que la deuxième partie de notre travail en a explicité le contenu.

Les auteures évoquent ensuite la pratique des mots clés payants, qui se traduit sur les pages de résultats par la présence de liens dits commerciaux ou « sponsorisés », lesquels « apparaissent à côté ou au-dessus des résultats issus de l'index du moteur ». Ce développement se poursuit en petits caractères : autour de la reproduction de la première page de résultats de Google en réponse à la requête **soutien scolaire en ligne**, on y précise d'une part le principe des annonces publicitaires, d'autre part le contenu de l'affichage d'un résultat naturel tel que Google le dispose. (L'adjectif *naturel* n'est pas utilisé par les auteures, qui parlent plus explicitement de « résultats issus de l'index de Google ».) Nous ne commenterons pas ici la tentative d'explicitation du fonctionnement des liens publicitaires, qui pose des problèmes subtils ¹⁶. Nous nous arrêterons en revanche un instant sur la présentation qui est faite des liens naturels. De manière un peu surprenante, les auteures les présentent ici comme « classés par ordre de pertinence », comme si elles validaient l'ordre proposé par Google. La description de ce qui apparaît sur la page mentionne d'abord le *titre* (par exemple, sur la copie d'écran ci-après, « Cyberpapy »), qui est, lit-on, « celui qu'a choisi le concepteur du site ».

[cyberpapy](#)
 Cyberpapy, le premier site gratuit de **soutien scolaire** entre generations, vos forums : aide aux devoirs et comment trouver un bon bouquin.
[www.cyberpapy.com/](#) - 13k - [En cache](#) - [Pages similaires](#)

Les auteures mentionnent ensuite l'URL, ici <http://www.cyberpapy.com/>. On notera que le commentaire ignore les liens [En cache](#) et [Pages similaires](#). Les bribes de texte figurant entre la ligne de titre et la ligne portant l'adresse de la page – soit ici : « Cyberpapy, le premier site gratuit de soutien scolaire entre générations, vos forums : aide aux devoirs et comment trouver un bon bouquin » – sont décrites par les auteures comme constituant « un bref descriptif », qui se trouverait « généré automatiquement par le moteur à partir du contenu de la page Web ». Ces notations appellent en fait plusieurs remarques. Ce que nous avons désigné comme des bribes de texte ne peut être tenu pour un descriptif de la page qu'à la condition de donner à ce mot un sens très large : ainsi qu'on le vérifiera sur la copie d'écran ci-après, le texte proposé a surtout pour caractéristique de mettre en évidence les mots de la requête, sans pour autant donner une idée toujours précise de ce que peut contenir la page – on songera ici aux cas de polysémie déjà évoqués, qui peuvent amener des résultats dont la non-pertinence n'est pas toujours immédiatement perceptible au vu des bribes de texte affichées.

Par ailleurs, nombre d'internautes ont été confrontés au petit mystère suivant : les mots clés mis en évidence (en gras) dans le texte affiché, ne se retrouvent pas toujours sur la page que vise le lien proposé. Cette énigme tient en général au fait que les mots qui ont conduit à

sélectionner l'URL se trouvaient sur une version de la page figurant dans le cache du moteur de recherche, alors que le contenu de cette page a changé (à moins qu'un script de re-direction n'amène le visiteur sur une toute autre page). Le manque que nous pointons ici n'est pas sans lien, bien évidemment, avec le silence de l'exposé sur la rubrique [En cache](#). Cela précisé, il est plus important sans doute de noter que même une question qui, dans un cursus de base de formation à la RII, devrait faire l'objet d'un traitement méthodique – « Comment comprendre les informations affichées par un moteur à propos d'un résultat de recherche ? » –, semble ne pas pouvoir rencontrer de réponse satisfaisante dans le cadre d'un tel exposé. Il y a là une difficulté sur laquelle nous reviendrons : la difficulté à donner à des questions « basiques » la place qui devrait leur être conférée.

[Soutien scolaire, Accompagnement scolaire, Maxicours.com](#)
soutien scolaire, accompagnement **scolaire**, cours, cours particuliers. ... Bienvenue sur Maxicours.com, le N°1 du **soutien scolaire en ligne** ...
[www.maxicours.com/](#) - 24k - [En cache](#) - [Pages similaires](#)

[Cyberprofs.com - Le pionnier de l'aide scolaire en ligne - soutien ...](#)
Cyberprofs, le pionnier de l'aide **scolaire en ligne**. Fiches de révision et conseils gratuits. 5000 réponses et corrections de devoirs, sujets et exercices ...
[www.cyberprofs.com/](#) - [Pages similaires](#)

[Accompagnement, aide et soutien scolaire en ligne : Ma Web Classe](#)
MaWebClasse propose un accompagnement **scolaire** aux collégiens cherchant de l'aide : exercices corrigés, qcms, devoirs à distance ... grâce à un **soutien** en ...
[www.mawebclasse.com/](#) - 16k - [En cache](#) - [Pages similaires](#)

[Aide, soutien scolaire et orientation](#)
Le service d'accompagnement **scolaire en ligne** le plus performant qui vous permettra de maîtriser les savoirs fondamentaux et de revoir les acquis. ...
[www.paraschool.com/](#) - 28k - [En cache](#) - [Pages similaires](#)

[Apprendre - cours, formations, exercices et soutien scolaire ...](#)
Des milliers de ressources éducatives, des cours, des exercices et du **soutien scolaire** gratuits en **ligne** pour le primaire, le collège, le lycée, ...
[www.foad-spirit.net/](#) - 112k - [En cache](#) - [Pages similaires](#)

[cyberpapy](#)
Cyberpapy, le premier site gratuit de **soutien scolaire** entre generations, vos forums : aide aux devoirs et comment trouver un bon bouquin.
[www.cyberpapy.com/](#) - 13k - [En cache](#) - [Pages similaires](#)

Questions d'URL

La sous-section suivante répond à la question « Comment décrypter une adresse URL ? ». On retrouve ici un « classique » du tout venant des exposés. Mais le « resserrement » déjà noté à propos de l'examen du contenu de l'affichage d'un résultat n'en continue pas moins d'engendrer des ambiguïtés. En l'espèce les auteurs prennent pour exemple l'URL suivante :

<http://www.clemi.org/spe/pedagogie/spedp.html>. Celle-ci est analysée selon le patron que voici :

protocole://nom-de-serveur/nom-de-répertoire/nom-de-fichier.extension

Selon une pente didactique bien connue, l'exposé semble céder à la tentation de sur-expliquer (à l'instar d'un comédien sur-jouant son rôle) certains éléments constitutifs de l'adresse, à la façon dont un maître d'école pourrait expliquer à de jeunes élèves que, dans l'écriture d'une URL, il ne faut pas oublier la double barre oblique (//) « qui est là pour bien séparer le début de l'adresse (http:) de ce qui suit » – alors que, ainsi qu'on l'a vu, Tim Berners-Lee lui-même regrettait son choix syntaxique, parfaitement inutile et dispendieux. C'est ainsi que, à propos de la présence du préfixe *www*, ayant expliqué que « *www.clemi.org/* est le nom du serveur choisi par son propriétaire », et même si elles observent qu'on se trouve « parfois sur des sites Web dont l'URL ne contient pas les trois “w” », les auteures croient devoir ajouter ¹⁷ : « Ici, il a inclus les trois “w” pour annoncer à l'internaute qu'il se trouve sur le World Wide Web... »

Sous de telles contraintes, qui conjuguent intention d'explication et resserrement du discours, les approximations abondent. Commentant le suffixe *.org* (dans *clemi.org*), l'exposé présente un univers bien rangé là où, en vérité, la pratique tend à brouiller les lignes : le suffixe *.org*, lit-on, « signifie qu'il s'agit d'une organisation non-commerciale » ; le suffixe *.fr* (par exemple) révèle « le pays d'origine » (ce serait la France, ici) ; et on peut supposer que *.com* se verrait gratifier de signe renvoyant à une organisation commerciale ¹⁸. Alors que le patron d'URL proposé distingue, au-delà du nom-de-serveur, le nom-de-répertoire puis le nom-de-fichier, l'adresse prise pour exemple ne se conforme pas à ce modèle ; on tient là peut-être ce qui a poussé les auteures à présenter (correctement) **spe** comme nom de répertoire mais à désigner **pedagogie** comme un nom de *fichier*, alors qu'il s'agit évidemment d'un nom de (sous-) répertoire, et à inventer de façon inattendue, pour le fichier **spedp.html**, le statut de « sous-fichier d'un fichier », alors qu'il s'agit d'un fichier dans un dossier ¹⁹.

Cette sous-section se termine par une question concernant la recherche de l'identité d'une page Web. L'exposé évoque le service *Whois*, qui peut être utile « pour connaître le propriétaire d'un site, ainsi que ses responsables administratif et technique », et en particulier le *Whois* de l'AFNIC, « pour les noms en *.fr* ». En interrogeant ce service à propos de « *clemi.fr* », on obtient ainsi (le 15 juillet 2008) la réponse que l'on découvrira ci-après (<http://www.afnic.fr/outils/whois>). On y apprend d'abord que le titulaire du nom de domaine *clemi.org* est le Centre national de documentation pédagogique (CNDP) et que le contact administratif est Isabelle Brida (*sic*).



En développant le premier titre (« Nom de domaine »), on découvre ceci.

Nom de domaine : clemi.fr

État : Actif (consultez aussi le [Site web](#))

Bureau d'enregistrement : [ACTION WEB](#)

Date de création : 15/03/2005
Date anniversaire : 15 mars

Qualification :

- Identifié grâce au numéro de SIREN 18004301000016 vérifié avec l'INSEE

Serveurs de noms (DNS) :

- Serveur n° 1 : ns1.actionweb.fr [62.210.166.1]
- Serveur n° 2 : ns2.actionweb.fr [62.210.166.2]
- Serveur n° 3 : ns3.actionweb.fr [88.191.51.193]

Le nom de domaine a été créé le 15 mars 2005 auprès du bureau d'enregistrement *Action Web*, lequel fournit la liste des serveurs de noms de clemi.fr, et sur lequel on dispose, en développant le quatrième titre, des renseignements suivants.

Contact technique : DNS AWEB

action-WEB
15, rue de Fontenay
92140 Clamart
France
Téléphone : +33 1 41 90 61 00
Courrier électronique : support@actionweb.fr

Les deux autres titres nous informent respectivement sur le CNDP et sur Isabelle Bréda (au moment du dépôt du nom de domaine).

 **Titulaire : CTRE NAT DE DOCUMENTATION PEDAGOGIQUE**

391bis, rue de Vaugirard
75015 Paris
France
Téléphone : +33 1 53 68 71 00
Courrier électronique : dns@clemi.org

 **Contact administratif : Isabelle Breda**

CLEMI
391bis, rue de Vaugirard
75015 Paris
France
Téléphone : +33 1 53 68 71 00
Courrier électronique : i.breda@clemi.org

Le site *Networksolutions* (<http://www.networksolutions.com/whois/index.jsp>) permet, de même, de s'informer à propos de clemi.org : on a reproduit ci-après les premières lignes relatives à ce nom de domaine, créé le 2 septembre 1997 par Isabelle Breda (*sic*).

```
Domain ID:D1731945-LROR
Domain Name:CLEMI.ORG
Created On:02-Sep-1997 04:00:00 UTC
Last Updated On:02-Sep-2007 01:24:41 UTC
Expiration Date:01-Sep-2008 04:00:00 UTC
Sponsoring Registrar:Namebay (R54-LROR)
Status:OK
Registrant ID:CL74791-NBAY
Registrant Name:isabelle breda
Registrant Organization:clemi
Registrant Street1:391 bis, rue de vaugirard
```

Recherchons maintenant si le nom de domaine clemi.com a été déposé, en consultant par exemple le site *Whois-Search* (<http://www.whois-search.com/whois/cleml.com>) ; on obtient ceci.

```
Registrant:
  Kim Seokjun
  1101-2004, Boram Maeul, Jungl-Dong
  Wonmi-Gu
  Bucheon, Gyeonggi 420-719
  KR

Registrar: DOTSTER
Domain Name: CLEMI.COM
  Created on: 20-JUN-02
  Expires on: 20-JUN-09
  Last Updated on: 08-MAY-08
```

On voit donc que clemi.com n'a rien à voir avec clemi.org ou clemi.fr. Le site *Eurid* (*The European Registry of Internet Domain Names*: <http://www.eurid.eu>) fournit ceci.

Domain	
Name	clemi
Status	REGISTERED (What this means)
Registered	February 15, 2007
Last update	February 15, 2007, 10:25 pm
Registrant	
Language	German
Email	w.-a.clemens@arcor.de
Registrar technical contacts	
Name	Hostmaster Strato Rechenzentrum
Organisation	Cronon AG Professional IT-Services
Language	German
Address	Emmy-Noether-Str. 10 D-76131 Karlsruhe

Là encore, le nom de domaine est réservé... Quant au nom clemi.info, voici un extrait de ce qu'on peut apprendre à son propos (à l'aide par exemple du site *Networksolutions*).

Domain ID:D13503266-LRMS
Domain Name:CLEMI.INFO
Created On:23-May-2006 19:26:41 UTC
Last Updated On:23-May-2008 22:24:09 UTC
Expiration Date:23-May-2009 19:26:41 UTC
Sponsoring Registrar:Key-Systems GmbH (R124-LRMS)
Status:CLIENT TRANSFER PROHIBITED
Registrant ID:P-20125975
Registrant Name:Clemens Hofbauer
Registrant Organization:
Registrant Street1:Statzingerstrasse 7/29
Registrant Street2:
Registrant Street3:
Registrant City:Luffenberg

Notons – ce que l'exposé ne mentionne pas – qu'il est possible à la création d'un nom de domaine de refuser que soient communiquées certaines des données précédentes. Ainsi en va-t-il par exemple du nom de domaine martinemahe.fr, comme on le découvrira ci-après sur la réponse obtenue de l'AFNIC. Le manque d'espace explique sans doute que l'exposé que nous suivons soit dénué de toute illustration sur la question examinée – contrairement au choix que nous avons fait ici. Un point mérite encore d'être souligné : la consultation des données fournies par les services *Whois*, précisent les auteurs, « requiert souvent une connaissance minimale de l'anglais ». De même que l'exposé avait touché (en se contentant de l'évoquer) à la frontière de la programmation Web, de même, ici, il ne peut éviter de mentionner la frontière de l'anglais – même s'il ne se risque guère au-delà.



Contrôler l'information : au-delà de l'URL

La sous-section qui suit apporte une réponse à une question que nous n'avons pas, jusqu'ici, considérée explicitement : « Comment s'assurer de la fiabilité d'une information ? » Le libellé même de cette question peut paraître exagérément ambitieux. Il n'en reste pas moins que la question est, en elle-même, nécessaire. Reformulons-la ainsi : étant donné une question Q , une recherche sur le Web fait apparaître une réponse R « brute », en général partielle, à la question Q ; quel crédit peut-on accorder à cette réponse R ? Telle est la question. La réponse générique apportée par la TAD tient en peu de mots (générateurs parfois d'un immense labeur) : la réponse R doit être soumise à l'épreuve d'une *dialectique des médias et des milieux* appropriée, qu'il reste en général à concevoir, à organiser, à réaliser, et qui permettra d'estimer la validité de R .

Ce que propose l'exposé semble *a priori* tout autre : les auteures y évoquent en effet la mise en œuvre de critères de validation *externe*, résumés dans la formule suivante : « Devant une page ou un site pertinent par rapport au sujet de sa recherche, il faut pouvoir répondre aux questions : qui a écrit cela ? quand ? et dans quel but ? » On recherchera donc – quand c'est possible – l'identité de la personne qui a créé le contenu (l'auteur) ainsi que celle de la personne (ou de l'organisme) qui « en a assuré la publication » : de fait, nombre de sites comportent explicitement une rubrique « Qui sommes-nous ? », censée renseigner l'internaute. Surtout, ce qui a motivé la rédaction et la publication du contenu doit être interrogé : « Devant une ressource en ligne (et hors ligne aussi d'ailleurs), lit-on, il faut toujours s'interroger sur ce que l'auteur cherche à faire savoir et pourquoi. » La question est

complexe, mais elle est présentée comme incontournable : « Il faut apprendre à découvrir les motivations et les valeurs sous-jacentes à l'information. »

Les éléments techniques utiles ne sont pas absents de l'exposé. Celui-ci indique par exemple que, pour identifier l'auteur, on cherchera sa signature « en haut ou en bas de page, ou encore dans un contact électronique » ; qu'on fera de même s'agissant des dates de création et de mise à jour de la page – en ajoutant que, si cela n'aboutit pas, « la validité des liens hypertextes peut donner une indication sur la fraîcheur des informations disponibles sur le site » (la présence de nombreux liens brisés tend à témoigner de la relative ancienneté de la conception de la page). Pour le reste l'exposé renvoie aux « nombreuses grilles d'évaluation en ligne », telles celles du site *Educnet*, dont on trouvera une reproduction partielle ci-dessous²⁰.

Évaluer l'information

Au minimum, répondre aux 3 questions suivantes :

Le site est-il fiable ?
L'information est-elle intéressante ?
Est-ce bien ce que je cherchais ?

Pour aller plus loin...

Qui ?	auteur, organisme...
Quoi ?	qualité de l'information, du document, du site
Quand ?	période traitée, date du document
Où ?	origine de l'information, limites géographiques
Quand ?	période traitée, date du document
Pourquoi ?	objectif(s) du document, du site
Comment ?	structure du document, navigation dans le site

QUI ?

Qui est l'auteur du document ?

L'auteur est-il identifié ? Peut-on le contacter ?
Est-ce un spécialiste du domaine ?
S'exprime-t-il au nom d'une institution ? A titre personnel ?

QUOI ?

Quelle est la nature du site ?

S'agit-il d'un site institutionnel ? d'un site associatif ? d'un site commercial ?
Sa compétence sur le sujet et/ou sa fiabilité sont-elles reconnues ?
Pointe-t-il vers des sites fiables ? Les sites fiables pointent-ils vers lui ?

Quelle est la pertinence des informations ?

Est-ce bien le type d'informations dont j'ai besoin ? (q. préalables)
Le niveau des informations est-il adapté ?
Est-il suffisamment simple ou au contraire suffisamment approfondi ?

Quel est l'intérêt du document ?

Le document est-il vraiment intéressant ? Qu'apporte-t-il de nouveau ?

Toutes ces manœuvres ne doivent cependant pas masquer l'objectif essentiel, dont la poursuite engendre ce en quoi nous pourrions voir un début d'entrée dans *une dialectique des médias et des milieux* adéquate, que les auteures dépeignent comme il suit.

Mais au final, on ne peut juger de la valeur des informations trouvées sur Internet qu'en référence à des choses déjà vues et interprétées. C'est donc en exerçant sa curiosité et en construisant une solide culture générale que l'on devient un internaute averti. En cas de doute face à une information, il ne faut pas hésiter à interroger son entourage ou d'autres internautes.

On retrouve ici une *partie* du système des milieux que l'on peut en effet songer à exploiter. Le premier milieu, c'est au fond la connaissance « clinique » que l'on peut avoir, ou que l'on peut acquérir, du domaine de réalité sur lequel porte l'information dont la fiabilité est examinée. Le second, c'est la « société des semblables », ses proches d'abord – son clan –, ou la tribu plus vaste, en partie anonyme, des internautes – ceux que l'on pourra par exemple interroger dans des forums de discussion.

Médias en ligne : une autre frontière

La troisième section de l'exposé s'intitule *Se tenir informé : les médias en ligne*²¹. Les auteures rappellent que les médias traditionnels (presse écrite, radio, télévision) ont depuis plusieurs années ouvert des sites Web où l'on peut retrouver une grande partie de leur contenu, enrichi souvent par l'intégration d'un contenu multimédia additionnel (texte, son, vidéo). À cela s'ajoutent des médias en ligne qui ne sont que cela : « On parle de webzine, de webradios et de webtélés », écrivent les auteures. La première question soulevée dans cette section est la suivante : « Qu'est ce qu'une newsletter ? » Des explications minimales sont données : une *newsletter*, une *lettre d'information*, que l'on reçoit dans sa boîte aux lettres électronique, est un service spécial auquel il convient de s'abonner. Il s'agit bien là, en effet, d'un dispositif de recherche d'information au long cours dans un secteur d'activité déterminé. La deuxième question abordée – un peu plus longuement – s'énonce ainsi : « Comment construire son actualité sur mesure avec les fils RSS ? » À l'instar des *newsletters*, les fils RSS constituent également une possibilité d'information utile à la RII. Des indications de base sont fournies par l'exposé : la lecture d'un fil RSS peut se faire avec un navigateur intégrant un « lecteur de fils RSS », tel Firefox, Opera, ou Internet Explorer 7.

On notera que l'exposé s'abstient, ici comme ailleurs, de donner méthodiquement l'original anglais du vocabulaire utilisé. L'article "RSS" de l'encyclopédie Wikipedia précise : "RSS is a family of Web feed formats used to publish frequently updated content such as blog entries, news headlines, and podcasts in a standardized format. An RSS document (which is called a 'feed', 'web feed', or 'channel') contains either a summary of content from an associated web site or the full text. RSS makes it possible for people to keep up with web sites in an automated manner..." Le même document indique encore : "The benefit of RSS is the aggregation of content from multiple Web sources in one place. RSS content can be read using software called an 'RSS reader', 'feed reader' or an 'aggregator', which can be web-based or desktop-based." C'est ici l'occasion de souligner que le développement du sigle RSS – Really Simple Syndication –, que propose classiquement l'exposé examiné, fait apparaître un mot – syndication – traditionnel dans le monde de la presse et des médias américains en un sens apparu, d'après le Online Etymology Dictionary (Harper, 2001), dès 1889 : "selling (an article or cartoon) for publication in many magazines or newspapers at the same time", indique pour Syndication la rubrique Quick Definitions du One Look Dictionary, en proposant cet exemple : "He received a comfortable income from the syndication of his work." Le dictionnaire Harrap's Media & Communication (Stevenson et al., 2002) rend le verbe to syndicate par « publier simultanément dans plusieurs journaux » et propose les illustrations suivantes : **she writes a syndicated column** « elle écrit une chronique qui est publiée dans plusieurs journaux », **the photograph was syndicated in all the local newspapers** « la photographie a été publiée dans toute la presse régionale », **a syndicated TV news programme** « des informations télévisées reprises par plusieurs chaînes ». Ce sens des mots syndication, syndicate (to), familier pour l'anglophone, a toutes chances de demeurer opaque pour le francophone non anglophone. Si la frontière de l'anglais n'est pas absente de l'exposé, elle y reste ici un horizon fort lointain.

À côté des lettres d'information et des fils RSS, l'exposé, on l'aura remarqué aussi, ne mentionne pas les « Alertes Google », peut-être parce que celles-ci, à la différence des lettres d'information et des fils RSS, sont engendrées par un robot (qui propose du vrac), et non par des humains (dont les choix sont en principe réfléchis). En revanche, sous l'intitulé « Qu'est-ce que le podcasting ? », il consacre une sous-section à un thème populaire chez les jeunes. Le mot de *podcasting* est lui-même polysémique ; à cet égard, l'article "Podcast" de l'encyclopédie Wikipedia précise utilement ceci.

A **podcast** is a series of digital-media files, which are distributed over the Internet using syndication feeds for playback on portable media players and computers. The term *podcast*, like *broadcast*, can refer either to the series of content itself or to the method by which it is syndicated; the latter is also called **podcasting**.

Le mot désigne ici la mise à disposition par certains sites Web de fichiers audio ou vidéo téléchargeables (par opposition au *streaming*), avec la possibilité de s'abonner à un flux RSS proposant les fichiers nouvellement disponibles. Là encore, un certain nombre d'indications pratiques sont apportées par l'exposé. On a bien, à nouveau, affaire à un dispositif de RII, notamment à moyen et long termes. Aux possibilités offertes par la *pull technology*, ou « technologie de recherche active », qui suppose « une démarche active de la part d'un internaute qui recherche de l'information au moyen de son logiciel de navigation » (Benensen & Juanals, 2004), s'ajoutent ainsi les potentialités de la *push technology*, ou « technologie de distribution personnalisée », qui « fait parvenir de façon automatique une information personnalisée à un internaute ».

Même si l'usage évoqué dans cette section des *newsletters*, fils RSS et autres *podcasts* n'y est pas explicitement articulé avec l'objectif d'une recherche d'information déterminée (ce qui, de notre point de vue, crée une certaine ambiguïté quant à son objet réel), cet exposé, dont l'ambition est certes limitée, fait ainsi rencontrer de *nouvelles frontières* de la RII : tout d'abord, on l'aura noté, celle du *référencement* – de la SEO, la *search engine optimization* ; ensuite, ici même, celle de l'utilisation d'outils d'information – *newsletters*, fils RSS, etc. – que l'internaute doit choisir et auxquels il doit s'abonner, et qui transporte ainsi l'utilisateur au-delà des seules recherches « instantanées », en l'obligeant à planifier une recherche d'information au long cours, installée dans une certaine durée ²².

CHAPITRE 16

Questions de RII

Dans tout ce qui précède, nous avons mis en œuvre, autour de l'étude de l'ensemble Π des praxéologies possibles de la RII, deux techniques apparentées d'observation des exposés mobilisés par notre enquête. La première – chronologiquement et logiquement – de ces techniques consiste à « écouter » ce que dit un exposé donné à propos de la RII, afin d'en dégager les composants praxéologiques, de quelque nature qu'ils soient, que cet exposé fait connaître d'une manière ou d'une autre : nous l'avons mise en œuvre dans la deuxième partie de notre travail et nous l'avons exploitée encore dans le chapitre précédent. La deuxième technique, mise en jeu dans la troisième partie essentiellement, adapte la première technique à des exposés qui, tout en portant sur des matières affines à la RII, ne parlent pas spontanément de RII : elle consiste à s'interroger sur ce qui pourrait s'en déduire pour la RII, à se demander en quoi cela peut importer à une « science de la RII », bref, à examiner la question de l'utilité pour la RII de certaines des praxéologies exposées ou évoquées. Dans un tel cas, nous nous trouvons dans la situation plus générale suivante : une certaine réponse R est apportée par l'exposé (à une question Q qui a trait ou non à la RII), et nous nous demandons si, à partir de R et de quelques autres ingrédients, on ne pourrait pas apporter une réponse R^* à une certaine question Q^* que R nous inspire ou que nous imaginons, et qui, elle, concerne la RII. En vérité, cette activité de questionnement *est partout présente* dans les parties précédentes : ainsi l'écoute d'un exposé, dans la deuxième partie de ce travail, nous a-t-elle amenée bien des fois à soulever des questions complémentaires, et à leur apporter au moins des esquisses de réponses. Le jeu des questions et des réponses apparaît ainsi comme le moteur essentiel de ce travail. Dans certains cas, nous avons la question et la réponse (fournies par un certain exposé) ; dans d'autres cas, nous avons une réponse qui nous inspirait une question à laquelle nous avons en règle générale tenté d'apporter réponse. Dans la plupart des cas, en même temps, nous avons soulevé des questions de notre cru, avec des réponses à l'avenant. Cette activité de questionnement, inspirée de façon en quelque sorte naturelle par les exposés examinés, était fortement appuyée sur notre propre pratique d'Internet, en matière de création de sites, de référencement de sites, de formation – d'abord en entreprise ensuite à l'université,

à propos de diverses matières du vaste champ des TIC. Dans ce qui suit la technique d'étude de l'ensemble II des praxéologies de la RII sera, sur un point, bien différente.

Quelles questions se poser ?

Face à un exposé, deux attitudes sont possibles. La première, que nous avons jusqu'ici adoptée, consiste à *l'écouter*, à s'efforcer d'entendre *ce qu'il dit*. Cette technique, évidemment indispensable, procède de l'idée commune de lecture. Elle a pourtant le désavantage de gommer ce que l'exposé *ne dit pas*, alors qu'il *pourrait* le dire ; de faire jouer à ce qui est dit un rôle d'écran qui nous cache ce qui est tu. Une seconde attitude consiste alors, non à laisser parler le texte, mais à *l'interroger*, à le questionner : répond-il à cette question ? Et à cette autre ? Et à telle autre encore ? C'est là ce qu'on nomme en TAD, en usant d'un néologisme *ad hoc*, une lecture *excriptrice* du texte considéré. C'est là aussi une technique que nous avons maniée dans tout ce qui précède. Mais l'usage de cette technique suppose *que l'on ait des questions à poser*. Nous avons rappelé ce qu'était, pour l'essentiel, la source de nos propres questions, notre pratique et notre réflexion autour de la RII et, plus largement, du monde Informatique & Internet. Dans ce qui suit, désormais, nous nous excentrerons de ce processus : nous demanderons à d'autres que nous-même, dotés généralement d'une pratique et d'une réflexion de haut niveau, de nous dire ce que sont pour eux les questions auxquelles une culture praxéologique de la RII, porteuse d'intelligibilité et d'opérationnalité, devrait pouvoir répondre. En d'autres termes nous interrogerons désormais des exposés sur les questions qu'on s'y pose plutôt que sur les réponses qu'on y apporte – à moins que ces réponses n'incorporent la question objet de notre enquête. Les exposés que nous examinerons constitueront en même temps un *milieu* permettant d'interroger, de valider, de rectifier si besoin est, et d'enrichir notre propre système de questions, développé dans les parties précédentes.

"How to search"

Le premier exposé que nous interrogerons est un ouvrage déjà plusieurs fois mentionné. Intitulé *Web Dragons* et paru en 2006, il est signé de Ian H. Witten, Marco Gori et Teresa Numerico. Pris ensemble, ces auteurs couvrent une très large partie du monde Informatique & Internet. Le premier est professeur de *computer science* à l'université Waikoto (Nouvelle-Zélande). Après des études de mathématiques à Cambridge, puis d'informatique à Calgary (Canada), il a obtenu un PhD en ingénierie électrique à l'université d'Essex (Angleterre). Ses

publications portent sur différents thèmes, dont les bibliothèques digitales (il est le co-auteur d'un ouvrage intitulé *How to build a digital library?* paru en 2003), l'apprentissage artificiel (*machine learning*), la compression de texte (il a cosigné en 1999 un ouvrage intitulé *Managing Gigabytes: Compressing and Indexing Documents and Images*), la fouille de données (il a fait paraître en 2005 la deuxième édition d'un ouvrage écrit en collaboration, intitulé *Data Mining*), à quoi on peut ajouter, parmi d'autres, un article un peu ancien (1985) intitulé "Elements of computer typography". Le deuxième auteur, Marco Gori, est lui aussi professeur de *computer science*, mais à l'université de Sienne (Italie). Lui aussi s'occupe d'apprentissage artificiel appliqué à la reconnaissance de formes (*pattern recognition*), ainsi qu'au *Web mining*. Ancien président de l'association italienne d'intelligence artificielle, membre de l'IEEE et de l'ECCAI²³, il assume des responsabilités au sein du chapitre italien de la *Computational Intelligence Society* de l'IEEE. La troisième cosignataire du livre, Teresa Numerico, enseigne à l'université de Salerne (Italie). Spécialiste de l'histoire et de la philosophie de l'informatique (sa thèse de doctorat s'intitulait *Le origini dell'informatica nell'opera di Alan Turing*), elle donne notamment un enseignement sur les nouveaux médias (*new media*), d'un double point de vue technique et théorique, et s'intéresse en particulier aux apports possibles de la philosophie et de l'éthique "for a better understanding of information and communication technologies"²⁴.

On voit donc que les auteurs de l'ouvrage que nous avons choisi comme pourvoyeur de questions de RII ont un champ de compétences théoriques et pratiques qui intègre, sans s'y réduire, le *Web searching*. Leur *opus* comporte sept chapitres dont le quatrième est intitulé *How to search*. C'est lui que nous examinerons ici du point de vue qui est le nôtre maintenant. Pour l'orientation du lecteur nous reproduisons ci-après les titres et sous-titres qui structurent le texte du chapitre 4.

Searching Text / Full-text Indexes / Using the Index / What's a Word? / Doing It Fast / Evaluating the Results

Searching in a web / Determining What a Page Is About / Measuring Prestige / Combining Prestige and Relevance / Hubs and Authorities / Finding Hubs and Authorities for a Query / Using Hubs and Authorities / Discovering Web Communities / Bibliometrics / Learning to Rank / Distributing the Index

Developments in Web Search / Searching Blogs / Ajax Technology / The Semantic Web

Birth of the Dragons / The Womb Is Prepared / The Dragons Hatch / The Big Five / Inside the Dragon's Lair

So What?

Notes and Sources

Quelle technique avons-nous appliquée à ce texte ? Considérons à titre d'exemple le fragment suivant, premier paragraphe d'un développement qui, sous le titre *Determining what a page is about*, en comporte trois.

There were other problems. One arose when people started to realize how important it was to get noticed and began putting words into their pages specifically for that purpose. Traditional information retrieval takes documents at face value: it assumes that the words they contain are a fair representation of what they are about. But if you want to promote your wares, why not include additional text intended solely to increase its visibility? Unlike paper documents, words can easily be hidden in web pages, visible to programs but not to human readers. (p. 113)

Comme pour tous les autres paragraphes du chapitre que nous examinons – ils sont 163, si nous ignorons les sections intitulées *So what?* et *Notes and sources* –, nous avons associé à ce paragraphe des questions pour lesquelles on peut considérer que le texte du paragraphe fournit une réponse ou un embryon de réponse. En l'espèce, nous avons formulé deux questions (contenant elle-même des sous-questions), dont la première s'énonce ainsi : « Est-il possible d'inclure dans une page Web du texte invisible du lecteur humain, et à l'adresse exclusive des moteurs de recherche ? Sous quelle forme ? » La seconde question annoncée est alors la suivante : « Quel peut être l'intérêt de placer sur une page Web du texte invisible du lecteur humain mais indexable par un moteur de recherche ? Cela peut-il donner lieu à des abus ? Lesquels ? » En règle générale nous ne prétendons pas que cette génération de questions à partir du texte soit exhaustive, ni même qu'elle soit plus pertinente qu'une autre. Il nous suffit, pour qu'une question soit associable au texte, qu'on trouve dans ce texte des déclarations qui, de manière explicite (à travers une ou plusieurs assertions) ou implicite (à travers des présupposés ou des renvois), puissent être regardées comme des fragments de réponses à ladite question. Dans le cas de la première question citée plus haut, la première sous-question trouve un fragment de réponse dans le passage suivant : “Unlike paper documents, words can easily be hidden in web pages, visible to programs but not to human readers.” La seconde sous-question, « Sous quelle forme ? », est plus fragile : le texte à partir duquel nous l'avons engendrée ne donne ainsi aucune indication de nature technique (par exemple sur le fait qu'il s'agirait de *metatags*), et se contente de parler de “additional text”, de texte qui se surajoute au texte visible par un lecteur humain. Le fait que le texte additionnel ne soit pas lu par le lecteur humain implique bien sûr qu'il n'est pas écrit comme le texte

« visible ». Cette implication évidente nous fait ainsi rencontrer un phénomène essentiel, le fait que le passage du texte aux questions engendrées à partir de lui *ne « boucle » pas sur ce texte*. Le texte n'est pas « saturé » par rapport aux questions qu'il suggère pourtant à travers son contenu.

Examinons à cet égard la deuxième question relative au premier paragraphe : « Quel peut être l'intérêt de placer sur une page Web du texte invisible du lecteur humain mais indexable par un moteur de recherche ? Cela peut-il donner lieu à des abus ? Lesquels ? » La première sous-question trouve son origine et a donc au moins une partie de sa réponse dans ce passage : "... people started to realize how important it was to get noticed and began putting words into their pages specifically for that purpose." Les mots (*words*) rajoutés au texte destiné au lecteur humain ont pour but de faire remarquer (*get noticed*) la page Web concernée – sous-entendu, remarquée par les moteurs de recherche (ou, comme le disent les auteurs, par les *programs*). En revanche, les deux dernières sous-questions ont une assise moins clairement dessinée dans le texte. C'est de manière surtout allusive, en effet, que les auteurs y suggèrent la possibilité d'abuser les *programs* : normalement, c'est-à-dire dans la recherche d'information « traditionnelle », celle d'avant le Web et les moteurs de recherche, laisse entendre leur texte, le contenu d'un document est pris à sa valeur faciale (*face value*), parce qu'on suppose qu'il respecte le principe de *fair representation* (représentation non biaisée). Le choix du vocabulaire, ici, est en lui-même informatif : le texte additionnel aurait pour objet de faire la promotion des *wares* (marchandises) que l'instigateur du texte additionnel souhaite « vendre ». En outre, l'adjonction de texte à destination des moteurs de recherche, nous est-il dit encore, a pour *seul* but d'accroître la visibilité de la page – il est *intended solely to increase its visibility*.

Cet exemple illustre un phénomène général. Un texte est lu comme s'il donnait à entendre un discours clos, assuré dans ses affirmations, ses présupposés ou ses allusions, et – en apparence – étranger autant qu'indifférent à tout ce qu'il ne contient pas de façon manifeste. Un texte, et en particulier ce qu'on peut appeler un texte *d'autorité*, se présente à nous comme un tissu serré, fidèle en cela à l'étymologie du mot (le latin *texere* signifie « tisser », *textus* désignant la chose tissée), qui exclut tout ce qu'il ne renferme pas. Du point de vue que nous venons de préciser, nous retrouvons ici une évidence : tout texte est une « œuvre ouverte », une *opera aperta*, selon l'expression popularisée dans un autre contexte par Umberto Eco (1965). En un sens, en effet, un texte fait accroire au lecteur qu'il en sait plus qu'il n'en dit, car si l'on entend ce qu'il dit comme des fragments de réponses à des questions, il apparaît nécessairement, aussi savant soit-il, démuné face aux questions qu'il évoque ainsi. Au reste, on le sait, interrogé sur telle question qu'il évoque, un texte peut

n'avoir qu'une réponse fort légère, voire indigente – un fait dont le comique a su autrefois tirer parti ²⁵. Même s'il promet plus qu'il ne pourra tenir, il n'en demeure pas moins qu'un texte doit, si l'on peut dire, être *pris au mot*. Dans le cas qui nous occupe ici, nous le verrons, ce choix peut paraître de prime abord audacieux étant donné la multiplicité des savoirs et savoir-faire portés par le texte et révélés par les questions que nous en avons tirées : en bien des cas, le lecteur dont l'opinion est faite en matière de RII se laissera surprendre par des questions que nous regarderons fermement, quant à nous, comme renvoyant (au moins en partie) à des praxéologies *déclarées utiles* en la matière par les auteurs que nous suivons. Il est vrai que le texte examiné ici tranche par rapport à certains exposés sur la RII, qui se réduisent à des listes plus ou moins longues de « trucs » ²⁶, en cela qu'il participe d'une volonté de construire une *culture* de la RII assumée dans sa profusion, sa diversité, son éclectisme, comme nous allons le voir maintenant.

Des trois premières pages du chapitre intitulé *How to search*, nous avons tiré l'ensemble de questions qu'on trouvera reproduites ci-après. Les indications figurant en gras donnent le numéro de la page suivi du numéro du paragraphe dans la page (un paragraphe étant compté sur la page où il commence). Ainsi l'indication **101, § 1** désigne-t-elle le premier paragraphe de la page 101 ; comme on le voit, à ce paragraphe nous avons, en l'espèce, associé trois questions, dont chacune est précédée du signe dièse (#), et dont la dernière est constituée de deux sous-questions.

101, § 1

Qu'appelle-t-on concordance d'une œuvre écrite ?

En 1829, Mary Novello, épouse de Charles Clarke, commença à établir une concordance des œuvres de Shakespeare. Avant cela, pour quel texte avait-on établi une concordance ?

Combien de temps fallut-il à Mary Novello pour établir une concordance de l'œuvre de Shakespeare ? Combien d'entrées comportait cette concordance ?

101, § 2

Qu'indiquait la concordance des œuvres de Shakespeare établie par Mary Novello pour chacun des mots extraits de ces œuvres ?

Quel était l'intérêt proclamé de disposer d'une concordance des œuvres de Shakespeare ?

102, § 1

Comment établissait-on l'index d'une œuvre ou d'un ouvrage avant l'apparition de l'informatique ?

Pour quelles raisons une concordance était-elle appelée ainsi ?

102, § 2

Pourquoi la concordance publiée en 1897 du Nouveau Testament en grec n'aurait pas pu être entièrement mécanisée ?

102, § 3

Quelle organisation humaine et matérielle fut adoptée pour établir la concordance de l'œuvre poétique de Wordsworth publiée en 1911 ? Combien d'entrées comportait-elle ? Combien de temps fut nécessaire pour l'établir ?

102, § 4

À quelle époque put-on établir des concordances à l'aide d'ordinateurs ?

De quels défauts souffraient les premières concordances établies à l'aide d'un ordinateur mais imprimées en capitales, sans ponctuation ?

À partir de quelle époque devint-il possible, à l'aide des ordinateurs personnels, de dépasser en qualité les concordances anciennes, faites à la main ?

103, § 1

Comment établir une concordance d'un texte dont on possède une version électronique, ou d'un ensemble de pages Web, ou un livre tombé dans le domaine public ?

Qu'es-ce que le projet Gutenberg ?

Qu'est-ce que l'Internet Archive ?

103, § 2

Les notions de concordance et de "full-text index" (index en texte intégral ou en plein texte) ont-elles à voir l'une avec l'autre ?

Quelles recherches peut-on faire concernant un texte pour lequel on dispose d'un index en texte intégral ?

Rappelons une fois encore que ce qui importe ici, *ce sont les questions*, qui définissent *en puissance* une culture praxéologique de la RII : nous ne poserons le problème des *réponses* qu'au chapitre suivant. Dans la série de questions ci-dessus, on voit apparaître, en fin de liste, une notion clé : celle d'*index* d'un texte, et plus précisément d'*index en texte intégral*. Posséder une culture de la RII, pour les auteurs, ne se conçoit donc pas sans la maîtrise praxéologique de cette notion (nous allons y revenir). Mais on perçoit aussi une exigence d'*approfondissement historique* de cette culture, ici par la mise en relation de la notion « moderne » d'*index* avec la notion ancienne de *concordance* et avec les travaux auxquels elle a donné lieu, depuis Mary Novello (1809-1898) jusqu'à nos jours, à travers la montée progressive de la mécanisation puis de l'informatisation²⁷. L'exigence portant sur la

connaissance des dynamiques historiques qui articulent le présent au passé, y compris au plan pratico-technique (comme le montre la question qui suit immédiatement la mention « 103, § 1 »), s'étend au monde contemporain : une culture adéquate de la RII suppose ainsi la connaissance des grands projets qui animent et structurent le Web, qu'il s'agisse du projet Gutenberg ou de l'*Internet Archive*.

“Searching Text”

Après une introduction que nous avons parcouru à travers les questions qui lui ont été associées, l'exposé auquel nous nous référons est scindé, on l'a vu plus haut, en quatre sections principales : *Searching Text*, *Searching in a web*, *Developments in Web Search*, *Birth of the Dragons*. C'est la première de ces sections que nous parcourons ici. L'amorce de programme d'étude dessinée par les questions déjà examinées va être abondamment *augmentée*, comme le montrent les questions que nous avons tirées des quelques paragraphes suivants.

104, § 1

Lorsqu'on lui soumet une requête, un moteur de recherche cherche-t-il les mots qui la composent dans l'ensemble des textes disponibles sur le Web ?

Full-text Indexes

104, § 2

Pourquoi un moteur de recherche utilise-t-il un index plein texte ?

Considérons la phrase suivante : *to be or not to be that is the question*. Que serait un index plein texte du texte suivant, en utilisant comme pointeur le rang du mot dans la phrase (comme indiqué ci-après) : to be or not to be that is the question ?

1 2 3 4 5 6 7 8 9 10

Que signifie le mot *pointer* dans la phrase suivante : *a full-text index lists all the words that appear in the text, along with a pointer to every occurrence* ?

Quelle relation y a-t-il entre le nombre de pointeurs d'un index plein texte et le nombre de mots du texte ?

105, § 1

Dans un index plein texte le pointeur d'une occurrence d'un mot est-il toujours le numéro du rang de ce mot dans le texte ? Comment faire lorsque l'index concerne un ensemble de textes (par exemple un ensemble de pages Web, ou les pièces de théâtre écrites par Shakespeare) ?

105, § 2

Dans un index relatif à un ensemble important de textes (par exemple 1 milliard de pages Web ou 2 000 ouvrages imprimés), y a-t-il plus (voire beaucoup plus) d'entrées, ou moins (voire beaucoup moins) d'entrées, que dans un dictionnaire de langue en plusieurs volumes ?

L'index de Google contient-il plusieurs milliards d'entrées ou plutôt seulement plusieurs millions d'entrées ? Pourquoi ?

105, § 3

Qu'appelle-t-on un *hapax legomenon* ?

Dans les *Complete works of Shakespeare*, qui contiennent environ 31 500 mots différents, quel est le pourcentage de mots qui n'apparaissent qu'une fois ? Quel est, approximativement, le nombre d'occurrences des mots les plus souvent utilisés ?

Google aurait déclaré que son index référence 21 630 000 000 pages Web contenant au moins une occurrence d'un certain mot de la langue anglaise. De quel mot peut-il s'agir ? Dans cet index, quel peut être le nombre de pointeurs relatif à ce mot ?

Dans un index, quelle allure a la distribution du nombre d'occurrences des différents mots indexés ?

105, § 4

Peut-on dire qu'un index est aussi gros que le corpus de textes qu'il indexe ? Pourquoi ?

Lorsqu'on applique une technique de compression, est-il vrai que l'index diminue sensiblement plus que le corpus de textes lui-même ? Pourquoi ?

Est-il exact que les moteurs de recherche disposent et d'un index plein texte et du texte lui-même des pages trouvées sur le Web ? Ces deux ensembles de données occupent-ils sensiblement le même espace ? Pourquoi ?

On retrouve ici l'importance accordée par les auteurs à la continuité culturelle par delà les discontinuités technologiques. La toute première question de cette suite peut apparaître comme ayant, dans une culture de la RII qui ne se réduit pas à quelques recettes, une place incontestable. Les questions suivantes font apparaître qu'il ne semble pas moins important de maîtriser les principes et les notions de base de la constitution et de l'utilisation des index, en même temps qu'il convient de disposer de connaissances « contextuelles », à propos de Google (et de ce qu'il révèle sur la langue anglaise) par exemple, ou encore de la distribution des mots dans un texte, tout cela se situant dans la proximité de connaissances relatives au fonctionnement des moteurs de recherche.

La série de questions suivante se place sous le titre *Using the index*. Comme l'indique cet intitulé, les auteurs se préoccupent ici de faire entrer leur lecteur dans le fonctionnement

de la recherche de documents à partir du moment où l'on dispose d'un index : pour l'utilisateur qui aurait franchi la frontière des opérateurs booléens, il s'agit en quelque sorte de passer de l'autre côté de l'écran, pour comprendre par exemple pourquoi les moteurs de recherche peuvent se permettre de lui offrir la possibilité de combiner logiquement des mots clés, de faire rechercher une expression exacte ou de pratiquer une recherche de proximité (*proximity search*). On a là, bien évidemment, le franchissement d'une frontière, tracée antérieurement à l'essor historique de l'Internet et des moteurs de recherche, et qui structure le champ de la RI, la *recherche d'information*, appelée IR (pour *information retrieval*) en anglais.

Selon l'article "Information retrieval" (§ 1) de l'encyclopédie Wikipedia, on peut définir la RI comme la "science of searching for documents, for information within documents and for metadata about documents, as well as that of searching relational databases and the World Wide Web". La RI a pris son essor dans les années 1950 mais a connu un développement exceptionnel dans les années 1990, avec l'avènement de l'Internet et du Web. C'est en 1992, ainsi, que le département de la défense américain a, conjointement avec le NIST (National Institute of Standards and Technology), parrainé la Text Retrieval Conference (TREC) : le but était de fournir à la communauté scientifique de la RI les moyens de procéder à des évaluations, sur de vastes collections de documents, des méthodologies existantes. Selon l'article déjà cité (§ 4), "this catalyzed research on methods that scale to huge corpora". Par la suite, l'apparition des premiers moteurs de recherche sur le Web "boosted the need for very large scale retrieval systems even further". La recherche sur les SRI (systèmes de recherche d'information, IR systems en anglais) est aujourd'hui très active, y compris en France, comme en témoigne l'ouvrage récent dirigé par Mohand Boughanem et Jacques Savoy (2008). La RI est un domaine à forte codisciplinarité, où interviennent (ou peuvent intervenir) l'informatique, les mathématiques, la statistique, la Library science, la linguistique, la psychologie cognitive, etc. Une culture raisonnée et bien fondée de la RII ne peut, croyons-nous, se développer que dans une relation attentive à la RI et aux domaines scientifiques que celle-ci sollicite. On ne s'étonnera donc pas de ce que plusieurs des notions évoquées ci-après, en suivant les auteurs de Web dragons, appartiennent typiquement à la RI.

On a reproduit ci-dessous les questions annoncées : on y verra que les auteurs vont plus loin encore que ce que nous mentionnions plus haut.

106, § 1

Lorsqu'on dispose de l'index d'un corpus de documents, comment est-il possible de trouver les documents du corpus contenant un certain mot ?

Est-il plus rapide pour un ordinateur d'examiner un texte à la recherche d'un certain mot ou de faire la même chose sur une liste ordonnée de mots ?

106, § 2

Comment procède-t-on pour répondre à une requête relative à plusieurs mots, c'est-à-dire pour choisir les documents indexés contenant chacun de ces mots ?

Comment procède-t-on pour répondre à une requête constituée d'une combinaison logique de mots (incluant tel mot ou tel autre, excluant tel autre mot encore, etc.) ?

106, § 3

Comment procède-t-on pour identifier les documents indexés contenant une expression exacte donnée ?

Comment procède-t-on pour identifier les documents indexés contenant deux mots qui, dans le document, sont proches l'un de l'autre, c'est-à-dire pour effectuer une recherche de proximité (*proximity search*) ?

106, § 4

Dans un corpus de documents indexés par un moteur de recherche, l'ordre des documents a-t-il un sens par rapport à une requête donnée ? Pourquoi ?

106, § 5

Étant donné une requête composée de plusieurs mots, quels critères peut-on utiliser pour estimer la plus ou moins grande pertinence des documents contenant ces mots en totalité ou en partie ?

Étant donné une requête composée de plusieurs mots, auxquels de ces mots est-il logique de donner le plus d'importance dans le choix des documents pour répondre à la requête ?

107, § 1

Étant donné une collection de documents, on peut représenter un document de la collection par un point dans un espace dont le nombre de dimensions (qui peut être alors de plusieurs milliards) est égal au nombre de mots, chaque mot déterminant un axe de l'espace et chaque document ayant pour coordonnée sur cet axe le nombre d'occurrences du mot dans ce document. Par exemple si une collection de documents ne comporte que cinq mots, m_1 , m_2 , m_3 , m_4 , m_5 , et si un certain document comporte 3 occurrences de m_1 , 24 occurrences de m_2 , 0 de m_3 , 68 de m_4 et 7 occurrences de m_5 , ce document sera représenté dans un espace à 5 dimensions par le point de coordonnées (3, 24, 0, 68, 7). Comment est-il logique de mesurer les distances sur chaque axe si l'on considère que, si un mot

est rare dans la collection de documents, une occurrence de plus ou de moins fait une plus grande différence que si le mot est plus fréquent ?

Lorsqu'on cherche un mot dans une collection de documents où ce mot n'a que de rares occurrences, on peut avoir l'impression que chacune de ces quelques occurrences nous informe sur le mot « en général ». Si, par contraste, ce mot a de nombreuses occurrences dans *l'un* des documents de la collection, on peut avoir l'impression que ce fait nous informe sur *ce document* (plutôt que sur le mot lui-même). Comment ces impressions sont-elles intégrées dans la modélisation classique d'une collection de documents par un espace dont chaque axe correspond à l'un des mots présents dans la collection ?

107, § 2

Lorsqu'on représente les documents d'une collection par un point dans un espace selon le modèle classique, deux documents représentés par des points distincts alignés avec l'origine comportent les mêmes mots dans les mêmes proportions (mais pas en même quantité) : du point de vue d'une requête donnée, ces deux documents n'ont pas à être distingués. Quand, au contraire, deux documents sont situés sur des droites différentes issues de l'origine, comment évalue-t-on la similitude (*similarity*) entre ces documents (cette similitude étant égale à 1, par définition, dans le cas de documents représentés par des points alignés avec l'origine) ?

107, § 3

Dans la modélisation classique d'une collection de documents par un espace dont le nombre de dimensions est égal au nombre de mots, la mesure de similitude à l'aide du cosinus traite comme identique un texte et tout autre texte issu du premier dont les mots ont été rangés dans n'importe quel ordre. Cela est-il préjudiciable à l'utilisation de cette mesure de similitude par un moteur de recherche ? Pourquoi ? Si oui, que faire ?

Comme on l'aura noté, à partir du paragraphe 5 de la page 106 est abordée la question clé de la *pertinence* d'un document vis-à-vis d'une requête. Ici encore, une autre frontière est franchie par les auteurs. Mais très vite, au-delà de certaines connaissances simples et quasiment de bon sens, eux-mêmes vont toucher à une frontière qu'ils se garderont bien de franchir : la frontière des *mathématiques*. La modélisation d'une collection de documents par ce que les auteurs nomment ici simplement un espace (*space*) est en fait connue dans le domaine de la RI sous le nom de *vector space model*, l'article princeps sur le sujet ayant plus de trente ans d'âge (Salton, Wong & Yang, 1975). Pour le contraste, alors que le texte que nous suivons ne fait droit à aucun formalisme mathématique (il ne contient pas même ce que nous nous sommes permis d'introduire dans la première question qui suit la mention **107, § 1**),

nous avons reproduit ci-après un bref passage du livre déjà mentionné dirigé par M. Boughanem et J. Savoy (2008, p. 31).

Le document (vecteur \vec{d}) et la requête (vecteur \vec{q}) sont représentés là encore dans un espace euclidien de dimension élevée engendré par tous les termes de l'index. La similarité est alors le cosinus de l'angle formé par les deux vecteurs :

$$\begin{aligned}
 RSV(\vec{d}, \vec{q}) &= \frac{\vec{d} \cdot \vec{q}}{|\vec{d}| * |\vec{q}|} \\
 &= \frac{\sum_{i=1}^n w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^n w_{i,j}^2} * \sqrt{\sum_{i=1}^n w_{i,q}^2}} \quad [1.4]
 \end{aligned}$$

D'autres fonctions de similarité ont été proposées dans la littérature, parmi lesquelles on peut citer les mesures de Jaccard et Dice.

Les auteurs de *Web Dragons*, eux, écartent tout symbolisme mathématique, en restent aux mots, et cheminent précautionneusement le long d'une frontière que, à l'évidence, ils ne souhaitent pas franchir, même si, parfois, leur pas se fait plus incertain, comme le montre le passage suivant.

To judge the similarity of two documents, we use the angle between their vectors. To judge the relevance of a document to a query, we treat the query as a tiny little document and calculate the similarity between it and the document. Technically, the cosine of the angle is used, not the angle itself, and this is called the *cosine similarity measure*. It ranges from 1 if the documents are identical to 0 if they have no words in common. (p. 107)

Trois conclusions doivent être soulignées en ce point. Tout d'abord, chez des auteurs pour qui ce qui se passe de l'autre côté de la frontière mathématique est d'une grande familiarité sans doute, on observe un parti pris apparemment irrévocable de n'y faire référence qu'allusivement. Ces auteurs, à l'instar de beaucoup d'autres, semblent avoir entériné un interdit touchant les mathématiques, qui serait l'analogue de l'interdit frappant le recours à la langue anglaise dans les textes français – l'utilisateur français se heurtant à la fois à la frontière de l'anglais (ce qui n'est plus le cas pour une grande partie de la population du monde) et à la frontière des mathématiques (ce qui semble être encore le lot de beaucoup sur la planète). Ensuite l'exemple précédent illustre un phénomène beaucoup plus général, que nous avons

déjà beaucoup rencontré : celui de l'existence de frontières qui traversent un domaine praxéologique – pour nous, la RII – pour l'amputer au nom de contraintes allogènes (sociales et culturelles), non spécifiques, alors même que de telles amputations ne trouvent nulle raison d'être dans le domaine praxéologique considéré. Enfin et surtout, on imagine aisément que le texte auquel nous avons associé les questions que l'on vient de parcourir ne saurait apporter à ces questions plus que des ébauches de réponse. Ici, proposer des réponses « solides » exigerait de faire quelques pas sur le territoire des mathématiques, ce que les auteurs, on l'a vu, se refusent à faire. Leurs textes nous *parlent* de ces réponses plutôt qu'ils ne nous les *montrent*. Bien entendu, qu'un texte ne réponde pas véritablement aux questions qu'il porte en lui *est un fait très général*, qu'il faut chercher à interpréter en termes de frontières – globales ou locales –, et qui sera au cœur des phénomènes analysés dans le chapitre suivant de ce travail.

Les auteurs s'engagent ensuite dans un développement qu'ils ont intitulé *What's a Word?* Les questions que nous lui avons associées concernent le traitement du texte d'un document en vue de l'indexer. Comme on le verra, les procédures les plus usuelles y sont mentionnées (pli de casse, lemmatisation, filtrage des mots vides). À nouveau, il s'agit d'aider l'utilisateur à dépasser la phénoménologie de son activité, en lui demandant de l'interpréter à la lumière d'une nouménologie qui, par ailleurs, est le pain quotidien des spécialistes de la RI – la question relative au mot « AT & T » explicitant concrètement le lien entre phénoménologie et nouménologie.

107, § 4

Quels problèmes posent l'utilisation de la notion de mot dans l'indexation d'un texte ?

108, § 1

Dans un texte, qu'est-ce qui peut constituer un mot, c'est-à-dire une chaîne alphanumérique d'un seul tenant qui puisse apparaître comme une requête adressée au moteur de recherche ?

108, § 2

Qu'est-ce que le *case folding* (pli de casse) ? Quels problèmes cela entraîne-t-il ?

108, § 3

Qu'est-ce que le *stemming* (lemmatisation) ? Quels en sont les avantages et les inconvénients dans l'utilisation d'un moteur de recherche ?

108, § 4

Qu'est-ce qu'un *stop word* (mot vide) ? Quelle est la proportion approximative des *stop words* dans un texte ?

Que font les moteurs de recherche avec les *stop words* ? Y a-t-il un intérêt à cela du point de vue de l'espace occupé par l'index ? Lequel ?

108, § 5

Quelles conséquences fâcheuses peuvent avoir pour l'utilisateur la décision d'ignorer tel mot lors de la constitution de l'index (c'est-à-dire de l'inclure dans la liste des *stop words*) ?

109, § 1

Que reste-t-il du mot « AT&T » (l'acronyme de *American Telephone and Telegraph*) si le logiciel d'indexation transforme les capitales en bas de casse et ignore : 1. les caractères spéciaux (qui ne sont pas des lettres) ; 2. les mots d'une lettre ; 3. le mot "at" ?

Sous un titre vigoureux, *Doing it fast*, les questions associées ensuite au texte que nous interrogeons participent d'une culture de la RII dans laquelle on ne se soucie pas seulement des *effets* – du point de vue de l'usager – de ce qui se passe dans les coulisses des moteurs de recherche : on s'y préoccupe aussi, comme on va le voir, des problèmes généraux rencontrés par le développement des moteurs de recherche, et l'on s'y familiarise avec quelques-unes des solutions les plus connues, tel le *binary searching*.

109, § 2

Quel est le principe mis en œuvre pour résoudre le problème de la construction d'un index en texte intégral ?

109, § 3

Pour fabriquer un index d'une importante collection de documents, le problème clé est-il plutôt un problème de programmation ou un problème de capacité de mémoire ? Ou encore de rapidité d'accès aux données ?

109, § 4

Lors de l'indexation d'une vaste collection de documents, comment peut-on concilier l'énormité de l'ensemble des données à traiter et le caractère nécessairement limitée de la mémoire principale (*main memory*) ?

La conception d'un logiciel d'indexation utilise-t-elle peu ou beaucoup de mathématiques ? Pourquoi ?

109, § 5

Pour rechercher une entrée dans un index, on utilise en général la technique de dichotomie (*binary search*). De quoi s'agit-il ? Pourquoi procéder ainsi ?

Lorsqu'un logiciel de recherche a identifié une entrée (mot clé) dans un index, il doit encore parcourir les pointeurs correspondants à cette entrée, ce qui peut être long. Comment faire pour fournir au plus vite des résultats à l'utilisateur ? Comment faire si la requête comporte plusieurs mots clés ?

110, § 1

Peut-on saisir dans la boîte de requêtes d'un moteur de recherche une longue suite de mots clés, voire un paragraphe entier d'un texte ? Si oui, est-il tout de même plus efficace de se limiter à deux ou trois mots clés ? Pourquoi ?

Pour quelles raisons la boîte de requêtes (*query box*) d'un moteur de recherche a-t-elle en général une taille réduite ?

110, § 2

Le classement en fonction de leur mesure de similitude des documents trouvés en réponse à une certaine requête est-elle une opération plus délicate que la recherche de ces documents elle-même ? Pourquoi ?

Quand on classe des documents en utilisant la mesure de similitude du cosinus, est-il possible de trouver les 10 ou 20 documents les plus pertinents sans comparer tous les documents entre eux ? Si oui, de quelle nature sont les connaissances clés qui permettent d'y parvenir ?

110, § 3

Quelles qualités peut-on attendre d'un bon système d'indexation (*indexing system*) ?

La construction d'un système d'indexation de qualité est une affaire complexe. Une fois la chose faite, de quel avantage essentiel bénéficie-t-on ?

Dans le paysage que l'on vient de parcourir, on l'aura remarqué, une petite place est allouée à la phénoménologie de la RII du point de vue de l'utilisateur : celui-ci découvrira peut-être que la requête qu'accepte un moteur de recherche *n'a pas* à être réduite à deux ou trois mots seulement ; et que, même, il ne devra pas hésiter à « coller » dans le « petit espace » de la boîte de requêtes tout un passage d'un texte qu'il recherche. La thématique principale n'en est pas moins celle du fonctionnement « réel » des moteurs de recherche et donc des problèmes que ce fonctionnement pose. À cet égard, à nouveau, les auteurs amènent leurs lecteurs jusqu'à une frontière que l'on explore mais qu'on ne franchira pas véritablement : celle de l'*informatique* des chercheurs et des étudiants en informatique, dont le territoire, nous disent-ils, est celui de ces choses que "you might learn in a graduate computer science course"²⁸.

La sous-section suivante du chapitre *How to search* est intitulée *Evaluating the Results*. Les questions formulées ci-après en écho à cette sous-section poursuivent le travail

engagé pour dessiner les contours d'une culture possible de la RII. Cette fois, l'accent est mis sur la question de l'évaluation de la pertinence des résultats. Là encore les notions les plus simples sont évoquées, telles les notions, classiques en RI, de *precision* (précision) et de *recall* (rappel). L'exposé se fait ici très retenu. Ainsi, contrairement à des auteurs déjà cités (Boughanem & Savoy, pp. 40-41), les auteurs de *Web Dragons* ne disent-ils mot, en cet endroit, des grandes évaluations des systèmes de recherche d'information, par exemple celles impulsées par la *Text Retrieval Conference* (TREC).

110, § 4

Parmi les qualités d'un système d'indexation, quelle est la plus délicate à évaluer ?

110, § 5

En supposant qu'on puisse dire qu'un document est pertinent ou n'est pas pertinent par rapport à une requête, ce qui simplifie la réalité, sur quels critères peut-on juger de la qualité des résultats affichés par un moteur de recherche en réponse à une requête ?

111, § 1

Si par exemple on considère les 20 premiers résultats affichés par un moteur de recherche en réponse à une requête, qu'appelle-t-on *precision* de ces résultats ? Qu'appelle-t-on *rappel* ?

111, § 2

Que mesure la précision ? Que mesure le rappel ?

111, § 3

Que signifierait un rappel de 100 % ? Que pourrait-on dire alors de la précision ?

Que se passe-t-il en général lorsque la précision augmente ? Lorsque le rappel augmente ?

111, § 4

Si l'on ne considère par exemple que les 20 premiers résultats envoyés par un moteur de recherche en réponse à une requête, est-il possible de calculer la précision de ces résultats ? Comment ? Est-il possible de calculer le rappel ?

Un vaste panorama praxéologique

De la liste des 321 questions que nous avons associées au chapitre *How to search*, nous avons parcouru jusqu'ici les 71 premières, soit un peu moins du quart, qui font écho aux 11 premières pages du chapitre, qui en représentent un peu plus du quart. Nous commenterons maintenant plus rapidement la suite des questions, en renvoyant le lecteur à l'annexe 4 à cette quatrième partie de notre travail.

Ainsi qu'on l'a rappelé, la seconde grande section du chapitre, après la section *Searching text*, que nous venons d'examiner, s'intitule *Searching in a Web*. Les premières questions (de 111, § 5 à 112, § 4) manifestent à nouveau l'exigence de connaissances du passé que nous avons déjà soulignée : on y demande par exemple ce que faisaient les internautes *avant* l'apparition des moteurs de recherche, et de quand date cette apparition s'agissant des moteurs de recherche « modernes ». Le bloc de trois questions attaché au paragraphe 5 de la page 112, que nous rappelons ci-après, montre bien, nous semble-t-il, par le rapprochement de questions très différentes en *apparence* le souci d'une culture qui rende raison de la RII comme *fait social total*.

- # De combien de mots se composent la majorité des requêtes formulées ? Un ? Deux ? Trois ?...
Quelle est la requête la plus fréquente ?
 - # Dans une collection de documents où un certain mot n'apparaît dans chaque document qu'au plus une fois, selon quels critères sont rangés les résultats affichés par un moteur de recherche s'il utilise l'algorithme du cosinus ?
 - # Lors de leurs travaux pionniers, les créateurs du moteur de recherche Google, Sergey Brin et Larry Page, ont utilisé la requête consistant en l'unique mot *university*. De quoi cela leur fournissait-il un exemple (ou un contre-exemple) ?

La première a trait aux pratiques « observables » des usagers des moteurs de recherche. La deuxième nous ramène à l'algorithme du cosinus et aux conséquences qu'il a dans un certain type de cas particuliers. De la première à la deuxième question on passe ainsi, sans transition, de la phénoménologie à la nouménologie. La troisième question pourra paraître relever d'une érudition historique inappropriée ; mais on doit la regarder ici comme participant d'un certain « folklore », d'une culture « populaire » de la RII ²⁹.

La première sous-section de la section *Searching in a Web* est intitulé *Determining what a page is about*. Les questions portent sur le « matériel » relatif à une page web qu'examine un système d'indexation : on y évoque les abus possibles de la part des auteurs de pages Web et on s'y interroge sur la pertinence de prendre en compte les liens arrivant sur la page. Ici apparaît *de facto* une problématique importante, selon laquelle, en pénétrant dans les « arcanes » du fonctionnement des moteurs de recherche, l'utilisateur devrait devenir plus lucide sur le fonctionnement global de la recherche sur Internet.

La sous-section suivante, *Measuring prestige*, est fort longue et scindée pour cela en trois parties : nous lui avons associé 32 questions. Le premier bloc de questions (de 113, § 4, à

115, § 2) est entièrement consacré à la notion de « prestige » telle qu'elle est cristallisée mathématiquement dans l'algorithme *PageRank*. Du point de vue de sa mise en œuvre, on s'y interroge sur ce qui peut augmenter le PageRank d'une page Web donnée et, corrélativement, sur les « manœuvres » que d'aucuns ont pu être tentés de réaliser pour augmenter artificiellement leur PageRank. La seconde partie de la série de questions (de **115, § 3** à **116, § 3**) fait écho à un développement intitulé sans détour *Calculating PageRank*. Selon un paradoxe déjà souligné, ces questions sont entièrement consacrées à l'évocation, sans utilisation explicite de mathématiques, des... mathématiques de l'algorithme PageRank. C'est ce qu'illustrent notamment les deux questions que nous avons associées au paragraphe 5 de la page 115, reproduites ci-après.

Pour représenter mathématiquement les pages du Web et les liens entre ces pages, on utilise une matrice de connexion (*connection matrix*). Comment est constituée cette matrice ? Combien de lignes et de colonnes possède-t-elle ? Que représente ce nombre ? Que représente chacune des ses lignes ? De ses colonnes ?

Pourquoi la plupart des coefficients (*entries*) de la matrice de connexion sont-ils nuls ?

Là encore, par l'interdit (pesant sur les mathématiques) qu'ils se sont imposés, les auteurs engendrent des questions auxquelles leur propre texte ne saurait apporter une véritable réponse. Pour le contraste, nous avons reproduit ci-après un court passage d'un livre auquel les auteurs que nous suivons renvoient par ailleurs, dans la section *Notes and sources* de leur chapitre (Langville & Mayer, 2006, p. 31 ; voir aussi Chevallard, 2008).

It Must Be Beautiful: Great Equations of Modern Science [Farmelo, 2002] is a collection of 11 essays about the greatest scientific equations, equations like $E = hf$ and $E = mc^2$. The contributing authors were invited to give their answers to the tough question of what makes an equation great. One author, Frank Wilczek, included a quote by Heinrich Hertz regarding Maxwell's equation: "One cannot escape the feeling that these mathematical formulae have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them." While we are not suggesting that the PageRank equation presented in this chapter,

$$\pi^T = \pi^T(\alpha \mathbf{S} + (1 - \alpha)\mathbf{E}),$$

deserves a place in Farmelo's book alongside Einstein's theory of relativity, we do find Hertz's statement apropos. One can get a lot of mileage from the simple PageRank formula above—Google

certainly has. Since beauty is in the eye of the beholder, we'll let you decide whether or not the PageRank formula deserves the adjective *beautiful*.

Peut-être plus cruellement qu'ailleurs pour le lecteur profane, le non-franchissement de la frontière des mathématiques fait ici ressentir ses effets. À côté de cela on retrouve des questions portant sur le « folklore » des moteurs de recherche (combien d'itérations fallait-il à Brin et Page pour calculer les valeurs du PageRank lors de leurs premiers essais ?) ou sur certains paramètres du fonctionnement actuel de ces moteurs (quel est le temps nécessaire aujourd'hui pour calculer les valeurs de PageRank pour l'ensemble du Web ?). Dans cette veine on citera encore la question ci-après, que nous reproduisons sans commentaire.

Lors du recalcul du PageRank, ce ne sont que certaines pages et certains liens qui entraînent des mises à jour (appelées respectivement *node updates* et *link updates*). Est-il possible de procéder à un recalcul du PageRank limité le plus possible à ces pages et à ces liens ? Quels problèmes cela pose-t-il ? Ces problèmes sont-ils actuellement résolus ?

La dernière partie (de **116, § 4** à **117, § 5**) de la sous-section que nous suivons ici à travers les questions engendrées est intitulée *The Random Surfer*. Pour l'essentiel on s'y interroge sur le modèle probabiliste du PageRank introduit par Brin et Page pour résoudre les problèmes d'origine mathématique rencontrés lors de la mise au point de leur algorithme, et notamment (mais pas seulement) les problèmes engendrés par l'existence de pages dépourvues de liens sortants (*outlinks*). Tout autant qu'à propos du calcul du PageRank, il s'agit ici de pénétrer des phénomènes essentiellement mathématiques sans le secours des mathématiques. En même temps, la tentative des auteurs permet de prendre un recul qu'on n'a pas forcément quand, devant un modèle mathématique, on est tenté de ne voir que les mathématiques, et non plus le modèle. Considérons ainsi la question suivante, associée au paragraphe 4 de la page 117.

Que se passe-t-il si la probabilité de téléportation du surfeur aléatoire de Brin et Page est prise égale à 1 ? Que se passe-t-il si on lui donne la valeur 0 ? Quelle est la valeur donnée à ce paramètre par Brin et Page au début de leurs travaux (vers 1998) ? Quel intérêt peut-il y avoir à augmenter légèrement cette valeur ?

La première sous-question peut-être comprise sans grande connaissance du modèle : si, chaque fois qu'il arrive sur une page, le « surfeur aléatoire » se fait téléporter vers une autre page choisie au hasard – ce qui se produit si la probabilité de téléportation est fixée égale à 1 – , alors il ne suit jamais aucun lien existant, en sorte que la structure des liens du Web est entièrement ignorée, comme si elle n'existait pas ! Dans ce cas le PageRank n'accomplit pas ce pour quoi il avait été créé : exprimer quantitativement la structure des liens organisant le Web. La deuxième sous-question constitue un petit piège : si la probabilité de téléportation est nulle, l'algorithme PageRank, en effet, ne peut pas être calculé : c'est pour qu'il puisse l'être que l'idée de téléportation a été introduite. La troisième sous-question appartient, comme certaines sous-questions précédemment envisagées, au folklore de la RII. La quatrième et dernière question est, si l'on peut dire, beaucoup plus pointue : elle a trait à la façon dont la rapidité de convergence du calcul du PageRank dépend de la valeur assignée à la probabilité de téléportation.

La sous-section suivante (de **117, § 5** à **118, § 3**) apparaît sous le titre *Combining prestige and relevance*. Les questions correspondantes reviennent sur le délicat problème du classement selon la pertinence en évoquant les différents critères pris en compte, notamment quant à ces parties du texte qui apparaissent « sensibles » pour évaluer la pertinence. La dimension « historique » – sur une décennie, il est vrai – est encore et toujours présente, à propos notamment des abus de certains *webmasters* dans l'usage des métatags. Tout cela est regardé, à l'évidence, comme une composante incontournable d'une culture solide de la RII.

Les sous-sections qui suivent (de **118, § 4** à **122, § 4**), intitulées respectivement *Hubs and authorities*, *Findings hubs and authorities for a query*, *Using hubs and authorities* et *Discovering web communities*, concernent des développements de la « science du Web » non évoqués jusqu'ici et qui s'articulent autour de l'algorithme HITS de Jon Kleinberg, des notions solidaires de *hub* et d'autorité, à quoi s'ajoute la notion de graphe de voisinage d'une requête (*neighborhood graph of a query*). Contrairement à l'algorithme de PageRank, les scores de *hub* et d'autorité attribués à une page par l'algorithme HITS dépendent de la requête. Les questions que nous avons retenues poursuivent plus avant cette comparaison, en évoquant en particulier le phénomène problématique du *topic drift* qui affecte l'algorithme HITS. Mais surtout, elles interrogent la notion de « communauté Web » (*web community*), sans pour autant abandonner la rubrique de l'érudition folklorique, ainsi que cette question l'atteste : « Un moteur de recherche a pris le nom de *Teoma*. Que signifie le mot *Teoma* ? En quelle langue ? »

La sous-section suivante de l'exposé des auteurs de *Web Dragons* (de **123, § 1** à **123, § 4**) s'intitule simplement *Bibliometrics*. On s'y interroge sur l'*impact factor* et sur

quelques notions affines (*connection strength, influence weight, influence rate*), en appliquant alors à cet univers les notions de *hub* et d'autorité, tout en parcourant l'épaisseur praxéologique qui va des concepts de base à leur mise en œuvre toute pratique, qu'atteste cette question par exemple : « À quels exemples peut-on penser d'autorités du Web qui éviteront de se citer l'une l'autre (en envoyant des liens l'une vers l'autre) ? »

La section *Searching in a web* comporte encore deux sous-sections (de **124, § 1** à **127, § 5**), intitulées respectivement *Learning to rank* et *Distributing the index*. Le titre de la première sous-section peut paraître un peu énigmatique : ceux qui apprennent (*learn*), en l'espèce, ce sont des machines. Franchissant une nouvelle frontière, on touche ici au territoire de l'*intelligence artificielle* et, plus précisément, à l'*apprentissage machine* (*machine learning*) : ce sont des machines qui « apprennent » à assigner un rang à des pages Web, pour une requête donnée. Là encore, l'exposé précise *de facto* qu'une culture de la RII doit contenir des connaissances sur la façon dont les sociétés de moteurs de recherche créent leurs algorithmes de classement par de tels apprentissages, fait souvent masqué à l'utilisateur, qui est en particulier évoqué par les auteurs à propos du moteur de recherche MSN de Microsoft (devenu depuis Live Search). La connaissance de ce secteur des industries du Web doit permettre à l'utilisateur de dépasser les commentaires d'allure uniquement paranoïaque à propos de l'intérêt qu'ont ces sociétés pour le *clickstream* de leurs usagers, ambition dont témoigne bien la question suivante : « De quels arguments un usager peut-il tenir compte pour accepter qu'un moteur de recherche retienne des informations concernant son comportement sur le Web ? »

La sous-section *Distributing the index* conduit à des questions sur l'organisation « matérielle » de l'utilisation de l'index d'un moteur de recherche. À nouveau, il s'agit d'un sujet souvent « oublié » dans la littérature de la RII, alors même que – ainsi que nous l'avons vu dans la troisième partie de ce travail – sa connaissance permet d'élucider certains mystères de la phénoménologie de la RII. À cet égard on présentera deux exposés de haut niveau chacun dans son registre, mais dont l'un fait le silence sur ce sujet alors que l'autre l'explicite sobrement mais efficacement. Dans son livre *The extreme searcher's Internet handbook*, Randolph Hock (2007, pp. 63-64) décrit dans les termes suivants les « parties » constitutives d'un moteur de recherche – les *spiders* ou *crawlers*, l'*indexing program* et l'index, le *search "engine" itself* et enfin la *search page*.

The search "engine" itself—This is the program that identifies (retriever) those pages in the database that match the criteria indicated by a user's query. Another important and more challenging process is also involved, that of determining the order in which the retrieved records should be displayed.

This “relevance-ranking” algorithm usually takes many factors into account, such as the popularity of the page (as measured by how many other pages link to it), the number of times the search terms occur in the page, the relative proximity of search terms in the page, the location of search terms (for example, pages where the search terms occur in the title of the page may get a higher ranking), and other factors.

Par contraste, nommant *guichetier* ce que R. Hock désigne par l’expression *search “engine” itself*, Jacques Savoy écrit ceci (Boughanem & Savoy, 2008, p. 49) : « D’un point de vue technique, comme le volume d’information est considérable, le guichetier devra travailler avec une centaine d’index répartis sur des milliers de machines. » Une note infrapaginale précise encore : « On peut estimer qu’un moteur de recherche dispose d’environ 3000 ordinateurs personnels et d’environ 300 à 1000 fichiers afin de répondre à une requête. » Les questions que nous avons associées au texte des auteurs de *Web Dragons* dessinent ainsi un parcours d’étude et de recherche qui fait dépasser le premier exposé (celui de R. Hock) pour nous faire pénétrer, si peu que ce soit, dans le territoire où se situe le second (J. Savoy).

L’exposé des auteurs de *Web Dragons* se poursuit alors par une troisième section (de 128, § 1 à 131, § 3) intitulée *Developments in web search*. Ici, les questions formulées ont trait à des sujets a priori plus proches de l’usager que ne le sont les principes et la réalité de l’organisation et du fonctionnement des moteurs de recherche. Sous l’intitulé *Searching blogs*, certaines questions invitent par exemple à clarifier la signification des mentions *trackbacks* ou *blogroll* que l’on peut trouver sur un blog. La présence d’une sous-section intitulée *Ajax technology* peut a priori surprendre ; mais cette brève plongée au cœur de la technologie Ajax se fait bien *du point de vue de l’usager*, qui se trouve interpellé parfois tout à fait directement, comme il en va dans les deux questions suivantes.

Quelle évolution du Web en général, du point de vue de l’usager, peut-on attendre de l’utilisation croissante d’AJAX ?

La généralisation de l’emploi d’AJAX va généraliser du même coup l’envoi par les serveurs Web aux navigateurs, non pas de pages Web entières, mais de petits bouts (*nuggets*) de ces pages. Que penser de l’éventualité que les moteurs de recherche indexent ces « petits bouts » individuellement, sans référence à la page Web dans laquelle ils sont prévus pour apparaître ?

La troisième sous-section a pour titre *The semantic web*, le « Web sémantique », ce qui désigne aujourd’hui encore plus un projet qu’une réalité, mais un projet dont les idées-

forces méritent sans doute de s'intégrer à une culture de la RII attentive aux évolutions du Web. À nouveau, il y a là un domaine que nous n'avons guère évoqué jusqu'ici. Son exploration pourrait par exemple se donner pour objectif de répondre aux quatre questions que voici.

Qu'appelle-t-on *Resource Description Framework* (RDF) ? Comment est-il possible de formuler des énoncés portant sur des ressources du Web ?

Considérons l'énoncé suivant, structuré sur trois lignes successives :

`<http://www.everwonder.com/david/felixthecat/>`

`<http://www.love.example.org/terms/reallyLikes>`

`<http://fishhead.homestead.com/>`

De quelle façon peut-on regarder cet énoncé comme signifiant que Félix aime le poisson ?

130, § 3

Un énoncé RDF a-t-il *a priori*, grâce à sa structure, plus de sens pour une machine que du texte sur une page Web ?

131, § 1

Étant donné une communauté d'agents programmés (sur le Web), qu'appelle-t-on une ontologie (*ontology*) ? Quelle est le but assigné à une ontologie ?

Les questions relatives au Web sémantique terminent la troisième section de l'exposé intitulée *Developments in web search*. La quatrième et dernière section (de **131, § 4** à **141, § 5**) est, on le sait, intitulée *Birth of the dragons*. Les Dragons, ce sont, bien sûr, les moteurs de recherche, à propos desquels les auteurs, dans leur préface, précisent ceci.

We refer to search engines as “web dragons” because they are the gatekeepers of our society’s treasure trove of information. Dragons are all-powerful figures that stand guard over great hoards of treasure. The metaphor fits. Dragons are mysterious: no one really knows what drives them. They’re mythical: the subject of speculation, hype, legend, old wives’ tales, and fairy stories. In this case, the immense treasure they guard is society’s repository of knowledge. (p. XV)

Les auteurs tirent parti de la métaphore zoologique des dragons pour segmenter leur exposé : tout d'abord, *The womb is prepared*, la matrice est préparée, ensuite *The dragons hatch*, les dragons sortent de l'œuf ; on voit alors apparaître *The big five*, la portée principale, avant de pénétrer *Inside the dragon’s lair*, dans l'ancre des dragons. Les questions, ici, interrogent

l'histoire de l'Internet, s'arrêtent sur le rôle du protocole FTP, l'apparition de Gopher, avant la naissance du World Wide Web : elles permettent de suivre les avancées et les vicissitudes d'une époque appréhendée à travers ce que nous avons rassemblé, dans la troisième partie de ce travail, au chapitre 9, sous l'intitulé *Personnes, institutions, dispositifs*. À nouveau on voit se dessiner ce qui est une exigence de toute culture, le partage de la connaissance de certaines figures, de certains événements, de certains commentaires sur ces figures et événements, ce dont témoigne par exemple cette question : « Le moteur de recherche OpenText a disparu très vite. Pour quelles raisons ? Quel enseignement a-t-on pu tirer de son échec ? » Plus généralement, c'est toute l'épopée des moteurs de recherche qui est ainsi activement interrogée, dans ses principales lignes de force comme dans ses détails qui, pour paraître parfois minuscules, peuvent pourtant éclairer l'utilisateur : quand et pourquoi, par exemple, "AskJeeves" a-t-il perdu son "Jeeves", et d'ailleurs qui était ce Jeeves ?... La visite de l'antre des dragons est, elle aussi, fort instructive : on peut y apprendre par exemple que le nombre de mots composant les requêtes que Google traitaient chaque jour vers le milieu de l'année 2006 représentaient plusieurs fois le nombre de mots contenus dans l'*Encyclopedia Britannica* – combien de fois, plus précisément, telle est la question posée. On pourra rapprocher une telle interrogation, dont la réponse est susceptible de « fixer les idées », de ce passage emprunté encore à J. Savoy.

... si l'information stockée sur papier augmente, celle mémorisée sur Internet représente un volume bien plus important. [...] On peut estimer qu'un livre comprend environ 1 million de caractères [Note. On peut estimer qu'une page comprend environ 2000 caractères (ou octets) et qu'un livre possède en moyenne 500 pages. Nous avons ainsi 1 000 000 caractères par livre, soit 1 MO] et, en mémorisant un caractère par octet, on obtient la valeur de 1 million d'octets ou 1 MO (10^6 octets). La production annuelle de livres sur la planète peut être estimée à 1 million (1 M), ce qui nous donne $1 \text{ MB} \times 1 \text{ M} = 1 \text{ TB}$ (10^{12} octets). Si l'on désire également mémoriser les journaux et autres documents administratifs d'une part et, d'autre part, conserver les cent dernières années, on peut multiplier le chiffre précédent par 1000, ce qui nous donne un volume total estimé à 1000 TB (ou 10^{15} octets) pour mémoriser toute l'information sur support papier.

On a estimé que l'information disponible sur la Toile en 2002 occupait un volume de 167 TB [...]. Cette valeur comprend uniquement la partie visible tandis que le « web invisible » renfermerait un volume nettement plus important, soit 91 850 TB. Le terme « invisible » indique bien que toute l'information n'est pas toujours facilement ou gratuitement accessible comme, par exemple, les pages générées dynamiquement (consultation d'un horaire). Dans cette partie, on retrouve également les multiples pages dont l'accès n'est pas autorisé à un moteur de recherche (comme, par exemple,

les collections d'articles des grandes maisons d'édition, les fonds documentaires des bibliothèques numériques comme celles de l'ACM ou de l'IEEE ou les multiples dépôts d'information des entreprises et administrations). Comparé au support papier, le volume de stockage sur le web s'avère dix fois plus important et cette tendance va encore s'accroître durant la prochaine décennie. (pp. 45-46)

Bien d'autres aspects pourraient encore être cités que nous laisserons le lecteur découvrir dans l'annexe 4. Terminons cette visite d'un panorama dont nous n'avons présenté que quelques vues par deux questions qui, l'une comme l'autre, font explorer la raison de l'emploi par les auteurs parlant des "web dragons" de l'expression *treasure trove* – laquelle est, au reste, une « anglicisation » de l'anglo-normand *tresor trouvé* (1190), qui lui-même traduisait le latin *thesaurus inventus*, « trésor trouvé ».

Derrière leur activité visible (répondre aux requêtes des usagers), les moteurs de recherche mémorisent toutes les requêtes qui leur sont adressées par chacun des usagers. Pourquoi ces données peuvent-elles être regardées comme constituant un trésor ?

Quelle différence peut-on voir entre les informations accumulées par les moteurs de recherche à propos des requêtes qui leur sont adressées et les informations publiées dans la presse et les ouvrages savants ? Sur quels aspects du fonctionnement de nos sociétés les moteurs de recherche ont-ils des informations uniques ?

Dans l'exposé que nous venons d'examiner, on ne trouve, sauf exception, de questions qu'en puissance. Nous en avons extrait – « excrit » – un certain nombre, dont le texte objet de cette opération ne saurait, croyons-nous, refuser la paternité : prendre le risque de parler, c'est prendre le risque d'être interrogé sur son dire. Nous poursuivrons maintenant le travail engagé sur un autre matériau, à l'aide d'autres techniques.

Points de vue de formateurs

L'ouvrage dont nous allons extraire des questions a plusieurs fois été mentionné dans ce qui précède : intitulé *Teaching web search skills*, avec pour sous-titre *Techniques and strategies of top trainers*, il est dû à Greg R. Notess (2006). Comme le titre l'indique, le problème pris en charge dans ce livre n'est pas tant celui d'établir un inventaire des "Web search skills" que d'examiner les points de difficulté – et les solutions qui peuvent leur être apportées – de l'enseignement des praxéologies de la RII. Le public auquel l'auteur s'adresse est donc formé

de ceux qui sont amenés à enseigner la RII, type d'enseignement qui semble aujourd'hui bien développé aux États-Unis. Significativement, la page de couverture de l'ouvrage reproduit une appréciation péremptoire autant qu'élogieuse due à Roy Tennant de TechEssence.Info³⁰ : "If you teach Web searching you need this book. Steal it if you must, but don't make the mistake of putting it down and walking away." La dédicace de l'auteur est elle-même tout à fait parlante quant à l'objet de son livre.

To all Web searching teachers, trainers, instructors, and workshop leaders, especially those who have talked with me about search engines, taken my classes, provided feedback on my Web site, and contributed ideas to this work. May you all have continued success in your training sessions!

Ouvrage d'un formateur adressé aux formateurs, il relève de la didactique « pratique » de la RII et comporte les treize chapitres suivants.

1. Web Searching and the Teaching Paradox / 2. Understanding Our Audience / 3. Choosing the Instructional Session Types / 4. Online Tutorials: Friend or Foe? / 5. Organization: Focus, Length, and Goals / 6. Terminology / 7. Content: Web Search Features / 8. Content: Primary Concepts / 9. Additional Content: Image Searching, News, Usenet, and More / 10. The Search Engines / 11 Creating a Framed Workshop Web Page / 12. Presentation Tips, Tricks, and Shortcuts / 13. Anecdotes, Examples, Exercises, Stories, and Tales from the Trenches

Nous reviendrons plus précisément sur le contenu de certains chapitres ; mais il doit être clair que les questions de RII n'y apparaissent qu'en tant qu'elles soulèvent des questions d'enseignement ou de formation. Comme le souligne encore une appréciation reproduite en quatrième de couverture, l'ouvrage regorge ainsi de *training tips*, plutôt que de *searching tips*. C'est donc par une observation « oblique » que l'on pourra y repérer les questions de RII sous-jacentes aux questions d'enseignements de la RII qui y sont invoquées.

Le livre de Greg Notess constitue un témoignage précieux sur la formation à la RII pratiquée outre-Atlantique au sein des institutions universitaires et autres : l'auteur a en effet tiré profit, non seulement de sa propre expérience, mais encore de celle de douze autres formateurs chevronnés en matière de *Web search*. L'auteur lui-même est *reference librarian*, « bibliothécaire de référence », à l'université d'État du Montana et, dans différents contextes et « formats », a enseigné le "Web searching" et plus généralement l'Internet depuis 1991³¹. Trois fois lauréat de l'*Information Authorship Award* – attribué chaque année à l'auteur d'un article publié dans divers magazines (*ONLINE*, *DATABASE*, etc.) –, il est aussi l'éditeur d'un

site à forte notoriété, *SearchEngineShowdown*, et a ouvert à l'automne 2006 un blog consacré au *screencasting* (capture vidéo d'écrans), outil dont l'importance va croissant dans les formations en ligne ³².

Qui sont les douze « invités » de Greg Notess ? Plusieurs d'entre eux sont des bibliothécaires d'université. Ainsi en est-il de *Joe Barker*, qui exerce son art depuis 1983 à l'université de Californie à Berkeley, où, étudiant, il avait obtenu en 1975 un *MLIS*, c'est-à-dire un *Master of Library and Information Science*. Joe Barker a une activité d'enseignement soutenue "about the Internet and how to find information on the World Wide Web" (p. 232). C'est ainsi que, du 26 août au 22 septembre 2008, il devait animer, sous le titre *Essential Tools & Tips for Successful Web Searching*, un atelier en ligne conçu autour des questions suivantes ³³.

What goes on in search engines? How do they "think" about searching? / What are today's best search engines after Google? Why are they the best? / What are the most useful search commands? Will they work beyond Google? / How can I create searches that effectively zero in on just what I need? / When is it time to turn from web search engines to subject directories (like lii.org)? / How can I find good, specific directories on subjects my users or I need to explore? / Is there a best search strategy for searching for a variety of information? / How do I deal with the need to evaluate all the stuff on the web quickly?

Son point de vue, tel que le rapporte Notess, mérite d'être cité.

In many ways, he feels that what he knows about Web searching is more directly valued in public libraries' reference and research than at the university, where students are ushered to "scholarly" sources and encouraged not to use the Web. Joe feels the Web contains valuable information that is often more up-to-date than published sources. (p. 233)

Paul Barron est également bibliothécaire ; il a exercé des fonctions de *reference librarian* et d'archiviste, ainsi que de directeur de la technologie pour un ensemble d'écoles dans l'État de Virginie où il réside et travaille. Il a conduit de nombreux ateliers sur le *Web searching* à l'adresse de publics divers (notamment d'enseignants et de bibliothécaires du secondaire et de l'université). Une page Web intitulée *Internet searching* (Barron, 2008) présente sur ce sujet des ressources qu'il recommande : il s'agit de sites ou d'ouvrages imprimés dont nous retrouverons certains des auteurs plus loin (Joe Barker, Phil Bradley, Randolph Hock, Gary Price, Danny Sullivan et bien sûr Greg Notess). *John Ferguson*, qui a pris sa retraite en 2003,

était *reference librarian*, et *teaching librarian* à la Richland College Library, après 19 ans d'intense activité professionnelle autour des pratiques de l'Internet. *Alice Fulbright* est depuis 1992 également *reference librarian* au Richland College Library (après avoir obtenu en 1990 un MLIS à l'université du Texas du nord). À son propos Notess écrit (p. 237): "In the past, Alice taught a continuing education course titled Internet Searching, geared to the needs and interests of adults in the community who wanted to learn Internet searching techniques and become acquainted with many valuable Web sites." *Jeff Humphrey* a, lui, obtenu un MLS (*Master of Library Science*) à l'université d'Indiana. Il a exercé comme *reference librarian* et comme formateur dans diverses bibliothèques publiques avant de rejoindre l'*Indiana Cooperative Library Services Authority* (INCOLSA), service public qui apporte son concours à plusieurs centaines d'institutions publiques de l'État d'Indiana, où exercent quelque 2200 bibliothécaires. Dans ce cadre, Jeff Humphrey intervient comme *Internet trainer*, "offering courses in searching, Web design, and HTML" (p. 238). Notess précise encore à son propos (pp. 238-239) : "He has delivered workshops and presentations at local, state, and national conferences covering all aspects of the Internet and its use in libraries." *Diane Kovacs*, actuellement présidente de *Kovacs consulting—Internet and Web training*, est titulaire notamment d'un MLIS de l'université de l'Illinois (1989) et a quelque 15 années d'expérience de l'enseignement des *Web research skills*. Elle a écrit de nombreux ouvrages, de 1995 (*The Internet trainer's guide*) à aujourd'hui : le dernier en date, paru en 2007, a pour titre *The virtual reference handbook*. *Sheila Webber* est actuellement *senior lecturer* au département d'*Information studies* de l'université de Sheffield (Angleterre). Spécialiste de l'*Information literacy*, elle appartient, comme les formateurs précédents, au monde des bibliothèques et est en particulier membre du *Chartered Institute of Library and Information Professionals* (CILIP). Elle enseigne le "Web searching as part of a large college level course" (Notess, p. 46) et, à ce propos, déclare : "I feel that Web search ought to be seen by the students as part of searching in general..." (p. 46). *Phil Bradley*, qui lui aussi vit et travaille en Angleterre (à Feltham, dans le Middlesex), s'est établi comme *Internet consultant* indépendant au milieu des années 1990. Ses activités sont multiples : à côté de son site Internet et de son blog, il assure par exemple une chronique mensuelle dans le magazine du CILIP, *Update*. De lui Greg Notess peut écrire (p. 234) : "He spends much of his time teaching groups (librarians, doctors, lawyers, etc.) about various aspects of the Internet and has probably instructed more people in the U.K. than anyone else in how to use the Internet." *Gary Price* a obtenu un MLIS de l'université d'État Wayne à Detroit (Michigan). De 1995 à 2001 il a été *reference librarian* à l'université George Washington. Il est actuellement directeur du service *Online Information Resources* de Ask.com. Son activité est multiple : il est l'éditeur du site *ResourceShelf* qui

propose des ressources dans le domaine des bibliothèques et de l'information, et en particulier sur les moteurs de recherche ; il a créé le site DocuTicker, qui fournit des informations quotidiennes en provenance des milieux institutionnels et il est en outre l'auteur de nombreuses publications, notamment dans le *Washington Post*, le *Guardian* et le *Chronicle of Higher Education*. Il est surtout, avec Chris Sherman l'auteur de *The Invisible Web* (2001), ouvrage préfacé par Danny Sullivan. Dans son dialogue avec Notess, Gary Price indique typiquement : "If you're doing a 'Web only' class, it is worth a few minutes... to focus on what the Web can't deliver and let folks know that libraries can offer many of the materials you can't find on the Web, such as remotely accessible databases and things commonly referred to as books" (p. 77). Pour ce qui la concerne, Rita Vine a été durant 20 ans bibliothécaire et formatrice dans ce domaine. Elle est co-fondatrice de Workingfaster.com, qu'elle préside, et dont le site dit ceci : "Workingfaster.com is a North American leader in web search skills training. We provide custom training solutions and high quality web search tools to help people become better internet searchers and trainers." Rita Vine est la créatrice d'une méthode intitulée "Keep It Simple Searching" ; elle a de plus contribué à la conception d'un outil remarqué, le *Search Portfolio*, que le site Workingfaster.com présente en ces termes (où l'on notera particulièrement la mention *hand-picked*) : "Simply the best starting points for effective web searching, hand-picked by leading web search experts." Dans un texte intitulé "Going Beyond Search Engines for Faster and Smarter Web Searching" (Vine, 2004), cette spécialiste propose neuf "tips to save time and find better information". À titre d'illustration, nous les reproduisons ci-après.

1. Use more search engine. / 2. Use high quality, filtered directories for resource discovery. / 3. If it isn't a search engine, don't search it like one. / 4. Be wary of sites with heavy use of advertising. / 5. Add qualifier words like *resources*, *links* or *directory* to a simple search engine search to bring link lists to the top of the results list. / 6. Use the advanced search features of search engines for selected searches. / 7. Follow the news about the search business. 8. Serious searchers need to understand the basics of how advertising impacts search results and the many ways in which the major commercial search properties each feed content and ads to secondary properties and even to each other. 9. Remember that just because it's popular doesn't mean it's good.

Nous avons déjà rencontré le nom de *Danny Sullivan*, le créateur de SearchEngineWatch.com, regardé aux États-Unis comme un « gourou des moteurs de recherche » (*search engine guru*). Cette réputation est fondée sur un travail pionnier (1996) intitulé *A Webmaster's Guide to Search Engines* (Sullivan, 2006). Bien que son activité concerne principalement la

connaissance des moteurs de recherche, notamment à travers une série de conférences intitulées *Search Engine Strategies*, il est aussi un formateur en matière de RII – Notess précise à cet égard (p. 242) que Danny Sullivan “gave training sessions on Web searching at the British Library for a number of years and continues giving training sessions for a large UK company.” *Randolph Hock*, a été *reference librarian* au MIT (Massachusetts Institute for Technology). Il est l’auteur de plusieurs ouvrages (dont Hock, 2007), ainsi que de nombreuses publications dans des revues professionnelles. “Randolph (Ran) Hock has trained more than 10.000 online researchers over the past 20 years”, écrit Notess (p. 237), qui ajoute (p. 238) : “Through his company, Online Strategies, Ran develops and presents customized on-site seminars in how to effectively and efficiently make use of Internet resources.”

Ce sont ces douze personnes, dont nous venons de dresser un rapide portrait, que Greg Notess a, indique-t-il (p. XVIII), “interviewed in depth”, ajoutant (p. XVII) : “I have talked extensively with these individuals and surveyed their approaches to teaching Web searching”. Plusieurs des interlocuteurs de Notess se connaissent non seulement de réputation – ils constituent, selon l’auteur (p. XVII), “a small but prominent group of expert Web search engine trainers” –, mais aussi de façon plus directe, pour avoir travaillé ensemble. C’est ainsi que les 1^{er} et 2 juin 2005 s’est tenu au CNIT (Centre National des Industries et Techniques, la Défense, Paris), sous le titre *WebSearch Academy*, un stage de formation sous-titré *Mastering Internet research* où se retrouvèrent comme formateurs Greg Notess, Ran Hock et Phil Bradley. Ce stage était présenté dans les termes suivants (Information Today, 2005) : “Designed for those whose job responsibilities focus on regularly using Internet research tools, WebSearch Academy features a faculty of search professionals who write about and teach Web searching.” Un tel aréopage issu essentiellement du monde des bibliothèques et de la *library science* reste marqué par le passé vénérable de ce domaine d’activité – on l’a vu notamment à l’importance donnée aux annuaires (par opposition aux moteurs de recherche) par Joe Barker ou Rita Vine. De même, on retrouvera un thème familier dans l’une des interventions (lors des deux journées parisiennes de la *WebSearch Academy*) de Phil Bradley, intitulée *Quality versus Quantity: Weeding out the Bad Stuff*, et que son auteur présente en ces termes : “Web search results can be misleading, biased, suspect, inappropriate, and even dangerous. This session spells out the warning signs that researchers should look out for.” Il n’en reste pas moins que Notess a réuni dans son livre quelques-uns des *formateurs* à la RII – et non, certes, des chercheurs en RI – sans doute les plus au fait des plus récentes innovations touchant les outils de la RII. Nous allons examiner maintenant sur un exemple ce que le rapprochement de leurs points de vue peut nous apprendre sur les conditions et contraintes de l’émergence possible d’une culture de la RII.

Synnomie ou idionomie ? Le cas des booléens

Ainsi que le suggère le sommaire de son livre, Greg Notess consacre plusieurs chapitres à la question des contenus d'une formation à la RII. Il convient de préciser ici qu'on ne doit pas confondre la question du contenu de *formations* à la RII avec une *culture* de la RII : les invités de Notess baignent sans doute dans une culture de la RII qui leur est largement commune ; et c'est plongés dans cette culture partagée – qui permet des échanges souvent à demi implicites – qu'ils vont répondre à la question des contenus d'une formation possible à la RII. Cette culture commune de la RII, à son tour, ne doit pas être confondue avec une culture de l'*enseignement* de la RII – qui n'est ici, nous le verrons, que très imparfaitement partagée. Cela souligné, voici la table des matières détaillée de l'un des chapitres consacrés aux contenus de formation, et sur lequel nous nous arrêterons maintenant.

Chapter 7 Content: Web Search Features

Search Features

Phrase searching / Boolean searching / Truncation / Field Searching / Limits / Stop Words and

Case Sensitivity / Sorting

The Teaching choice

C'est la question du *Boolean searching*, de la recherche booléenne, que nous examinerons plus précisément à titre d'exemple. Greg Notess indique d'abord que l'enseignement des opérateurs booléens a dû être envisagé depuis deux décennies dans des conditions et sous des contraintes changeantes. Alors que la connaissance des booléens était une partie indispensable de la "online searching instruction in the 1980s and early 1990s" (p. 120), l'avènement des moteurs de recherche a modifié la situation, au point que, "in the 1990s, some pundits argued for getting rid of Boolean searching altogether, painting glowing pictures of the day when Boolean concepts could be abandoned by all searchers" (p. 121). Ces variations, qu'ont connues la plupart des formateurs interrogés par Notess, ne les empêchent nullement d'être les porteurs de partis pris individuels souvent tranchés vis-à-vis de la question à laquelle l'auteur donne la formulation concise suivante : "what can we teach users about Boolean searching on the Web?" Pour ce qui la concerne, *Alice Fulbright* dit enseigner les opérateurs booléens "because students need to know how to broaden and narrow their searches", explication qui a le mérite de préciser la raison d'être – à ses yeux – des praxéologies correspondantes. De son côté, *John Ferguson* aboutit à la même conclusion,

mais pour des raisons différentes, davantage ancrées dans le monde des bibliothèques, à savoir que les opérateurs booléens sont “still used to search both subscription databases and online catalogs” ; pour le reste, il admet que les choses changent, précisant à cet égard : “We still teach basic Boolean searching, but less and less when it comes to using Web search engines.” Et d’ajouter cette observation définitive : “Advanced Boolean searching—such as nesting—is no longer taught.” Plus nuancé, *Phil Bradley* adopte simplement des choix différents selon qu’il s’adresse à un public de gens de bibliothèques ou à des étudiants qui ne sont pas des “information professionals” : dans le premier cas, il enseigne les booléens “if required, and sometimes it is” ; dans le second, il se limite à l’enseignement de l’interface de « recherche avancée » des moteurs. *Ran Hock*, lui, se montre plus franchement partisan de ne pas écarter les booléens du corpus enseigné, considérant que “the concept of Boolean, even if AND, OR, and NOT are not used, is still essential to effective serious research”, et que les opérateurs + et -, en tout cas, sont indispensables “in order to understand the full capabilities of certain tools” (p. 123). Face à ces positions relativement proches, d’autres parmi les « invités » de Greg Notess sont beaucoup plus réticents. *Joe Barker* a longtemps enseigné les booléens car, pensait-il, “people had to know this to survive” (p. 122). Mais parce que tant de personnes semblent incapables de formuler correctement une requête booléenne, il a réduit son enseignement au seul opérateur OR (tout en proposant une fiche pour ceux qui veulent aller plus loin). *Rita Vine* est, quant à elle, franchement opposée à l’enseignement des opérateurs booléens, comme elle l’avoue abruptement (p. 122) : “I know that I’m probably the only one out there who doesn’t teach Boolean, not even to expert searchers. We used to in the old days (mid-’90s) but then stopped.” À ses yeux, l’usage des opérateurs booléens est souvent un symptôme de ce que l’outil de recherche utilisé est simplement mal choisi : “If you are picking around with keywords and combinations and synonyms, [...] chances are you shouldn’t be using a search engine anyhow, and you are better off with a resource discovery tool.” *Greg Notess* lui-même semble plus flexible. S’il reconnaît que “most of us agree that sophisticated Boolean abilities are rarely needed for Web searching”, il fait une exception dans ses stages de formation pour les usagers « avancés » qui expriment le souhait de pouvoir en certains cas utiliser les commandes booléennes. Pour ces usagers, précise-t-il, il continue “to teach how and when to use Boolean searching and which search engines support it”. Le cas de *Danny Sullivan* est autre : il n’enseigne pas les opérateurs booléens, mais simplement, sous la dénomination *search engine math*, les opérateurs + et - déjà évoqués. Sur une page de son site SearchEngineWatch.com intitulée “Boolean Searching” et datée du 26 octobre 2001, il écrit au reste ce que voici (Sullivan, 2001b, § 1).

Boolean search commands have been used by professionals for searching through traditional databases for years. Despite this, they are overkill for the average web user. The commands described on the Search Engine Math page provide the same basic functionality as Boolean commands and are also supported by all the major search services. If you are new to searching, start off learning how to search better by first reading the Search Engine Math page, rather than trying to learn Boolean commands. I'm certain you'll find it easier.

À cela il ajoute aussitôt (§ 2) : “In fact, many professionals might benefit by abandoning Boolean commands when using web search engines.” L'introduction de la page “Search Engine Math” est, de ce point de vue, non moins directe (Sullivan, 2001c, § 1).

Forget power searching. Don't worry about learning to do a “Boolean” search. All most people need to know is a little basic “search engine math” in order to improve their results. Come learn how to easily add, subtract and multiply your way into better searches at your favorite search engine. The information below works for nearly all of the major search engines.

Cette assurance – comme celle de plusieurs des interlocuteurs de Notess – est frappante, alors même que, pourtant, les positions des protagonistes divergent entre elles ou changent dans le temps. À la fin des années 1990, observe Notess (p. 125), “many Web trainers considered the + - system a convenient way to teach Boolean, but some absolutely hated it”. Lors de l'enquête effectuée pour son livre, il sera encore surpris par les réactions véhémentes – et définitives – de certains de ses collègues à propos de leur utilisation éventuelle des opérateurs + et -. *Joe Barker*, par exemple, s'exclame : “Never ! Useless unless the default is OR. I wish + had never been invented...” Quant à *Rita Vine*, elle accueille la question par un “Ptooi!” – « Grand Dieu ! »³⁴.

Ce qui frappe ici, notamment par rapport à notre questionnement – quelles questions de RII installer au cœur d'une culture partagée ? –, c'est, précisément, l'apparent égotisme de ces formateurs : là où la construction d'une culture commune, acceptée de tous suppose un processus de *synnomie* – le fait de bâtir une loi (*nomie*) que tous reconnaissent (*syn*) –, on rencontre ce qui est au mieux le point de départ éventuel d'un tel processus, état dans lequel chacun reste assujéti à sa propre loi, en revendiquant même fortement, pour certains, cette *idionomie*³⁵. Le *forum* de questions (d'enseignement de la RII) est ici artificiellement créé (par Greg Notess) et n'est nullement spontané. Mais, au-delà, un autre aspect doit maintenant être souligné.

Dans son livre, Notess soulève un certain nombre de questions d'enseignement à la RII – c'est en cela que son livre peut être regardé comme un *forum* en la matière – puis recueille des réponses R^\diamond auprès de ces « institutions » que sont les douze formateurs qu'il interroge. À partir de là, en combinant ces réponses avec ses propres choix, il esquisse des réponses R^\heartsuit , toujours à propos de questions Q d'enseignement à la RII. S'agissant de la question de l'enseignement des booléens, ainsi, il évoque sa réponse personnelle (p. 124) : “ At a very basic level, to teach the automatic default AND used by most search engines, I instruct students to add more words to find fewer results. My motto here is ‘More words equals less results.’” Ce slogan, ajoute Notess, marche habituellement très bien auprès des étudiants. Mais il avoue aller plus loin sur le sujet, cette fois à propos du booléen OR.

Even so, I like to point out that other search systems (library catalogs, bibliographic databases, directories, or site search engines) may behave differently and find more results as more query terms are added. Here is a simple technique students can use to check a given search tool: First, search on one word and note the number of results. Next, add a second word and search again. If the number goes down, the search system is performing like a Web search engine (and hence using Boolean AND). If the number increases, then it is finding more results and thus doing a Boolean OR operation. (p. 124)

Ainsi que l'illustre ce développement, et selon le terme repris dans l'*Introduction* à cette partie de notre travail, le livre de Notess est une *boutique* offrant des réponses – des praxéologies – en matière d'enseignement de la RII. En revanche, son but, on l'a dit autrement plus haut, n'est nullement d'offrir des réponses à des questions *de RII*, non plus d'ailleurs que d'être un *forum* de questions de RII. Forum et boutique en matière de RII, ce livre l'est pourtant, mais *marginale*ment ; ou, plus précisément, il peut être *utilisé* comme forum et boutique en la matière. C'est ainsi que le lecteur dont l'équipement praxéologique serait à cet égard pris en défaut pourra apprendre à la lecture de ce qui précède comment explorer, sur un point très simple, le comportement d'un moteur de recherche : fonctionne-t-il avec des AND ou des OR ? Dans un développement que Notess a intitulé le “The + and - dilemma”, à propos du remplacement, dans beaucoup de moteurs de recherche, de l'usage implicite de OR par l'usage implicite de AND, il est amené à écrire ceci.

On top of the change in the default operator, the + began to be used for other functions. Google was the first to use it to force the inclusion of stop words (frequently occurring words such as *of*, *the*, and *is*, which are usually ignored by the search engine), with other search engines following suit. (p. 125)

De telles indications, qui apportent des éléments de réponse à des questions de RII, sont motivées ici par les besoins de l'analyse des réponses R^\diamond recueillies touchant des questions d'enseignement de la RII ; mais elles apportent aussi des lumières au lecteur qui, sur le point évoqué, en manquerait : le signe + présente ainsi l'intérêt de rendre les *stop words* (les mots vides) "searchable". En fait, le paysage praxéologique brossé par Notess, dont des éléments pourront s'intégrer à l'équipement praxéologique du lecteur, se laissent décrire comme suit.

So, depending on the search engine being used, the functionality of the + has changed. It may require a term to be present, make a stop word searchable, or have no effect at all. Also keep in mind that some specialized search engines, including site search tools, still use the + - system and default to OR. (p. 125)

Pour le dire en termes familiers, on peut apprendre de la RII dans un livre sur *l'enseignement* de la RII : la boutique ouverte par Notess, qui offre principalement des praxéologies d'enseignement de la RII, propose aussi une petite offre en fait de praxéologies de la RII elle-même. Pourtant cette observation ne doit pas masquer un phénomène essentiel au fonctionnement des institutions didactiques : si la matière à enseigner est problématique, elle est supposée ne l'être *que* pour les enseignés eux-mêmes. Dans la noosphère, la fiction prévaut que, pour les formateurs, ce caractère problématique n'existerait pas ou, du moins, n'aurait pas de pertinence. On suppose par exemple que chacun connaît l'utilité du signe + dans le maniement des mots vides : la chose irait de soi, ferait partie de la culture commune aux formateurs, en sorte que, entre eux, il suffirait de s'y référer allusivement, à demi-mot.

Il résulte de cela que, ce qui n'advient jamais ou presque, dans ces circonstances, c'est ce que nous nommerons le *moment socratique*, celui du *Qu'est-ce que... ?*, où l'on interroge les mots pour connaître les choses auxquelles ils renvoient. À aucun moment, par exemple, la question « Mais qu'est-ce qu'au juste que le *Boolean searching* ? » n'est posée ; encore moins fait-elle l'objet d'une réponse dont la validité et les limites seraient explicitement interrogées. Et il en va de même de cette autre question solidaire : « Qu'est-ce qu'au juste qu'une combinaison booléenne (*Boolean combination*) ? » La fiction, rappelons-le, est que ces choses-là, pour cette raison qu'on en parle, qu'on formule des assertions assurées à leur propos, sont supposées connues, « bien connues » même, en sorte qu'un interdit pèse sur leur

mise en question. Il en découle que, sauf exception, les questions correspondantes – qui seraient des questions *de RII* – ne seront pas posées ; et que, subséquemment, les réponses éventuelles à de telles questions n'apparaîtront pas comme telles, même s'il est possible au lecteur de regarder certains développements du texte comme des réponses à des questions qu'il aura dû produire par lui-même. Nous reviendrons au chapitre suivant sur la question des *réponses* et de l'offre praxéologique en matière de RII ; notons ici simplement, en manière d'illustration, que la page consacrée par Danny Sullivan au *Boolean searching* (Sullivan, 2001b) ne donne ainsi que quelques indications minimales sur les commandes OR, AND, NOT, NEAR et sur le *nesting* le plus simple, pour lequel il fournit l'exemple suivant : **impeachment AND (clinton OR johnson)**. Nous nous arrêterons maintenant sur la question des *questions* de RII.

Questions de RII

On a dit que la noosphère de ce qui existe en matière d'enseignement de la RII semble n'offrir que très rarement des forums de questions (au sens de la TAD) relatifs à la RII. De tels forums existent pourtant, par exemple dans le cadre de « forums » en ligne (qui ne sont cependant pas si nombreux à faire une place explicite aux questions de RII). C'est ainsi que le site *Abondance* d'Olivier Andrieu propose un forum intitulé « Recherche d'informations sur le Web ». L'examen des questions qui y sont abordées montre que beaucoup de ces questions relèvent moins de la RII que du référencement proprement dit. Cela noté, un petit nombre de questions auxquelles un tel forum permet d'être formulées sont en effet typiquement des questions de RII. Ainsi en va-t-il avec celle que nous reproduisons ci-après ³⁶.

Bonjour,

J'ai constaté (avec Google) que l'ordre des mots clés influence sur le nombre de résultats de recherche.

Exemple : jardin versailles (1 350 000 résultats)

versailles jardin (1 360 000 résultats)

Quelqu'un pourrait m'expliquer pourquoi?

Merci d'avance

Lila1

Cet exemple illustre un fait que nous prendrons dans la suite comme un trait fondamental de l'écologie sociale de la RII : d'abord, il y a des personnes à qui se posent des questions de

RII ; ensuite, il y a des personnes qui se posent des questions de RII ; enfin il y a des personnes (en nombre moindre très évidemment) qui tentent de soulever ces questions dans des forums (ce mot étant pris ici au sens de la TAD), c'est-à-dire s'efforcent de les faire reconnaître dans un tel cadre institutionnel comme des *problèmes*.

À côté des forums du Net, il existe aussi des forums « médiatiques » dans lesquels des journalistes, se substituant aux usagers proprement dits, dont ils tentent de se faire les représentants, soulèvent des questions de RII. Mais outre que la chose nous semble peu fréquente, elle donne à entendre bien souvent des questions « convenables », en quelque sorte lissées, déjà un peu à distance des questions brutes que porte en lui l'utilisateur, avec leur caractère « rugueux » et peut-être inconvenant... Voici, à titre d'exemple de ce type de forum (au sens de la TAD), les questions posées par Dominique Filippone (Mesguich & Thomas, 2006b), du *Journal du Net*, aux auteurs d'un ouvrage déjà mentionné, *Net recherche* (Mesguich & Thomas, 2006a), dans la période qui suit la sortie de leur livre.

- D'où vous est venue l'idée de ce livre ?
- A-t-on une idée précise du nombre de pages existantes aujourd'hui sur Internet ?
- Face à cette prolifération des pages consultables, comment vérifier la véracité de l'information ?
- Existe-t-il des différences techniques entre les annuaires, les moteurs et les métamoteurs ?
- Que pensez-vous de la cartographie de l'information sur Internet (grise / blanche...) ?
- La recherche sur des moteurs spécialisés tels que *LawCrawler* est-elle forcément plus pertinente que sur un moteur généraliste ?
- Qu'appellez-vous dans votre livre le « Web invisible » ?
- Comment accéder à ce Web invisible ?
- Comment rechercher de l'information sur les blogs et s'assurer de la fiabilité des informations ?
- Avez-vous exploré les recherches sur les contenus liés aux fils RSS ?
- Pourra-t-on voir très prochainement des moteurs de recherche sur les flux RSS ?
- En quoi les flux RSS sont-ils un complément à la recherche via les outils classiques de recherche sur Internet ?
- Que pensez-vous du projet *Quaero* ?
- Que pensez-vous de l'initiative visant à financer le moteur de recherche Exalead ? A-t-elle une réelle chance d'aboutir à un concurrent de Google ?
- Qu'appelle-t-on Web sémantique ?
- Des adaptations des pages seront-elles nécessaires à l'adoption du Web sémantique par les outils de recherche ?
- En quoi le Web 2.0 risque de changer la structure de l'information sur Internet ?

- Les moteurs se classent-ils en différentes générations ? Si oui, quelles différences entre eux ?
- Les principaux moteurs se livrent une guerre féroce à coup de nouveautés. Quels outils sont vraiment utiles à la recherche ?
- La pertinence des résultats dépend-elle de l’interrogation ou de l’indexation de la page ?
- Existe-t-il des outils permettant des classements automatiques de contenus Web ou de résultats de recherche sur des moteurs en fonction d’ontologies générales ou spécifiques ?
- Est-il nécessaire de parler anglais pour chercher de l’information sur Internet ?
- La recherche par métamoteur est-elle plus rapide ?
- Si vous aviez 5 conseils à donner pour bien chercher, quels seraient-ils ?
- Comment peut-on faire pour être dans les 5 premiers résultats de Google ?
- Vous-même, comment organisez-vous vos recherches ?
- Où peut-on se procurer votre ouvrage ?

On voit que les questions soulevées dans ce « forum » ont une certaine généralité : pour des raisons que l’on peut comprendre, on ne s’y arrête pas sur ces questions de détail qui font en grande partie l’ordinaire des difficultés de l’usager. L’apparente faiblesse du questionnement (ou plutôt de la problématisation) touchant la RII que l’on peut parfois rencontrer dans des forums de questions de toute nature témoigne déjà du faible développement des cultures de la RII. Cela nous a conduite, on le sait, à engendrer des questions à partir d’un exposé – celui des auteurs des *Web dragons*. C’est ce que nous ferons à nouveau maintenant en nous mettant à l’écoute des questions plus ou moins implicites dans l’ouvrage de Greg Notess.

Contrairement aux questions du livre *Web dragons* (où s’explicite une culture de la RII familière à ceux qui, du côté de la *computer science* et de la RI, « font » le Web), à distance aussi des questions médiatiques proposées aux auteures de *Net recherche*, le livre de Notess peut permettre de saisir, sinon les questions que *se posent* les usagers ordinaires ou « professionnels » du Web, du moins ces questions qui, selon l’expérience qu’en ont les formateurs interrogés, *se posent* objectivement aux usagers. On trouvera dans l’annexe 5 à cette partie une telle liste de questions – engendrée, donc, à partir du texte intitulée *Teaching Web Search Skills* (Notess, 2006). Les 67 questions composant cette liste ont été mises par écrit au fur et à mesure de leur extraction au cours d’une lecture grosso modo linéaire : leur ordre n’est donc en rien significatif d’une intention de notre part. Quant à leur contenu, ces questions apparaissent très diverses, ainsi qu’il est normal dans une culture vivante. Certaines sont des questions premières, incontournables : ainsi en va-t-il sans doute des questions 19 (*Comment disposer d’un navigateur sur son ordinateur personnel ?*), 18 (*Qu’est-ce qu’un navigateur ? Est-ce cela qu’on appelle en anglais un browser ?*) ou 52 (*Windows est-il un*

moteur de recherche ? Sinon, quel est le moteur de recherche « associé » ? Est-ce Internet Explorer ?). D'autres formulent des interrogations qui, en principe, se font jour à une étape ultérieure d'une initiation éventuelle : ainsi en va-t-il de la question 33 (*Quand on formule une requête dans la barre de recherche d'un moteur, doit-on se limiter à deux ou trois mots clés ? Pourquoi ?*), ou même de la question 60 (*Quand on demande à un moteur de rechercher les documents ayant un certain mot dans leur titre, de quel titre s'agit-il ? Celui du document ? Ou d'autre chose ?*). Certaines questions peuvent procéder du désir de comprendre certains mystères du monde Internet : ainsi en est-il des questions 27 (*Pourquoi une adresse Web commence-t-elle toujours par http:// ?*), 28 (*Pourquoi dans une adresse Web trouve-t-on généralement, après http://, les « trois w » [www], et parfois non ?*), ou encore 59 (*Un site Web a-t-il une URL ?*). Quelques questions pourront sembler plus avancées, telle la question 62 (*Pour apprécier la « fiabilité » des contenus d'une page Web, on peut penser à examiner les pages Web qui ont un lien vers cette page. Mais comment connaître ces pages ?*) ; d'autres pourront procéder d'un doute devenu irritant à partir d'un certain niveau de familiarité avec le monde Internet : c'est le cas sans doute de la question 51 (*Y a-t-il une différence entre une base de données et un index ? Si oui, laquelle ?*). Certaines encore pourront venir à l'esprit auxquelles on n'avait pas de prime abord pensé. Tel est le cas certainement de la question 4 (*On dit que, lorsqu'on lui présente une requête, un moteur de recherche comme Google ne recherche pas dans tout le Web les pages qu'il va afficher ; si c'est vrai, que fait-il alors ?*), dont nous savons qu'elle ouvre sur un vaste ensemble de questions...

En dépit de leur diversité ces questions procèdent en principe de l'expérience vécue de formateurs dans leur fréquentation d'usagers en formation. Pour cela, elles sont en principe toutes « naïves », brutes, ce dont témoigne parfois des redondances qui n'en sont en fait que pour l'expert – telles celle qu'on aura peut-être notée entre la question 3 (*Qu'est-ce que le « Web invisible » et comment y effectuer des recherches ?*) et la question 11 (*Existe-t-il sur le Web des données consultables mais auxquelles les moteurs de recherche ne permettent pas d'avoir accès ? Pourquoi ?*). Le problème que nous voudrions poser ici – pour l'examiner dans le chapitre suivant – se ramène alors à ceci : en considérant qu'il s'agit là de questions auxquelles une culture solide de la RII (et plus encore la culture-mère dont nous parlions plus haut) se doit d'apporter une réponse adéquate, est-il possible, voire facile, une fois ces questions énoncées, de leur trouver des réponses dans « le monde qui nous entoure » ?

CHAPITRE 17

Vers des réponses ?

Dans le *Discours préliminaire* qu'il rédige pour l'*Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers* (1751), d'Alembert brosse un tableau magistral de connaissances de toute nature qui sont l'objet même de l'*Encyclopédie*. À ce propos, il est amené à souligner le contraste entre l'heureux état de développement des sciences et l'état de grave sous-développement de ce qu'il nomme les arts mécaniques. Cette situation, explique-t-il, a conduit les maîtres d'œuvre de l'*Encyclopédie* à enquêter directement auprès des artisans eux-mêmes – des « artistes », dans le vocabulaire du temps.

On a trop écrit sur les sciences : on n'a pas assez bien écrit sur la plupart des arts libéraux ; on n'a presque rien écrit sur les arts mécaniques ; car qu'est-ce que le peu qu'on en rencontre dans les auteurs, en comparaison de l'étendue et de la fécondité du sujet ? Entre ceux qui en ont traité, l'un n'était pas assez instruit de ce qu'il avait à dire, et a moins rempli son sujet que montré la nécessité d'un meilleur ouvrage. Un autre n'a qu'effleuré la matière, en la traitant plutôt en grammairien et en homme de lettres, qu'en artiste. Un troisième est à la vérité plus riche et plus ouvrier : mais il est en même temps si court, que les opérations des artistes et la description de leurs machines, cette matière capable de fournir seule des ouvrages considérables, n'occupe que la très petite partie du sien. [...] Tout nous déterminait donc à recourir aux ouvriers.

Que vont trouver les enquêteurs du projet encyclopédique ? La réalité que décrit d'Alembert est sans doute celle d'un autre temps ; elle n'en est pas moins révélatrice de l'état d'incurie dans lequel peut persister une « culture professionnelle » dans laquelle le minimum de communication utile semble même parfois ne pas être assuré. Voici ce qu'écrit d'Alembert à ce propos.

On s'est adressé aux plus habiles de Paris et du royaume : on s'est donné la peine d'aller dans leurs ateliers, de les interroger, d'écrire sous leur dictée, de développer leurs pensées, d'en tirer les termes propres à leurs professions, d'en dresser des tables et de les définir, de converser avec ceux de qui on avait obtenu des mémoires, et (précaution presque indispensable) de rectifier dans de longs et

fréquents entretiens avec les uns, ce que d'autres avaient imparfaitement, obscurément, et quelquefois infidèlement expliqué. Il est des artistes qui sont en même temps gens de lettres, et nous en pourrions citer ici; mais le nombre en serait fort petit. La plupart de ceux qui exercent les arts mécaniques, ne les ont embrassés que par nécessité, et n'opèrent que par instinct. À peine entre mille en trouve-t-on une douzaine en état de s'exprimer avec quelque clarté sur les instruments qu'ils emploient et sur les ouvrages qu'ils fabriquent. Nous avons vu des ouvriers qui travaillent depuis quarante années sans rien connaître à leurs machines. Il a fallu exercer avec eux la fonction dont se glorifiait Socrate, la fonction pénible et délicate de faire accoucher les esprits, *obstetrix animorum*.

La situation décrite paraît sans doute accablante. Pourtant, toutes proportions gardées, elle est peut-être plus proche qu'on ne pourrait le croire de celle qui plaque au sol une culture quand ne s'est pas produit encore le *take-off* (le décollage) libérateur, pour emprunter ici un terme à une certaine théorie du développement économique³⁷. Face à l'état actuel de la culture courante de la RII telle qu'elle s'incarne dans les usagers ordinaires, nous pourrions, à l'instar de d'Alembert et de ses amis, aller voir, non les « ouvriers », mais les usagers utilisant les ressources du Web d'une manière importante, régulière et donc, si l'on peut dire « professionnelle ». Nous pendrons deux exemples. Dans le chapitre précédent, nous avons reproduit les questions posées aux auteures de l'ouvrage *Net recherche* lors d'un entretien publié dans le *Journal du Net* ; l'avant dernière question était la suivante : « Vous-même comment organisez-vous vos recherches ? » À cette question, Armelle Thomas répond ceci³⁸.

J'ai très souvent une approche source très poussée, même si je ne connais rien au sujet au préalable. Je travaille aussi parallèlement sur les mots-clés, en me faisant sur un papier un « mini-thésaurus » que j'affine au fur et à mesure. J'use énormément de mon NetSnippets, je navigue beaucoup plus que je n'interroge. Une fois la recherche bien engagée, je change volontiers d'outils ou de technique de recherche.

Il y a là certainement un schéma de technique qu'on ne trouve pas facilement dans la « littérature » sur le sujet, et plus généralement dans ce que nous avons appelé l'offre praxéologique en matière de RII. Il en va de même de la réponse de l'autre auteure, Véronique Mesguich.

Globalement, j'utilise la même technique qu'Armelle. J'utilise de plus en plus les moteurs comme un « guide de sources ». Une fois que j'ai trouvé une ou plusieurs pistes intéressantes, j'explore à fond les sources, je navigue de lien en lien, j'essaie de trouver à la fois des sources classiques (sites

traditionnels, bases de données) et le cas échéant des sources plus originales (blogs, newsletters, experts...). Comme disent les anglais, “sit back and think”. Finalement, le meilleur outil de recherche, c’est peut-être encore le cerveau !

Nous ne nous avancerons pas beaucoup plus loin à la rencontre des « artistes » de la RII (pour parler comme d’Alembert). Mais nous donnerons un exemple supplémentaire un peu exceptionnel (pour des raisons que l’on va voir). Nous avons acceptée, alors que cette recherche était en cours, de répondre à une enquête par écrit portant sur la question même à laquelle nous venons de voir A. Thomas et V. Mesguich répondre. Cette enquête ayant été abandonnée (faute, semble-t-il, d’enquêtés « ordinaires » ou « professionnels » mais capables d’apporter une réponse un tant soit peu construite), nous nous permettons de reproduire ici – sans commentaires – la réponse que nous avons rédigée à l’époque, en précisant que nous nous y référons à des recherches « à long terme », portant sur des sujets non ponctuels, comme le fonctionnement des moteurs de recherche, le référencement et ses techniques, ou encore diverses sociétés commerciales (Google, la société allemande SAP, etc.).

Voie 1 : parcourir les moteurs de recherche généralistes comme Google, Yahoo!, MSN, ASK, Exalead sur un certain nombre de mots clés et expressions que je fais varier.

Voie 2 : vérifier les listes de liens éventuels sur des annuaires généralistes et spécialisés (en passant éventuellement par www.enfin.fr) ou des sites spécialisés dans le domaine (dont je connais souvent déjà le nom).

Voie 3 : vérifier les forums (principalement ceux de Google et Yahoo!), les blogs (en passant par Technorati), les bookmarks des membres sur Del.icio.us, Digg.com.

Voie 4 : mettre en place des alertes (Google), des abonnements à des lettres d’information et des fils RSS (que je peux chercher sur Yahoo!).

Voie 5 : vérifier dans les livres et revues spécialisés ou non (ou des noms de sites entendus à la télévision ou la radio), les sites et les mots clés proposés.

Voie 6 : si j’ai une question précise à laquelle je n’ai pas pu trouver de réponse je poste la question sur un forum et active l’alerte pour être avertie en cas de réponse.

Tout au long du travail de recherche je récolte les résultats, selon le temps dont je dispose et l’endroit où je suis (sur mon ordinateur ou non). Si j’ai du temps j’enregistre les résultats dans un fichier Word ou sur un de mes sites, sinon je les enregistre dans un blog, sur Del.icio.us ou dans les favoris de mon navigateur.

La question à laquelle nous nous attacherons dans ce qui suit est cependant autre : c'est celle de l'offre praxéologique qu'un usager ordinaire peut trouver, dans l'état actuel des choses, « dans la culture », c'est-à-dire dans la culture où il évolue ou peut être amené à évoluer. Pour fixer les idées, reprenons ici la question ci-après, qui nous avait servie d'illustration au chapitre précédent.

Bonjour,

J'ai constaté (avec Google) que l'ordre des mots clés influence sur le nombre de résultats de recherche.

Exemple : jardin versailles (1 350 000 résultats)

versailles jardin (1 360 000 résultats)

Quelqu'un pourrait m'expliquer pourquoi?

Merci d'avance

Lila1

Le problème que nous voulons poser est le suivant. Peut-on trouver à cette question Q une ou plusieurs réponses R^\diamond dans l'ensemble Σ des exposés relatifs à la RII ? Pour le dire autrement, est-il possible de « s'approvisionner » en réponses à cette question dans une « boutique de RII » dont l'accès ne soit pas déraisonnablement difficile ? Dans cette étape, nous ignorons volontairement la *valeur* (notamment sous l'angle de leur validité) des réponses R^\diamond éventuellement rencontrées. Dès lors, dans le cas de la question soulevée par Lila1, nous devrions rencontrer facilement une réponse à notre propre question, celle apportée par les intervenants sur le forum de discussion où la question a été posée. En l'espèce, à la date du 5 juin 2007, trois intervenants avaient fait connaître leur réaction³⁹. Le premier, *landeau*, se contente d'apporter ce commentaire : « Bonne question... » Le deuxième, *Dorian*, apporte une réponse déjà quelque peu construite : « Tout simplement parce que le premier mot donne plus de pertinence dans l'expression de recherche recherche. Et selon les phrases ça n'a pas du tout le même sens. » Le troisième enfin, « fosseene », propose une réponse qui pourra paraître énigmatique à Lila1 et que voici.

tout à fait d'accord avec toi

tu as un site sur la location et qui s'appelle machin tu mettre

location : toutes les locations sur machin.com

deux mots clefs avec ou sans « s » et surtout le premier mot qui ressort en premier

fosseene

Le butin est maigre, sans doute ⁴⁰. Mais c'est bien de cela qu'il sera question dans ce qui suit.

Témoigner de l'offre praxéologique ?

Pour introduire la première technique exploratoire de l'offre praxéologique de RII que nous avons mise en jeu, nous partirons de cette description due à Y. Chevallard (2008, séance 8).

... soit *S* un certain ensemble de personnes et d'institutions ; on se demande si, chez ces personnes ou dans ces institutions, donc « dans *S* », existe une réponse à une certaine question *Q*. Pour répondre à cela, on utilise des témoins qui ont fréquenté (toujours partiellement) ces personnes et ces institutions, en leur demandant s'ils ont jamais, dans ce cadre, été en contact avec une telle réponse *R* – et cela, éventuellement, en précisant les circonstances de ce « contact ». Bien entendu, on ne leur demande pas de *restituer cette réponse* (ce serait une autre question), et surtout on ne leur demande pas *s'ils posséderaient* eux-mêmes (ce qu'ils croient être) une réponse à *Q* ! Là encore, l'exercice de la subjectivité demande à être contrôlé, en éloignant le narcissisme si fréquent en pareilles circonstances, et en s'adonnant alors à un travail de *remémoration* qui participe souvent d'une véritable anamnèse.

Ainsi le lecteur des quelques pages précédentes pourra-t-il se souvenir qu'il a bien été en contact avec une réponse R^\diamond à cette question *Q* : « Lorsqu'on interroge un moteur de recherche par le truchement d'une requête faite de plusieurs mots successifs, l'ordre des mots influe-t-il sur les documents proposés par le moteur et sur l'ordre dans lequel il les propose ? » La technique d'exploration utilisée, qui n'existe encore qu'à l'état naissant, appelle des études approfondies auxquelles nous n'apporterons ici qu'une contribution modeste, liée à l'usage que nous en ferons. L'idée clé, qui est elle-même complexe, est celle de *témoin*. Cette notion recouvre ici le cas bien connu de « témoins » *non humains* dont l'exemple type est celui des marqueurs radioactifs utilisés en géologie ou en médecine par exemple ⁴¹. Il existe également des façons classiques de transformer les animaux en « témoins ». Un exemple récent et spectaculaire est celui de l'exploration de l'océan Austral (que les moyens océanographiques traditionnels ne permettaient guère, du fait de la présence d'une mer de glace le recouvrant) par le truchement d'éléphants de mer munis de capteurs, à propos desquels les auteurs de l'étude pionnière sur ce sujet concluent (Charrassin et al., 2008) : "... elephant seals fill a 'blind spot' in our sampling coverage, enabling the establishment of a truly global ocean-observing system." Lorsqu'on passe à des témoins

humains, en leur demandant d'explicitier une partie de leur expérience vécue, et donc en faisant appel à leur mémoire, les choses, on le sait, se compliquent ⁴². Nous ne rappellerons ici que l'expérience ancienne réalisée par Édouard Claparède (1873-1940) alors qu'il enseignait à la faculté de droit à Genève. Dans une étude intitulée *La fragilité du témoignage*, un psychiatre, Régis Pouget (2002), rapporte dans les termes suivants quelques-unes des conclusions du travail de Claparède.

Il avait rédigé un questionnaire qu'il proposait aux étudiants de son université, rassemblant une quinzaine de questions banales relatives aux locaux de cette même université. Les réponses, au nombre de 54, étaient édifiantes. Aucune ne donnait l'ensemble des réponses exactes. Les étudiants qui fréquentaient l'université depuis plusieurs années ne fournissaient pas de meilleures réponses que les étudiants nouvellement arrivés. Le plus frappant était la discordance entre la quantité de témoignages obtenus pour un objet donné et l'exactitude de ces témoignages. Ainsi 52 étudiants sur 54 ont apporté une réponse à la question simple sur une fenêtre qui faisait face à la loge du concierge mais 44 sur les 52 niaient l'existence de cette fenêtre qui existait pourtant. On voit que la concordance des dires de certains témoins n'est pas un critère de vérité. (pp. 1-2)

Le même auteur tire de la revue de travaux à laquelle il procède plusieurs conclusions très générales (pp. 4-5). Tout d'abord, note-t-il, « un témoignage exact en totalité est l'exception ». Ensuite, « les renseignements inexacts sont donnés par les témoins avec autant d'assurance que les renseignements exacts ». Il énonce encore le principe suivant, qui met en garde contre l'influence de la situation *actuelle* sur le souvenir de la situation remémorée : « Les témoins ont tendance à percevoir les faits et à reconstituer leurs souvenirs en fonction de ce qui leur paraît vraisemblable plus qu'en fonction de ce qui s'est réellement passé. » Après avoir indiqué encore que « la concordance des témoignages entre plusieurs témoins n'est pas toujours un critère valable », il conclut par cette prudente observation : « en matière de reconnaissance, l'erreur est plus fréquente que l'exactitude. » En dépit de ces difficultés bien connues, nous avons sollicité, à propos de l'offre praxéologique en matière de RII dans la culture où ils évoluent, le témoignage de publics étudiants divers, dont nous allons maintenant examiner l'apport.

Une enquête auprès de « témoins »

Nous avons demandé à plusieurs groupes d'étudiants en science de l'éducation de l'université de Provence de témoigner de leur rencontre éventuelle avec des réponses à une série de

questions relevant du domaine de la RII. Voici d'abord, sans commentaire pour le moment, la série de 20 questions à propos desquelles leur souvenir était sollicité : ces questions ont été choisies parmi les 67 que, dans le chapitre précédent, nous avons tirées du livre de Greg Notess (et qui figurent, on l'a dit, dans l'annexe 5 à cette partie de notre travail).

1. Quand on utilise un moteur de recherche, y a-t-il autre chose à savoir au-delà du fait de taper des mots clés et de lancer la recherche ?
2. Quelle différence y a-t-il entre un « site Web » et une « page Web » ?
3. Qu'est-ce qu'un portail ? Est-ce utile pour la recherche d'information ?
4. Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ?
5. Le moteur de recherche Google présente deux boutons, « Recherche Google » et « J'ai de la chance ». Si l'on clique sur le bouton « J'ai de la chance », on arrive directement sur une page Web. Comment cette page est-elle choisie ?
6. Comment peut-on copier tout ou partie du contenu d'une page Web apparaissant à l'écran afin de coller une telle image dans un fichier .doc ou .ppt ?
7. Qu'est-ce exactement que le Web ? Est-ce différent d'Internet ? En quoi ?
8. Comment ouvrir un lien (présent sur une page Web) dans une nouvelle fenêtre ?
9. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?
10. Existe-t-il des moteurs de recherche qui soient spécialisés dans la recherche d'images ?
11. Qu'est-ce que « le cache de Google » et à quoi cela peut-il être utile ?
12. Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ?
13. Il arrive que, lorsqu'on tape l'URL d'un site Web dans la barre d'adresse d'un navigateur, on arrive sur un autre site. Pourquoi cela ? Comment peut-on l'éviter ?
14. Qu'est-ce qu'un annuaire ? Quand est-ce utile ?
15. Avec un moteur de recherche, peut-on chercher les pages Web qui contiennent tel mot mais pas tel autre ?

16. Lorsqu'un moteur de recherche propose un résultat, que peut-on tirer de ce qu'il indique sur la page de résultats ? Que signifient les différentes indications qu'il affiche ?
17. Dans un travail universitaire (mémoire, présentation, etc.), comment citer une page Web ?
18. On dit qu'un moteur de recherche « indexe » les pages du Web. En quoi cela consiste-t-il et à quoi cela sert-il ?
19. Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ?
20. Quand, par erreur, on tape une requête dans la barre d'adresse d'un navigateur, on obtient souvent des résultats comme s'il s'agissait d'un moteur de recherche (et non d'un navigateur) ? Pourquoi ?

À propos de cet ensemble de questions, nous ne pouvons guère à ce stade de l'analyse exprimer qu'un *sentiment*. Il nous semble que ces 20 questions, dans leur diversité, sont susceptibles de « parler » à une majorité des personnes que nous connaissons pour en être proche professionnellement, parce que ces questions parlent d'un univers qui ne leur est pas inconnu ; mais elles en parlent sur un « ton » qui, en revanche, *ne leur est nullement familier*. S'interroger ou être interrogé, explicitement, directement, à propos du Web ne constitue pas, aujourd'hui encore, en effet, une situation banale ni même normale, c'est-à-dire qui suive une quelque norme institutionnelle connue. À l'occasion de la présentation de cette liste de questions dans le cadre d'un séminaire de didactique dont les participants n'étaient à l'évidence ni incultes dans le domaine, ni supérieurement informés (à de rares exceptions près), nous avons ainsi pu observer comme un *silence* dans le déroulement de la séance à l'instant où les questions sont apparues sur l'écran, silence dû sans doute en grande partie à une lecture attentive, certes, mais aussi, en même temps, à la réception d'un ensemble d'énoncés d'un type inattendu, et presque étrange...

Venons-en donc au premier groupe que nous avons sollicité : il était en l'espèce composé de 11 étudiants de deuxième année du master professionnel de sciences de l'éducation, option « Ingénierie de la formation par alternance ». Sur le questionnaire transmis à chacun d'eux, intitulé *À propos de la recherche d'information sur Internet*, chacune des questions apparaissait selon le format que l'on trouvera illustré ci-après à propos de la question 2. Le deuxième score que chaque enquêté devait attribuer à chacune des questions – relatif au « désir actuel de disposer d'une réponse » – n'était présenté et commenté qu'oralement. (Notons en outre que, pour des raisons matérielles, lors de cette passation,

seules les 19 premières questions – sur les 20 reproduites ci-dessus – ont été proposées aux étudiants.)

QUESTION 2. *Quelle différence y a-t-il entre un « site Web » et une « page Web » ?*

– Importance du contact avec une réponse :	0	1	2	3	4
– Votre désir actuel de disposer d’une réponse :	0	1	2	3	4

La consigne écrite en tête du questionnaire, commentée oralement lors de la passation, était la suivante.

À propos de la recherche d’information sur Internet

En choisissant dans chaque cas l’un des scores de 0 (contact nul) à 4 (contact approfondi), précisez l’importance du contact que vous avez pu avoir (dans le cadre d’un cours, d’un stage de formation, d’une lecture personnelle, d’un échange informel, etc.) avec une réponse éventuelle à chacune des n questions suivantes.

N. B. Le fait d’avoir éventuellement oublié tout ou partie d’une réponse avec laquelle vous avez été en contact autrefois ne doit pas intervenir dans le choix du score. Par ailleurs, l’ordre de présentation des questions, ci-après, est volontairement arbitraire.

Nous nous arrêterons ici sur le premier score, relatif à « l’importance du contact » avec une réponse éventuelle. L’examen « de près » des scores assignés par les enquêtés semble révéler de multiples singularités. Observons en premier lieu que le nombre de scores nuls attribués est de 22, ce qui représente environ 10,5 % des scores assignés (le nombre de scores assignés est égal à $19 \times 11 = 209$). Sept des 11 étudiants ne choisissent jamais le score nul : les 22 zéros attribués se partagent entre 4 personnes qui ont attribué respectivement 9, 7, 2 et 4 zéros, et dont le total des scores représente respectivement 17 %, 34 %, 42 % et 41 % du maximum théorique du total des scores assignés (qui est ici égal à $19 \times 4 = 76$). Le total des scores attribués par les 7 autres étudiants est respectivement, en pourcentage du maximum possible des scores, égal à 67 %, 45 %, 83 %, 53 %, 57 %, 68 %, 51 % : on voit donc que les 7 qui n’ont jamais attribué de score nul ont un total des scores supérieur ou égal à 51 % du total maximal, alors que les 4 autres étudiants ne dépassent pas 42 % ; ceux-ci déclarent ainsi

n'avoir eu que des contacts limités ou nuls avec des réponses, tandis que ceux-là auraient eu des contacts non nuls et souvent plus approfondis.

D'une façon plus générale, la distribution des scores sur l'ensemble des 209 scores assignés est la suivante.

Score	0	1	2	3	4
Effectif	22	53	55	55	24

On voit que cette distribution est quasi symétrique. Sur les 24 scores 4 attribués – il y en a un peu plus que de scores nuls, donc –, 21 proviennent du « groupe des sept », avec les effectifs suivants : 0 ; 0 ; 10 ; 5 ; 1 ; 5 ; 0. Les trois scores 4 qui échappent à ce groupe se répartissent comme suit au sein du « groupe des quatre » : 0 ; 1 ; 2 ; 0. On aura observé que l'un des participants, que Greg Notess appellerait peut-être – mais peut-être à tort – un “know-it-all” (Notess, 2006, p. 14), a attribué *dix fois* le score 4 : il déclare ainsi avoir eu un contact « approfondi » avec une ou des réponses aux questions 2 (« Quelle différence y a-t-il entre un “site Web” et une “page Web” ? »), 3 (« Qu'est-ce qu'un portail ? Est-ce utile pour la recherche d'information ? »), 4 (« Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ? »), 6 (« Comment peut-on copier tout ou partie du contenu d'une page Web apparaissant à l'écran afin de coller une telle image dans un fichier .doc ou .ppt ? »), 7 (« Qu'est-ce exactement que le Web ? Est-ce différent d'Internet ? En quoi ? »), 8 (« Comment ouvrir un lien (présent sur une page Web) dans une nouvelle fenêtre ? »), 12 (« Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ? »), 14 (« Qu'est-ce qu'un annuaire ? Quand est-ce utile ? »), 15 (« Avec un moteur de recherche, peut-on chercher les pages Web qui contiennent tel mot mais pas tel autre ? ») et 18 (« On dit qu'un moteur de recherche “indexe” les pages du Web. En quoi cela consiste-t-il et à quoi cela sert-il ? »). Tout cela, au fond, n'a rien d'impossible, même si nous sommes dubitative devant certains des six scores 3 attribués par la même personne, par exemple à la question 19 (« Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ? »).

Ces accidents du relief statistique s'estompent si l'on regarde d'une distance suffisante les données recueillies. Le total des scores assignés à chacune des 19 questions est respectivement le suivant (le score total maximal est égal à $11 \times 4 = 44$) : 26, 32, 29, 30, 12,

25, 26, 27, 13, 27, 10, 23, 16, 19, 24, 25, 29, 16, 15. Si l'on classe alors les questions par score croissant, on obtient le rangement suivant (en désignant les questions par leur numéro) : $11 < 5 < 9 < 19 < 13 < 18 < 14 < 12 < 15 < 6 < 16 < 1 < 7 < 8 < 10 < 3 < 17 < 4 < 2$. Les questions auxquelles des réponses ont été le moins nettement rencontrées sont alors les suivantes (on se limite ici aux questions dont le total des scores ne dépasse pas 50 % du total maximal).

11. Qu'est-ce que « le cache de Google » et à quoi cela peut-il être utile ?

5. Le moteur de recherche Google présente deux boutons, « Recherche Google » et « J'ai de la chance ». Si l'on clique sur le bouton « J'ai de la chance », on arrive directement sur une page Web. Comment cette page est-elle choisie ?

9. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?

19. Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ?

13. Il arrive que, lorsqu'on tape l'URL d'un site Web dans la barre d'adresse d'un navigateur, on arrive sur un autre site. Pourquoi cela ? Comment peut-on l'éviter ?

18. On dit qu'un moteur de recherche « indexe » les pages du Web. En quoi cela consiste-t-il et à quoi cela sert-il ?

14. Qu'est-ce qu'un annuaire ? Quand est-ce utile ?

Ce « choix collectif » n'est guère surprenant : pour nous, seule la présence de la question 5 et sa position dans le classement a été une surprise⁴³. Mais d'autres épisodes survenus à la même époque nous ont confirmé qu'il ne s'agissait pas là d'une anomalie : le bouton « J'ai de la chance » semble en effet très peu connu dans l'univers où nous-même évoluons, en partie pour cette raison que nombre d'utilisateurs accèdent à l'interface avec le moteur de recherche Google proposée par le navigateur Firefox, interface où ce bouton... n'apparaît pas.

Le deuxième groupe d'étudiants sollicité était également composé de 11 personnes, inscrites en licence de sciences de l'éducation et en formation à Forcalquier. Le questionnaire comportait cette fois les 20 questions annoncées et, comme il en ira dans les passations ultérieures, comportait une consigne légèrement modifiée, que nous reproduisons ci-après.

Pour chacune des 20 questions suivantes, on ne vous demande ni de formuler une réponse à cette question, ni de dire si vous croyez connaître la (ou une) réponse. Mais on vous demande de préciser

– en choisissant l’un des scores de 0 (contact nul) à 4 (contact approfondi) – *l’importance du « contact » que vous avez pu avoir avec une réponse éventuelle* à cette question (dans le cadre d’un cours, d’un stage de formation, d’une lecture personnelle, d’un échange informel, etc.), et cela même si vous avez tout oublié de cette réponse. On vous demande en outre, dans tous les cas, d’estimer (toujours sur une échelle de 0 à 4) votre désir actuel de disposer d’une réponse solide à la question considérée. Notez que l’ordre de présentation des questions, ci-après, est volontairement arbitraire.

On a noté déjà – dans l’annexe 1, à la fin de notre *Introduction générale* – que le groupe de Forcalquier présentait quelques traits atypiques. Le tableau des effectifs des différents scores (de 0 à 4), ci-après, illustre bien cette remarque : comme on le verra, la distribution est fortement dissymétrique, avec un nombre de zéros quatre fois supérieur à celui attribué par le groupe précédent. L’asymétrie évidente se complique du fait que, alors que les effectifs décroissent lorsqu’on passe de 0 à 3, ils croissent lorsqu’on passe de 3 à 4, atteignant pour ce dernier score une valeur supérieure à celle observée chez le premier groupe. (On notera que la somme des effectifs, soit $88 + 52 + 33 + 16 + 31 = 220$, est bien égale au nombre des scores assignés, à savoir $11 \times 20 = 220$.)

Score	0	1	2	3	4
Effectif	88	52	33	16	31

Les membres du groupe ont eu manifestement, vis-à-vis de la RII, des passés parfois fort différents. À une extrémité, la personne dont les scores ont le plus fort total, représentant 74 % du total maximal, a exercé comme documentaliste et a eu dans ce rôle à former à la RII des élèves de collège : elle attribue 8 fois le score 4, jamais le score nul, et n’assigne un prudent score 1 qu’à la question 19 (« Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ? »), ce qui se comprend, ainsi qu’à la question 16 (« Lorsqu’un moteur de recherche propose un résultat, que peut-on tirer de ce qu’il indique sur la page de résultats ? Que signifient les différentes indications qu’il affiche ? »), ce qui est plus étonnant mais constitue peut-être une indication sur une faiblesse actuelle de la culture scolaire de la RII. Le deuxième plus fort total des scores s’établit à 64 % du total maximal : il est le fait d’une étudiante dont le profil se rapproche du profil précédemment évoqué et qui attribue 9 fois le score 4 ; toutefois, elle fait usage par trois fois du score nul, utilisé à propos des questions 19, 20 (« Quand, par erreur, on tape une requête dans la barre d’adresse d’un navigateur, on obtient souvent des résultats comme s’il s’agissait d’un moteur de recherche (et

non d'un navigateur) ? Pourquoi ? ») et 9 (Comment faire une recherche portant spécifiquement sur les groupes de discussion [newsgroups] ?), ce qui ne surprend guère. À l'extrême opposé, on trouve une étudiante qui, 17 fois, a attribué un 0, deux fois un 1 et, étonnamment, une fois un 3, attribué à la question 12 : « Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ? » Les différentes questions examinées ont, en règle générale, obtenu un score total bien évidemment inférieur au score correspondant attribué par le premier groupe examiné, ainsi que le montre le tableau ci-après ⁴⁴.

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Premier groupe	59	73	66	68	27	57	59	61	30	61	23	52	36	43	55	57	66	36	34
Groupe de Forcalquier	45	39	34	41	18	50	39	52	9	34	20	59	18	27	18	27	68	45	2
Différence des scores	14	34	32	27	9	7	20	9	21	27	3	-7	18	16	37	30	-2	-9	32

Ce tableau, on l'aura remarqué, montre aussi des exceptions. Pour la question 17, l'excès de 2 points en pourcentage du groupe de Forcalquier sur le premier groupe s'explique aisément par le fait que le groupe avait eu, au cours des semaines précédentes, une petite formation sur un certain nombre de questions touchant le monde Internet, en sorte par exemple qu'aucun des 11 enquêtés n'a attribué à cette question de score nul : elle reçoit 4 fois le score 4, 2 fois le score 3, 3 fois le score 2 et 2 fois le score 1 (l'étudiante aux 17 zéros a ainsi attribué à cette question le plus petit score non nul). De même, on peut sans doute voir l'effet de la formation reçue dans le cas des deux autres questions où le groupe de Forcalquier déclare un contact plus approfondi que le premier groupe, à savoir la question 12 (la formation suivie par ce groupe a dû faire une place à la « recherche avancée », et c'est d'ailleurs à cette question 12, on l'a noté, que l'étudiante aux 17 zéros attribue un 3) et la question 18 (« On dit qu'un moteur de recherche "indexe" les pages du Web. En quoi cela consiste-t-il et à quoi cela sert-il ? »), à laquelle il paraît raisonnable d'appliquer, *mutatis mutandis*, la même remarque.

En dépit des singularités de la vie du groupe et de l'histoire de ses membres, un regard « de loin » fait-il apparaître des difficultés voisines de celles manifestées par le premier groupe dans l'accès aux réponses aux questions proposées ? Voici le classement des questions

par ordre croissant du total des scores qui leur ont été assignés : 19 < 9 < 20 < 5 < 13 < 15 < 11 < 14 < 16 < 3 < 10 < 2 < 7 < 4 < 1 < 18 < 6 < 8 < 12 < 17. S'agissant du premier groupe, les huit questions avec lesquelles le contact déclaré par le groupe est le plus faible sont (par ordre croissant d'importance du contact) les questions 11, 5, 9, 19, 13, 18, 14, 12. Si l'on ignore l'ordre mis sur les questions, les huit premières questions sont donc, dans le cas du premier groupe, les questions 5, 9, 11, 12, 13, 14, 18, 19, et, dans le cas du groupe de Forcalquier, les questions 5, 9, 11, 13, 14, 19, 20. Si l'on tient compte de ce que la question 20 ne pouvait pas figurer dans le classement réalisé de fait par le premier groupe, et que les questions 12 et 18 étaient l'objet d'un contact récent pour le groupe de Forcalquier, l'accord paraît remarquable : au-delà de la dissemblance des témoignages exprimés par les scores attribués par les membres de l'un et l'autre groupe, *un même fait émerge*, celui d'une rencontre incertaine avec des réponses à un certain nombre des questions proposées, en l'espèce les questions 5, 9, 11, [12], 13, 14, [18], 19, [20], dont l'accessibilité culturelle, donc, ne va de soi.

Offre et besoins praxéologiques : une rencontre délicate

Pour vérifier le résultat ainsi mis en évidence, nous avons étendu l'enquête à d'autres groupes de « témoins ». Outre deux groupes supplémentaires, le premier constitué de 36 professeurs stagiaires de l'IUFM d'Aix-Marseille, le second de 10 étudiants de deuxième année du master professionnel de sciences de l'éducation ayant choisi soit l'option « Responsable de formation », soit l'option « Ingénierie de la formation multimédia », nous avons recueilli des « témoignages » *en ligne*, d'une part auprès de 15 élèves de l'Institut de formation des cadres de santé (IFCS) d'Aix-en-Provence, d'autre part auprès de 59 personnes ayant rempli le questionnaire en ligne à la suite d'une annonce relayée par divers collègues⁴⁵. En incluant le groupe de Forcalquier, on arrive ainsi à un ensemble de 131 réponses complètes au questionnaire. Le classement des 20 questions par le total de leurs scores est alors le suivant : 19 < 9 < 20 < 11 < 13 < 5 < 18 < 12 < 7 < 3 < 16 < 15 < 14 < 2 < 10 < 1 < 4 < 17 < 8 < 6. *L'accord est remarquable* avec le résultat précédemment obtenu : alors que les huit questions pour lesquelles la rencontre avec des réponses apparaissait la plus fragile étaient les questions 5, 9, 11, [12], 13, 14, [18], 19, [20], on voit apparaître ici, dans ce « rôle », les questions 5, 9, 11, 12, 13, 18, 19, 20. Seule la question 14, qui figurait *in extremis* parmi les questions à réponses peu rencontrées, remonte véritablement dans le classement, avec un score total qui, toutefois, ne représente que 45 % du maximum possible (alors par exemple que la question 6, la mieux « servie » dans l'offre praxéologique explorée, voit le total de ses scores s'établir à

65 % du maximum) : peut-être y a-t-il là une conséquence du fait que les 59 enquêtés en ligne évoqués plus haut, qui sont souvent des enseignants, des formateurs ou des enseignants chercheurs, ont une moyenne d'âge sensiblement plus élevée (autour de 42 ans) que les témoins « étudiants ».

On voit ainsi que plusieurs des questions de RII choisies pour cette enquête ne sont guère prises en charge par l'offre praxéologique effectivement accessible aux enquêtés dans les conditions usuelles de leur activité. Au demeurant, la technique que nous avons mise en œuvre nous semble être de nature à *surestimer* cette offre. Il est vraisemblable en effet que les « témoins » sollicités, invités en principe à consulter leurs souvenirs, et cela uniquement, auront pour certains eu tendance à remplacer, plus ou moins subrepticement, l'idée de rencontre d'une réponse par celle de *connaissance personnelle*, présente ou passée, d'une réponse. En soi, la chose ne devrait pas avoir de conséquence inflationniste sur les scores assignés : si l'on connaît une réponse, c'est qu'on l'a rencontrée sous une forme ou une autre (même si l'on a oublié les circonstances et modalités précises de cette rencontre), à moins de l'avoir créée soi-même *ex nihilo*, ce qui est sans doute rare ; en sorte que la substitution de « Je connais ou j'ai connu » à « J'ai rencontré » pourrait plutôt conduire à *sous-estimer* l'offre praxéologique, parce que les enquêtés oublieraient les « rencontres » avec des réponses qu'ils considèrent n'avoir pas « vraiment » connues et ne pas connaître. Il semble pourtant que cette substitution laisse la voie libre, chez au moins certains des enquêtés, à une surestimation, non de leur contact avec une réponse, mais de leur « maîtrise » d'une réponse, et cela moins parfois par « vanité » que par souci de l'institutionnellement correct – quand il semble à l'enquêté qu'il ne serait pas correct d'ignorer entièrement la réponse à telle ou telle question, de sorte que le score qu'il attribuera, qui aurait pu être 0 ou 1 par exemple, glissera peut-être vers 1 ou 2 (ou plus).

On a vu que les enquêtés étaient invités à assigner des scores appréciant *deux* ordres de réalité : tout d'abord, le *contact* qu'ils avaient pu avoir avec une réponse ; ensuite, le *désir* qu'ils pouvaient éprouver de disposer d'une réponse. Signalons d'abord que les deux variables définies ainsi sur l'ensemble des 20 questions, soit le *score de contact* et le *score de désir*, apparaissent presque toujours faiblement ou très faiblement corrélées. Sur le premier groupe étudié (pour lequel nous ne disposons des scores que pour les 19 premières questions), la série des scores de contact assignés par le groupe et celle des scores de désir ont ainsi un coefficient de corrélation égal à 0,09. Pour le groupe de Forcalquier, ce coefficient vaut 0,12. Quant aux deux séries de scores de contact et de désir correspondant au groupe des 131 enquêtés, leur corrélation est très légèrement négative, égale à -0,21.

Cela noté, comment le groupe des 131 enquêtés classe-t-il selon l'intensité du désir les 20 questions proposées ? Voici la réponse : $8 < 2 < 19 < 9 < 10 < 6 < 20 < 14 < 4 < 7 < 3 < 1 < 18 < 5 < 15 < 16 < 12 < 11 < 13 < 17$. Les réponses les plus « désirées » sont ainsi celles relatives aux questions 18, 5, 15, 16, 12, 11, 13, 17, soit, en oubliant l'ordre, 5, 11, 12, 13, 15, 16, 17, 18. Ainsi donc, sur le petit échantillon de 20 questions que nous avons choisi, parmi les huit questions déclarées être celles avec lesquelles les enquêtés ont eu les contacts les plus limités (5, 9, 11, 12, 13, 18, 19, 20), cinq (5, 11, 12, 13, 18) se classent parmi les questions pour lesquelles des réponses sont le plus vivement désirées (5, 11, 12, 13, 15, 16, 17, 18). On observe ici un phénomène qui nous semble au cœur du problème de la culture de la RII aujourd'hui : l'offre praxéologique est d'une retenue – sinon d'une indigence – qui ne conduit pas aisément à la satisfaction des besoins praxéologiques exprimés par les usagers, pour peu qu'on les interroge.

Une offre praxéologique accessible ?

Une culture porte en elle une *offre* praxéologique plus ou moins généreuse. Mais elle aide aussi – plus ou moins – à y accéder en allégeant le coût d'accès à cette offre, d'une part parce qu'elle rend légitime, normal, voire banal le projet d'y accéder, diminuant ainsi le coût « psychologique » de s'engager dans un tel projet, mais aussi parce qu'elle « outille » praxéologiquement ce projet en rendant techniquement familiers les gestes à accomplir pour l'exécuter. À l'inverse, l'absence d'une culture idoine peut provoquer une « stase » (Chevallard, 2008, séance 5) de la diffusion praxéologique dans les institutions qui en pâtissent : une personne qui se trouve assujettie à une telle culture pourra par exemple se poser une question, motif potentiel d'une enquête et point d'un départ d'un parcours d'étude et de recherche, puis abandonner là sa quête en croyant se souvenir d'une certaine réponse jamais vérifiée, jamais vraiment « travaillée ».

Dans la culture courante actuelle de la RII, il est sans doute encore peu commun qu'une personne passe des heures à tenter de rencontrer des réponses R^\diamond à des questions Q *touchant à la RII* : l'obstacle est ici à la fois dans *l'outil* – la RII – et dans *l'objet* auquel on doit l'appliquer – des questions *de RII*. Pour cette raison notamment, nous avons provoqué un tel épisode en demandant à des personnes de bien vouloir participer, en équipe de trois, à une séance de trois heures environ de recherche de réponses R^\diamond : cette sollicitation a été adressée à des formateurs de mathématiques de l'IUFM d'Aix-Marseille. Deux équipes de trois formateurs ont ainsi été constituées ; les deux séances de travail correspondantes ont eu lieu dans un local de l'IUFM où chaque membre de l'équipe pouvait disposer d'un ordinateur

connecté à l'Internet. Nous parlerons ci-après de l'équipe 1 et de l'équipe 2 en référence à la chronologie des séances. Ce choix de « témoins », des formateurs aguerris, devait d'abord permettre de provoquer une séance de travail dense et bien remplie, ce qui s'est en effet produit : comme le montre l'enregistrement vidéo des séances (les deux séances ont été intégralement filmées), les deux équipes ont travaillé tout au long du temps alloué sans lâcher prise, selon un habitus professionnel à l'évidence profondément incorporé ! Par ailleurs, nous savions que ces personnes avaient une familiarité réelle avec la RII, mais cela, sauf exception, sans avoir jamais cultivé de façon systématique et approfondie ce domaine mis par eux au service de leurs besoins professionnels.

Le travail à accomplir a été présenté à chaque équipe à l'aide d'un document distribué à chacun de ses membres, doublé d'une présentation orale et d'un échange avec les participants. Le document comportait quatre rubriques successives : 1. L'objet de la séance / 2. Le corpus documentaire / 3. Les questions / 4. Un relevé de conclusions. Voici d'abord le texte présentant *l'objet de la séance*.

Il s'agit pour l'équipe réunie

- de mettre en évidence la présence éventuelle, dans un certain corpus de documents (précisé ci-après), de réponses à un ensemble de questions indiquées dans ce qui suit,
- pour chaque question examinée, d'estimer globalement, à l'arrêt de la recherche, la facilité d'accès à une réponse, selon l'échelle suivante :

0 : réponse introuvable ;

1 : réponse très difficile à trouver ;

2 : réponse difficile à trouver ;

3 : réponse assez facile à trouver ;

4 : réponse très facile à trouver.

Attention ! Il ne s'agit pas de juger de la validité ou de la valeur de ce qui se présente, explicitement ou implicitement, comme une réponse. Il ne s'agit pas non plus d'estimer la complexité de la réponse trouvée ou la difficulté qu'il y aurait à la « comprendre ». Il s'agit seulement de jauger la difficulté d'y accéder dans le corpus de documents indiqué.

Quel était alors le *corpus documentaire* dans lequel devaient être localisées par l'équipe les réponses supposées pertinentes ? Le document de présentation le définissait ainsi.

Les documents dans lesquels il est demandé de tenter de mettre en évidence l'existence éventuelle d'une réponse (ou de fragments de réponse) sont les suivants.

a) Le chapitre II du petit ouvrage d'Isabelle Bréda et Carole Hourt intitulé *Les bons usages de l'Internet* (2007). Le texte de ce chapitre est disponible sous la forme d'un fichier pdf (afin de faciliter la recherche de chaînes de caractères).

b) Trois ouvrages (en français) sur Internet contenant chacun des chapitres ou parties portant sur la recherche d'information sur Internet :

– *Internet pour les nuls*, de John R. Levine, Carol Baroudi et Margaret Levine Young (First Interactive, 2006) ;

– *Internet. Le guide complet*, de Laurent Bécalséri (Micro application, 2007) ;

– *Le livre de l'Internet*, de Henri Lilen (First Interactive, 2008).

c) Les sites (en français) dont les adresses suivent :

• <http://membres.lycos.fr/patderam/docu.htm>

• <http://documentaliste.ac-rouen.fr/spip/spip.php?article75>

• <http://www.ext.upmc.fr/urfist/cerise/p7.htm>

• <http://www.commentcamarche.net/utile/research.php3>

• http://www.precisement.org/internet_docu/methode_rech.htm

• <http://www.ebsi.umontreal.ca/jetrouve/internet/index.htm>

• http://www.sciencepresse.qc.ca/recherche/intro_recherche.html

• <http://www.rechercheinternet.ca/>

• <http://clioweb.free.fr/peda/chercher.htm>

• <http://www.agropolis.fr/ist/guide/impression/gfri-chercher-internet.pdf>

• <http://www.dsi-info.ca/>

• <http://www.bibl.ulaval.ca/vitrine/giri/>

• <http://www.esen.education.fr/fr/ressources-par-type/outils-pour-agir/guide-de-recherche-d-information/>

• <http://www.educnet.education.fr/dossier/rechercher/default.htm>

d) Plus généralement, l'ensemble des documents accessibles sur le Web.

Plusieurs points doivent être soulignés. Tout d'abord, les quatre ouvrages mis à la disposition de l'équipe totalisent ensemble quelque 1800 pages : si une lecture questionnante était possible, une lecture inventoriante exhaustive ne l'était guère ! Ensuite et surtout, en dépit de la liste particulière de sites Web proposée, la recherche pouvait se développer librement sur

l'ensemble du Web, sans qu'il y eût de priorité imposée dans l'usage des outils documentaires proposés (ou suggérés). L'idée était de créer une situation ressemblant raisonnablement à une situation quasi usuelle de recherche documentaire – même si, sans doute, il est rare qu'un formateur non spécialiste se donne les moyens de disposer même de la minuscule « bibliothèque » que nous avons sélectionnée. Voici alors les questions Q auxquelles les équipes devaient rechercher des réponses R^\diamond : il s'agit de dix questions choisies parmi les vingt utilisées ci-dessus pour explorer l'offre praxéologique de RII.

Les questions pour lesquelles il est demandé d'examiner la présence éventuelle d'une réponse dans le corpus de documents indiqué (ainsi que d'estimer la difficulté d'accès à une telle réponse) sont les suivantes.

Questions

1. Quelle différence y a-t-il entre un « site Web » et une « page Web » ?
2. Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ?
3. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?
4. Qu'est-ce que « le cache de Google » et à quoi cela peut-il être utile ?
5. Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ?
6. Il arrive que, lorsqu'on tape l'URL d'un site Web dans la barre d'adresse d'un navigateur, on arrive sur un autre site. Pourquoi cela ? Comment peut-on l'éviter ?
7. Avec un moteur de recherche, peut-on chercher les pages Web qui contiennent tel mot mais pas tel autre ?
8. Dans un travail universitaire (mémoire, présentation, etc.), comment citer une page Web ?
9. Comment faire pour utiliser Google News sans être chaque fois renvoyé vers Google Actualités ?
10. Quand, par erreur, on tape une requête dans la barre d'adresse d'un navigateur, on obtient souvent des résultats comme s'il s'agissait d'un moteur de recherche (et non d'un navigateur) ? Pourquoi ?

Le travail demandé consistait à *repérer* des réponses R^\diamond , non à en faire l'analyse, et encore moins à les « évaluer ». Bien entendu, le repérage suppose une analyse minimale, qui permet de s'assurer du fait qu'il s'agit bien d'une réponse possible à la question examinée ; mais rien de plus n'était demandé. La dernière rubrique du document guide, *Un relevé de conclusions*, précisait ce qui était concrètement attendu de l'équipe.

Pour chaque question examinée, et après accord entre les membres de l'équipe, il est demandé de consigner par écrit brièvement (sous forme électronique de préférence) :

- les références des passages (imprimés ou en ligne) où un élément de réponse apparaît ;
- le « score » (sur l'échelle de 0 à 4 précisée plus haut) attribué par l'équipe à la facilité d'accéder à l'ensemble des éléments de réponse qui aura été identifié. (Ce score pourra faire éventuellement l'objet d'un bref commentaire.)

Les séances se sont déroulées jusqu'à leur terme selon la consigne annoncée : chaque équipe a élaboré un relevé de conclusions dans lequel on trouve notamment un score de difficulté pour chacune des dix questions. Voici le tableau de ces scores.

Question (sur les 10)	Question (sur les 20)	Équipe 1	Équipe 2
1	2	4/4	3-4/4
2	4	2/4	2/4
3	9	4/4	4/4
4	11	4/4	4/4
5	12	3/4	4/4
6	13	1/4	1/4
7	15	3/4	4/4
8	17	4/4	4/4
9	19	2/4	0/4
10	20	1/4	2/4

On peut simplifier ce tableau en ne retenant que deux catégories de questions : « difficiles » pour les scores 0, 1, 2, affectés ci-après d'un signe ♠, et « faciles » pour les scores 3 et 4, marqués d'un signe ♥ (voir ci-après). On voit immédiatement le *bon accord* entre les

jugements portés par chaque équipe, qui ont pourtant travaillé de manière indépendante l'une de l'autre.

Question (sur les 10)	Question (sur les 20)	1 ^{re} équipe	2 ^e équipe
1	2	♥	♥
2	4	♠	♠
3	9	♥	♥
4	11	♥	♥
5	12	♥	♥
6	13	♠	♠
7	15	♥	♥
8	17	♥	♥
9	19	♠	♠
10	20	♠	♠

Les questions « difficiles », au nombre de 4 (sur les 10 proposées), sont les questions 4, 13, 19 et 20. La présence des questions 13, 19 et 20 corrobore le résultat de l'étude précédente, relative au contact avec des réponses aux 20 questions, qui conduisait au classement suivant : $19 < 9 < 20 < 11 < 13 < 5 < 18 < 12 < 7 < 3 < 16 < 15 < 14 < 2 < 10 < 1 < 4 < 17 < 8 < 6$. En revanche, il n'en va pas de même pour la question 4 (« Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ? »), réputée « connue » par le groupe des 131 (avec un total des scores représentant 54 % du total maximal possible). On voit aussi qu'une « anomalie » de sens inverse affecte les questions 9 et 11 : déclarées par le groupe des 131 témoins comme ayant des réponses plutôt insaisissables (dans les deux cas, le total des scores de contact représente seulement 21 % du maximum possible), elles apparaissent, dans le jugement des deux équipes, comme « faciles », et même « très faciles », puisque leur score de difficulté est dans les deux cas de 4/4. En revanche, une réponse à la question 4 est simplement jugée « difficile à trouver ». Que peut-on conclure de ces accords et de ces désaccords ?

Si l'on rapproche les jugements portés par les deux équipes d'investigation des classements induits par les scores assignés par les 131 enquêtés, on obtient le tableau suivant montrant les quatre cas de figure possibles ⁴⁶.

	Recherche facile	Recherche difficile
Contact fort	2, 15, 17	4
Contact faible	9, 11, 12	13, 19, 20

Nous nous arrêterons d’abord sur les questions à propos desquelles une « anomalie » se produit, en commençant par les questions 9, 11, 12, pour lesquelles le contact avec des réponses déclarées par le groupe des 131 est *faible* et qui sont pourtant jugées par les deux équipes d’investigation comme faciles à trouver (les deux équipes leur accordent le score de 4/4, sauf pour la question 12 où l’équipe 1 donne le score de 3/4). Le contenu de ces questions peut expliquer la situation constatée ; rappelons-le.

9. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?

11. Qu’est-ce que « le cache de Google » et à quoi cela peut-il être utile ?

12. Peut-on demander à un moteur de recherche de n’afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ?

La faiblesse du contact déclaré semble indiquer que ces questions sont à la périphérie de la culture courante en matière de RII – ce qui paraît évident à l’observation clinique informelle de cette culture. En même temps, ce que nous disent les deux équipes d’investigation, c’est que, dès lors qu’on se met en peine de chercher, on trouve aisément des réponses R^\diamond – dont, rappelons-le, il n’est pas envisagé ici de juger la « valeur ». Arrêtons-nous sur le cas de la question 9. L’équipe 1 signale que, sous le titre « Rechercher efficacement sur les newsgroups », le *Guide complet* consacre une page et demie au thème correspondant. L’équipe 2 renvoie à une page du site de Christophe Asselin ainsi qu’à une page du site *Learn the Net*⁴⁷. Pour la question 11, l’équipe 1 propose une page du site *WebRankInfo*, ainsi que la section intitulée « Utiliser les pages de mise en cache » du *Guide complet* déjà cité⁴⁸. L’équipe 2, quant à elle, renvoie à un billet de Lauren Baker du 10 avril 2006 dans le *Search Engine Journal* intitulé “Google cache : searching Wikipedia when it’s down” (l’adresse n’est pas fournie, mais le document est aisé à trouver) ; en même temps elle signale la section du *Guide complet* que mentionnait déjà l’équipe 1.

Examinons maintenant le cas de la question 4, à propos de laquelle le contact avec des réponses est déclaré fort, alors que de telles réponses seraient « assez difficiles » à trouver. Rappelons-en le contenu : « Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ? » On peut imaginer qu'il apparaisse « culturellement correct » de déclarer que des réponses à cette question ne nous sont pas étrangères, alors même que, en sens inverse, un certain nombre d'acteurs du Web ont intérêt à ne pas trop s'appesantir sur la question : double conjecture qui irait dans le sens du résultat observé. La difficulté reconnue par les équipes d'investigation pour trouver des réponses à la question 4 s'exprime dans leur relevé de conclusions : deux des adresses proposées d'abord apparaissent non pertinentes ; une autre adresse (sur le site du *Journal du Net*) est celle d'un... *pop-up*, et une quatrième adresse (toujours sur le *Journal du Net*) est, elle aussi, peu pertinente. L'équipe 1 a eu plus de chance avec les ouvrages imprimés : elle mentionne à juste titre la page 67 du *Guide complet* ainsi que la page 47 des *Bons usages d'Internet*. L'équipe 2 renvoie d'un côté à un papier non signé publié sur le site de *L'Atelier*, qui est le « Centre de veille technologique de BNP Paribas », d'un autre côté au centre d'aide de Google, sur une page où figure une section intitulée « Liens commerciaux »⁴⁹. À cela s'ajoutent d'autres références dont plusieurs sont signalées dans le relevé de conclusions comme ne contenant « pas grand-chose » ou n'étant « pas très pratiques ». Le « butin » est complété par la mention d'une page du site de l'Urfist intitulée *Internet : aspects critiques* comportant un développement sur le thème « Liens naturels ou commerciaux »⁵⁰. Le sentiment qui se dégage de l'ensemble des indications relatives à la question 4, c'est que, si l'on trouve dans le corpus documentaire certains éléments de réponse à cette question, ce n'est qu'assez difficilement et, en fait, sans qu'on parvienne à mettre la main sur une réponse « franche » (même si elle se trouvait être erronée). Les fragments de réponse mis au jour paraissent en effet n'aborder le sujet que de façon quelque peu oblique, allusive, ou du moins qui peut paraître telle à un investigateur un rien soupçonneux. Si la question 4 est bien une question à laquelle d'aucuns pensent qu'il est nécessaire de disposer d'une réponse, le sentiment d'en disposer *effectivement* paraît fragile, peut-être parce que la question posée est vécue comme légèrement sulfureuse.

Les questions 13, 19 et 20 apparaissent, dans notre tableau simplifié, comme susceptibles de réponses qui n'auraient que peu été rencontrées et qui seraient en même temps difficiles à trouver. La question 13 a reçu deux fois la mention « très difficile à trouver », la question 20 a été jugée par l'équipe 1 « très difficile à trouver » et par l'équipe 2 « difficile à trouver ». La question 19 a reçu, elle, des jugements plus contrastés : pour l'équipe 1, une réponse à cette question est « difficile à trouver » alors que pour l'équipe 2 une telle réponse

est tout simplement *introuvable*. En pratique, le relevé de conclusions de cette dernière équipe assigne à cette question le score 0 *sans aucun commentaire* : dans le temps et les conditions qui lui étaient imposées, l'équipe a apparemment renoncé à trouver. Qu'en est-il alors de l'équipe 1 ? Son relevé de conclusions propose un unique document⁵¹, dont l'examen ne permet pas de voir qu'il fournisse une réponse à la question posée : le score aurait dû être en ce cas 0, et non 2.

Construire des réponses R^\heartsuit ?

Les quatre types de questions que nous avons dégagés jouent chacun un rôle clé dans l'organisation et le développement d'une culture. Les questions de la catégorie « Contact fort » (celles de la deuxième ligne du tableau ci-dessus) constituent, à un moment donné, « l'intérieur » d'une culture de la RII. Pour fixer les idées, nous reproduisons une fois encore celles que nous avons vu entrer dans cette catégorie parmi les dix que nous avons « sélectionnées ».

2. Quelle différence y a-t-il entre un « site Web » et une « page Web » ?
4. Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ?
15. Avec un moteur de recherche, peut-on chercher les pages Web qui contiennent tel mot mais pas tel autre ?
17. Dans un travail universitaire (mémoire, présentation, etc.), comment citer une page Web ?

Il s'agit là, en principe, de questions auxquelles des réponses sont déclarées suffisamment « connues » (le total des scores assignés à ces questions vaut respectivement 50 %, 54 %, 45 % et 59 % du total maximal possible) pour que ces questions soient « reconnues », c'est-à-dire apparaissent légitimes (comme questions) dans la culture où évoluent les enquêtés – en gros, la culture universitaire française des usagers non spécialistes de la RII –, en sorte que le problème de la construction d'une réponse R^\heartsuit de la culture considérée *pourrait* être posé à leur propos. Dans tous les cas, pourtant, l'inexistence d'un *espace de problématisation* légitime, à forte puissance d'investiture épistémologique, peut laisser jouer la propension toujours active au *déni de problématicité*, si bien que l'on se contentera des réponses (supposées) que l'on croit posséder – en gros, celles dont devaient se remémorer l'existence les enquêtés ayant attribué de forts scores aux questions concernées (ils sont 40 % pour la question 2, 46 % pour la question 4, 35 % pour la question 15 et 53 % pour la question 17 à

avoir attribué les scores 3 ou 4). Pour le type de questions que représente la question 4 dans la catégorie « Contact fort », l'évitement du problème de construire une réponse R^\heartsuit est d'autant plus favorisé que le matériau utilisable pour construire R^\heartsuit , à savoir des réponses R^\diamond , *manque*. Pour le type de questions représenté par les questions 2, 15, 17, où les réponses R^\diamond ne manquent pas, la tentation peut être forte de s'en tenir à telle réponse « toute faite », à « poser $R^\heartsuit = R^\diamond$ », pour parler comme les mathématiciens, en faisant l'économie du travail d'analyse et d'évaluation de ces réponses R^\diamond préalable à la construction d'une réponse R^\heartsuit , et du travail de cette réponse proprement dite. Tel est, schématiquement exprimé, le danger d'une *stase praxéologique* dans la culture de la RII considérée.

Les questions de la catégorie « Contact faible » (celles de la troisième ligne du tableau) composent à un moment donné ce qui est en quelque sorte la *frontière* et *l'extérieur* de la culture de la RII considérée, zones que, semble-t-il, on peut identifier respectivement au type « Recherche facile » pour la première et au type « Recherche difficile » pour la seconde. Pour ce qui est de la frontière, les questions qui en relèvent (parmi les dix choisies) sont les suivantes.

9. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?

11. Qu'est-ce que « le cache de Google » et à quoi cela peut-il être utile ?

12. Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc.) ? Comment le faire ?

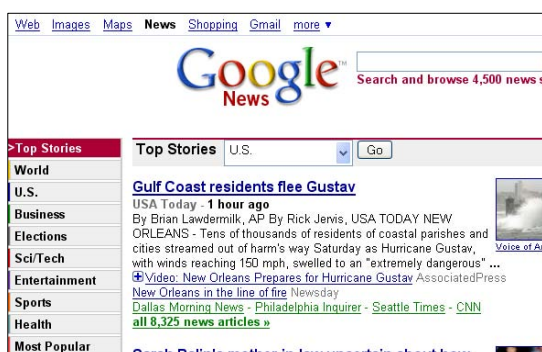
Ce sont là des questions qui, en principe, pourraient être intégrées à l'intérieur de la culture, pour peu que celle-ci voie ses frontières se déplacer : d'après l'ensemble des analyses réalisées dans ce travail, et à voir les spécimens de questions ci-dessus, on imagine que ce n'est pas là un phénomène qu'il est facile de provoquer et de réguler ! Les questions du type « Recherche difficile » semblent avoir moins de chance encore de s'intégrer dans une culture de la RII. Rappelons-les également.

13. Il arrive que, lorsqu'on tape l'URL d'un site Web dans la barre d'adresse d'un navigateur, on arrive sur un autre site. Pourquoi cela ? Comment peut-on l'éviter ?

19. Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ?

20. Quand, par erreur, on tape une requête dans la barre d'adresse d'un navigateur, on obtient souvent des résultats comme s'il s'agissait d'un moteur de recherche (et non d'un navigateur) ? Pourquoi ?

Ici, conditions économiques et contraintes écologiques (en matière de RII) semblent faire barrage. Considérons ainsi la question 19, la seule déclarée « introuvable » par nos enquêteurs. Si l'on se fie à leur jugement, cela signifie essentiellement que cette question *n'est pas posée* et n'est pas « reconnue » comme question – ni *a fortiori* comme *problème* – dans l'étendue du corpus documentaire qu'ils ont explorée. De fait, cette question semble tellement absente de la culture ambiante de la RII que sa bonne compréhension de la part des enquêteurs n'est pas assurée : elle signifie que si, depuis un ordinateur « français », on tape l'URL <http://news.google.com> présentée par Notess (2006, p. 159) comme celle de *Google News*, on voit s'afficher la page *Google Actualités* – en français, à propos de l'actualité « française ». Comment faire alors pour accéder à l'actualité états-unienne, celle que Notess voit s'afficher sur *son* ordinateur ? Telle est la question. Or, en vérité, il suffit de *poser* cette question pour que la réponse saute aux yeux. Sur la copie d'écran ci-après (capturée le 31 août 2008), on aperçoit, à côté de la mention *À la une*, une boîte de liste déroulante (*drop-down list box* ou *combo box*) où l'on peut choisir entre une longue liste de pays : on a reproduit ci-dessous, à titre d'exemples, le site « américain » et le site espagnol.



Dans ce « blanc » du discours sur la RII ⁵², on peut sans doute voir les effets d'un certain confinement culturel, alors même que le fait de pouvoir ainsi disposer à tout instant des « nouvelles » venant de quelque cinquante pays différents peut constituer en certains domaines un outil de travail extrêmement précieux. À tout le moins, on retrouve ici, *in fine*, que le *problème des questions* est l'alpha et l'oméga de toute dynamique « herbartienne » de la culture de la RII.

Conclusion

Entrer dans la culture d'un univers institutionnel donné nous permet d'accéder à des équipements praxéologiques qui, sinon, auraient pour nous un coût d'accès exorbitant : une culture rend proches les praxéologies qu'elle porte en elle, auxquelles elle permet à ses membres de s'initier selon des procédures réglées. De ce point de vue, une culture n'est jamais si efficace que lorsqu'elle devient *folklore*, diffusant alors des praxéologies réputées « bien connues », et qui « vont de soi » pour les acteurs de cette culture. Bien entendu, à cet avantage essentiel répond un inconvénient non moins certain, dont témoigne peut-être le sens familier que l'adjectif *folklorique* a pris en français depuis un demi-siècle⁵³ : celui qui tend à substituer, à un régime praxéologique « contrôlé », un régime « rumoral » diffusant, sans ostentation mais sans façon, astuces techniques « simples », justifications lapidaires, « théorisations » bricolées, toutes à fort pouvoir de percolation dans le complexe institutionnel concerné – à moins que ses produits ne s'y heurtent à un barrage écologique, que celui-ci prenne la forme d'un refus idéologique, d'un rejet « savant », ou de quelque autre forme de résistance institutionnelle. D'une façon générale, une culture est affectée par une double dynamique : d'un côté, elle est le fruit d'un « travail » spécifique engendrant des dynamiques herbartiennes diverses ; d'un autre côté, elle se trouve l'objet d'une folklorisation rampante.

L'analyse d'une culture en tant que support de diffusion praxéologique doit ainsi l'interroger tant sur son *étendue* – quelle fraction des institutions et des personnes concernées « baigne-t-elle » *de quelles praxéologies* ? – que sur sa *profondeur* – définie par la longueur moyenne des chaînes explicatives des praxéologies dont elle soutient la diffusion. Quand on se tourne alors vers la culture de la RII de « l'utilisateur ordinaire » telle que semblent vouloir la façonner les exposés du type de ceux étudiés dans la deuxième partie de notre travail et encore dans le premier chapitre de cette quatrième partie, ou même vers ce que nous avons appelé la « culture universitaire française des usagers non spécialistes de la RII », on doit conclure que cette profondeur est limitée presque artificiellement, et que ces cultures de la RII se trouvent coupées en grande partie de leurs racines technologico-théoriques, condamnant ainsi l'utilisateur qui s'y trouve confiné à demeurer à la surface d'une « science de la RII » qu'il faudrait dès lors aller chercher ailleurs, plus près des concepteurs et des créateurs des entités qui peuplent le Web.

L'étendue, on l'a vu, se définit par deux paramètres non indépendants : *quelles praxéologies*, et *pour qui* ? À propos du premier paramètre, la limitation de l'étendue est illustrée *en particulier* par les praxéologies entrant dans la catégorie désignée dans cette partie par l'étiquette « Contact faible, recherche facile ». On constate ainsi que tel groupe d'utilisateurs peut se trouver dans l'ignorance totale et durable de tel point de la vie du Web, sur lequel il est pourtant – grâce au Web ! – on ne peut plus facile de s'informer, dès lors bien sûr que la question adéquate a été posée. À titre d'illustration, nous nous permettrons de témoigner ici d'une « découverte » faite au cours de notre recherche : dans la troisième partie de ce travail, nous avons été amenée à mentionner les « treize serveurs de la racine » ; or nous avons pu observer, de façon certes informelle mais non moins convaincante, que nombre de familiers du Web, regardés dans leur environnement institutionnel comme des plus compétents en la matière, n'en avaient jamais entendu parlé. En d'autres termes, ils avaient eu jusque-là, avec cette réalité cruciale de la vie du Web, un contact *nul*.

Personne n'ignore comment il est convenu de rendre compte de ce type de faits : en invoquant une *spécialisation*, qui, au reste, peut être plus ou moins douce ou plus ou moins brutale : telle connaissance serait en effet réservée à tel groupe de personnes définies par leur « spécialité » (par exemple les informaticiens universitaires ou les référenceurs professionnels), les autres étant décrétées « ne rien avoir à en faire ». On notera en passant que, si une culture augmente fortement l'accessibilité à certains équipements praxéologiques, ainsi qu'on l'a souligné, elle diminue considérablement, dans le même temps, les possibilités d'accéder à d'autres praxéologies, regardés alors, culturellement, comme à peu près hors de portée. La gestion commune de la diffusion praxéologique dans une société par la construction de « spécialités » participe d'une raison économique en matière praxéologique qui cherche à assurer la satisfaction de certains besoins sans toucher pour l'essentiel aux contraintes écologiques qui sembleraient devoir empêcher une diffusion de masse non spécialisée. Or ces logiques de spécialisation, inévitables sans doute dans leur principe, mais que l'on regarde trop, nous semble-t-il, comme indépassables, ne sauraient s'articuler immédiatement au projet dans lequel nous situons notre travail, celui d'élucider les conditions de possibilité d'une culture *citoyenne* de la RII, qui, selon la formule utilisée autrefois par l'historien Lucien Febvre (1878-1956) à propos de l'*Encyclopédie française* (1935-1966), doit « aller droit aux problèmes que l'homme non spécialisé porte en lui », ce qui ne se réduit pas, nous le savons, aux problèmes que cet usager non spécialisé *se pose*, spontanément, et de façon explicite (Chevallard, 1998b).

À côté de l'assèchement que le « tout spécialisation » peut provoquer dans un projet de culture de la RII *pour tous* (qui, du point de vue l'élémentation des savoirs, doit nous

permettre d'accéder à *toute* culture de la RII, si sophistiquée soit-elle, au lieu de nous en éloigner), nous noterons aussi un phénomène moins évident, mais bien réel, en l'éclairant d'un exemple dont l'un des mérites est d'être d'une actualité récente. Un billet intitulé "Yahoo! drops Boolean support", mis en ligne sur le site *Pandia Search Engine News* le 15 août 2008 (Koch & Koch), rend compte d'un article publié par Greg Notess dans le numéro de juillet-août de la revue *Online*. On y annonce que, selon Notess, Yahoo! a cessé de donner son sens (booléen) au mot NOT employé (en majuscules) dans une requête, même si l'utilisateur de Yahoo! peut quand même continuer d'utiliser le signe moins comme on le sait). Le même billet précise encore : "Yahoo! has also dropped support for the AND operator. This is not equally serious, as Yahoo! will do an AND search by default. The OR operator still works, but nesting with parentheses fails." L'auteur du billet oppose alors les pratiques des usagers ordinaires à celles des « spécialistes ».

Yahoo! has apparently come to the conclusion that since so few uses Boolean anyway, they may as well not support it. That is a mistake. Power-users like librarians, journalists and researchers use them, and they are exactly the kind of people that influences the public opinion in these matters.

Bring back Boolean, Yahoo!

On voit ici typiquement que le non-franchissement de la frontière des booléens par le grand public a des effets à long terme – ici, la décision prise par Yahoo! – contre lesquels le fait d'être un « spécialiste » ne prémunit pas ! Les « spécialistes » seraient même, en l'espèce, les seules vraies victimes de ce « délestage » réalisé par Yahoo!. Les groupes spécialisés, professionnellement ou non, sont ici, comme en bien des cas, solidaires de l'utilisateur ordinaire, avec lequel, trop souvent, on ne compte que négativement.

On aura noté peut-être combien est en vérité subtile, parfois, la cartographie de l'espace praxéologique de la RII : passer de l'usage de NOT à celui du signe moins (-) nous ferait ici franchir une frontière ! Le premier problème dont peut pâtir une culture de la RII est bien celui de son enfermement dans des frontières qui tendent à s'imposer *contre* les besoins des usagers, par pure inertie « culturelle ». De tels tracés, d'apparence arbitraire quand on les regarde dans une perspective fonctionnelle, sont souvent les vestiges d'un passé (récent) que la pression des besoins des usagers n'a pas subvertis parce que cette pression ne trouve pas à s'exprimer – faute de « forums » adéquats – sous forme de *problèmes* largement reconnus comme tels. Tel est, on l'a suggéré, l'un des problèmes cardinaux de la construction d'une culture de la RII. Si les forums existent peu, les « boutiques », elles, existent ; mais, précisément, leur offre praxéologique n'est pas mise en question comme il conviendrait, à

l'aune de besoins déclinés en questions : l'apparent « plein » de l'offre en dissimule alors les lacunes innombrables, au détriment de ce qui serait une culture développée et partageable, qui offre des réponses adéquates aux questions que l'utilisateur « porte en lui ». L'absence de questions masque la pénurie de réponses. Lorsque, ainsi, on découvre l'existence de métamoteurs de recherche, on peut trouver l'idée excellente. Mais cette découverte devrait aller de pair avec une question évidente : « Comment, en consultant divers moteurs (tels Google, Yahoo!, Exalead ou Live Search), le “métamoteur” constitue-t-il la liste de résultats qu'il affiche ? » C'est là, nous semble-t-il, une question *minimale* dans une culture de la RII soucieuse de l'outillage intellectuel et du confort psychologique de ses sujets. Considérons alors deux des « exposés » évoqués jusqu'ici. Dans celui signé d'Isabelle Bréda et Carole Hourt (2007), on lit ceci, qui peut conduire à penser qu'interroger un métamoteur équivaut à interroger *plusieurs* moteurs *d'un seul coup* : « Quant au métamoteur, il permet d'interroger plusieurs moteurs de recherche à la fois, en un seul clic, et présente l'ensemble des résultats sur une page. On accède ainsi à une plus grande partie du Web. » De l'entretien avec Véronique Mesguich et Armelle Thomas (2006b), de même, on retiendra cette information, compatible avec la « réponse » précédente (bien qu'elle franchisse en catimini une frontière, en parlant des index des moteurs) : « Les méta-moteurs, quant à eux, ne maintiennent pas d'index, mais permettent de rechercher simultanément dans les index de plusieurs outils. » Mais voici maintenant un autre fragment de réponse, *qui change tout* : parlant des *metasearch engines*, Ran Hock (2007, p. 74) écrit en effet sans détours que “most only return the first 10 or 20 records from each source. If record No. 11 in one of the search engines was a great one, you may not see it.” Le *travail* possible sur la question soulevée – travail que nous ne poursuivrons pas ici – commence alors à prendre tournure.

Le problème des *ateliers*, auquel renvoie l'exemple précédent, est peut-être celui qui, au terme de cette recherche, restera le plus obscur. S'il y a peu de questions, il y a tout de même quelques réponses, dont on peut faire provision dans des boutiques dont la variété de l'offre n'est pas, certes, le trait le plus remarquable. Mais en RII comme ailleurs, l'arrière-boutique où se fabrique ce qui sera offert à l'utilisateur reste dans la pénombre. Il n'est pas facile, ainsi, de connaître la genèse des réponses offertes, le passage, entre les mains d'un auteur, de la question Q à la réponse R^\heartsuit qu'il « livrera » à l'utilisateur – et qui ne sera peut-être qu'une copie, affadie ou rehaussée selon le cas, de quelque réponse R^\diamond . C'est là un problème profond, qui marque une limite volontaire de notre recherche, sur laquelle nous reviendrons brièvement dans notre *Conclusion générale*.

ANNEXE 4

“How to search” : un exposé en questions

101, § 1

Qu'appelle-t-on concordance d'une œuvre écrite ?

En 1829, Mary Novello, épouse de Charles Clarke, commença à établir une concordance des œuvres de Shakespeare. Avant cela, pour quel texte avait-on établi une concordance ?

Combien de temps fallut-il à Mary Novello pour établir une concordance de l'œuvre de Shakespeare ? Combien d'entrées comportait cette concordance ?

101, § 2

Qu'indiquait la concordance des œuvres de Shakespeare établie par Mary Novello pour chacun des mots extraits de ces œuvres ?

Quel était l'intérêt proclamé de disposer d'une concordance des œuvres de Shakespeare ?

102, § 1

Comment établissait-on l'index d'une œuvre ou d'un ouvrage avant l'apparition de l'informatique ?

Pour quelles raisons une concordance était-elle appelée ainsi ?

102, § 2

Pourquoi la concordance publiée en 1897 du Nouveau Testament en grec n'aurait pas pu être entièrement mécanisée ?

102, § 3

Quelle organisation humaine et matérielle fut adoptée pour établir la concordance de l'œuvre poétique de Wordsworth publiée en 1911 ? Combien d'entrées comportait-elle ? Combien de temps fut nécessaire pour l'établir ?

102, § 4

À quelle époque put-on établir des concordances à l'aide d'ordinateurs ?

De quels défauts souffraient les premières concordances établies à l'aide d'un ordinateur mais imprimées en capitales, sans ponctuation ?

À partir de quelle époque devint-il possible, à l'aide des ordinateurs personnels, de dépasser en qualité les concordances anciennes, faites à la main ?

103, § 1

Comment établir une concordance d'un texte dont on possède une version électronique, ou d'un ensemble de pages Web, ou un livre tombé dans le domaine public ?

Qu'es-ce que le projet Gutenberg ?

Qu'est-ce que l'Internet Archive ?

103, § 2

Les notions de concordance et de "full-text index" (index en texte intégral ou en plein texte) ont-elles à voir l'une avec l'autre ?

Quelles recherches peut-on faire concernant un texte pour lequel on dispose d'un index en texte intégral ?

Searching Text

104, § 1

Lorsqu'on lui soumet une requête, un moteur de recherche cherche-t-il les mots qui la composent dans l'ensemble des textes disponibles sur le Web ?

Full-text Indexes

104, § 2

Pourquoi un moteur de recherche utilise-t-il un index plein texte ?

Considérons la phrase suivante : *to be or not to be that is the question*. Que serait un index plein texte du texte suivant, en utilisant comme pointeur le rang du mot dans la phrase (comme indiqué ci-après) : to be or not to be that is the question ?

Que signifie le mot *pointer* dans la phrase suivante : *a full-text index lists all the words that appear in the text, along with a pointer to every occurrence* ?

Quelle relation y a-t-il entre le nombre de pointeurs d'un index plein texte et le nombre de mots du texte ?

105, § 1

Dans un index plein texte le pointeur d'une occurrence d'un mot est-il toujours le numéro du rang de ce mot dans le texte ? Comment faire lorsque l'index concerne un ensemble de textes (par exemple un ensemble de pages Web, ou les pièces de théâtre écrites par Shakespeare) ?

105, § 2

Dans un index relatif à un ensemble important de textes (par exemple 1 milliard de pages Web ou 2 000 ouvrages imprimés), y a-t-il plus (voire beaucoup plus) d'entrées, ou moins (voire beaucoup moins) d'entrées, que dans un dictionnaire de langue en plusieurs volumes ?

L'index de Google contient-il plusieurs milliards d'entrées ou plutôt seulement plusieurs millions d'entrées ? Pourquoi ?

105, § 3

Qu'appelle-t-on un *hapax legomenon* ?

Dans les *Complete works of Shakespeare*, qui contiennent environ 31 500 mots différents, quel est le pourcentage de mots qui n'apparaissent qu'une fois ? Quel est, approximativement, le nombre d'occurrences des mots les plus souvent utilisés ?

Google aurait déclaré que son index référence 21 630 000 000 pages Web contenant au moins une occurrence d'un certain mot de la langue anglaise. De quel mot peut-il s'agir ? Dans cet index, quel peut être le nombre de pointeurs relatif à ce mot ?

Dans un index, quelle allure a la distribution du nombre d'occurrences des différents mots indexés ?

105, § 4

Peut-on dire qu'un index est aussi gros que le corpus de textes qu'il indexe ? Pourquoi ?

Lorsqu'on applique une technique de compression, est-il vrai que l'index diminue sensiblement plus que le corpus de textes lui-même ? Pourquoi ?

Est-il exact que les moteurs de recherche disposent et d'un index plein texte et du texte lui-même des pages trouvées sur le Web ? Ces deux ensembles de données occupent-ils sensiblement le même espace ? Pourquoi ?

Using the Index

106, § 1

Lorsqu'on dispose de l'index d'un corpus de documents, comment est-il possible de trouver les documents du corpus contenant un certain mot ?

Est-il plus rapide pour un ordinateur d'examiner un texte à la recherche d'un certain mot ou de faire la même chose sur une liste ordonnée de mots ?

106, § 2

Comment procède-t-on pour répondre à une requête relative à plusieurs mots, c'est-à-dire pour choisir les documents indexés contenant chacun de ces mots ?

Comment procède-t-on pour répondre à une requête constituée d'une combinaison logique de mots (incluant tel mot ou tel autre, excluant tel autre mot encore, etc.) ?

106, § 3

Comment procède-t-on pour identifier les documents indexés contenant une expression exacte donnée ?

Comment procède-t-on pour identifier les documents indexés contenant deux mots qui, dans le document, sont proches l'un de l'autre, c'est-à-dire pour effectuer une recherche de proximité (*proximity search*) ?

106, § 4

Dans un corpus de documents indexés par un moteur de recherche, l'ordre des documents a-t-il un sens par rapport à une requête donnée ? Pourquoi ?

106, § 5

Étant donné une requête composée de plusieurs mots, quels critères peut-on utiliser pour estimer la plus ou moins grande pertinence des documents contenant ces mots en totalité ou en partie ?

Étant donné une requête composée de plusieurs mots, auxquels de ces mots est-il logique de donner le plus d'importance dans le choix des documents pour répondre à la requête ?

107, § 1

Étant donné une collection de documents, on peut représenter un document de la collection par un point dans un espace dont le nombre de dimensions (qui peut être alors de plusieurs milliards) est égal au nombre de mots, chaque mot déterminant un axe de l'espace et chaque document ayant pour coordonnée sur cet axe le nombre d'occurrences du mot dans ce document. Par exemple si une collection de documents ne comporte que cinq mots, m_1 , m_2 , m_3 , m_4 , m_5 , et si un certain document comporte 3 occurrences de m_1 , 24 occurrences de m_2 , 0 de m_3 , 68 de m_4 et 7 occurrences de m_5 , ce document sera représenté dans un espace à 5 dimensions par le point de coordonnées (3, 24, 0, 68, 7). Comment est-il logique de mesurer les distances sur chaque axe si l'on considère que, si un mot est rare dans la collection de documents, une occurrence de plus ou de moins fait une plus grande différence que si le mot est plus fréquent ?

Lorsqu'on cherche un mot dans une collection de documents où ce mot n'a que de rares occurrences, on peut avoir l'impression que chacune de ces quelques occurrences nous informe sur le mot « en général ». Si, par contraste, ce mot a de nombreuses occurrences dans *l'un* des documents de la collection, on peut avoir l'impression que ce fait nous informe sur *ce document* (plutôt que sur le mot lui-même). Comment ces impressions sont-elles intégrées dans la modélisation classique d'une collection de documents par un espace dont chaque axe correspond à l'un des mots présents dans la collection ?

107, § 2

Lorsqu'on représente les documents d'une collection par un point dans un espace selon le modèle classique, deux documents représentés par des points distincts alignés avec l'origine comportent les mêmes mots dans les mêmes proportions (mais pas en même quantité) : du point de vue d'une requête donnée, ces deux documents n'ont pas à être distingués. Quand, au contraire, deux documents sont situés sur des droites différentes issues de l'origine, comment évalue-t-on la similitude (*similarity*) entre ces documents (cette similitude étant égale à 1, par définition, dans le cas de documents représentés par des points alignés avec l'origine) ?

107, § 3

Dans la modélisation classique d'une collection de documents par un espace dont le nombre de dimensions est égal au nombre de mots, la mesure de similitude à l'aide du cosinus traite comme identique un texte et tout autre texte issu du premier dont les mots ont été rangé dans n'importe quel ordre. Cela est-il préjudiciable à l'utilisation de cette mesure de similitude par un moteur de recherche ? Pourquoi ? Si oui, que faire ?

What's a Word?

107, § 4

Quels problèmes posent l'utilisation de la notion de mot dans l'indexation d'un texte ?

108, § 1

Dans un texte, qu'est-ce qui peut constituer un mot, c'est-à-dire une chaîne alphanumérique d'un seul tenant qui puisse apparaître comme une requête adressée au moteur de recherche ?

108, § 2

Qu'est-ce que le *case folding* (pli de casse) ? Quels problèmes cela entraîne-t-il ?

108, § 3

Qu'est-ce que le *stemming* (lemmatisation) ? Quels en sont les avantages et les inconvénients dans l'utilisation d'un moteur de recherche ?

108, § 4

Qu'est-ce qu'un *stop word* (mot vide) ? Quelle est la proportion approximative des *stop words* dans un texte ?

Que font les moteurs de recherche avec les *stop words* ? Y a-t-il un intérêt à cela du point de vue de l'espace occupé par l'index ? Lequel ?

108, § 5

Quelles conséquences fâcheuses peuvent avoir pour l'utilisateur la décision d'ignorer tel mot lors de la constitution de l'index (c'est-à-dire de l'inclure dans la liste des *stop words*) ?

109, § 1

Que reste-t-il du mot « AT&T » (l'acronyme de *American Telephone and Telegraph*) si le logiciel d'indexation transforme les capitales en bas de casse et ignore : 1. les caractères spéciaux (qui ne sont pas des lettres) ; 2. les mots d'une lettre ; 3. le mot "at" ?

Doing It Fast

109, § 2

Quel est le principe mis en œuvre pour résoudre le problème de la construction d'un index en texte intégral ?

109, § 3

Pour fabriquer un index d'une importante collection de documents, le problème clé est-il plutôt un problème de programmation ou un problème de capacité de mémoire ? Ou encore de rapidité d'accès aux données ?

109, § 4

Lors de l'indexation d'une vaste collection de documents, comment peut-on concilier l'énormité de l'ensemble des données à traiter et le caractère nécessairement limitée de la mémoire principale (*main memory*) ?

La conception d'un logiciel d'indexation utilise-t-elle peu ou beaucoup de mathématiques ? Pourquoi ?

109, § 5

Pour rechercher une entrée dans un index, on utilise en général la technique de dichotomie (*binary search*). De quoi s'agit-il ? Pourquoi procéder ainsi ?

Lorsqu'un logiciel de recherche a identifié une entrée (mot clé) dans un index, il doit encore parcourir les pointeurs correspondants à cette entrée, ce qui peut être long. Comment faire pour fournir au plus vite des résultats à l'utilisateur ? Comment faire si la requête comporte plusieurs mots clés ?

110, § 1

Peut-on saisir dans la boîte de requêtes d'un moteur de recherche une longue suite de mots clés, voire un paragraphe entier d'un texte ? Si oui, est-il tout de même plus efficace de se limiter à deux ou trois mots clés ? Pourquoi ?

Pour quelles raisons la boîte de requêtes (*query box*) d'un moteur de recherche a-t-elle en général une taille réduite ?

110, § 2

Le classement en fonction de leur mesure de similitude des documents trouvés en réponse à une certaine requête est-elle une opération plus délicate que la recherche de ces documents elle-même ? Pourquoi ?

Quand on classe des documents en utilisant la mesure de similitude du cosinus, est-il possible de trouver les 10 ou 20 documents les plus pertinents sans comparer tous les documents entre eux ? Si oui, de quelle nature sont les connaissances clés qui permettent d'y parvenir ?

110, § 3

Quelles qualités peut-on attendre d'un bon système d'indexation (*indexing system*) ?

La construction d'un système d'indexation de qualité est une affaire complexe. Une fois la chose faite, de quel avantage essentiel bénéficie-t-on ?

Evaluating the Results

110, § 4

Parmi les qualités d'un système d'indexation, quelle est la plus délicate à évaluer ?

110, § 5

En supposant qu'on puisse dire qu'un document est pertinent ou n'est pas pertinent par rapport à une requête, ce qui simplifie la réalité, sur quels critères peut-on juger de la qualité des résultats affichés par un moteur de recherche en réponse à une requête ?

111, § 1

Si par exemple on considère les 20 premiers résultats affichés par un moteur de recherche en réponse à une requête, qu'appelle-t-on *précision* de ces résultats ? Qu'appelle-t-on *rappel* ?

111, § 2

Que mesure la précision ? Que mesure le rappel ?

111, § 3

Que signifierait un rappel de 100 % ? Que pourrait-on dire alors de la précision ?

Que se passe-t-il en général lorsque la précision augmente ? Lorsque le rappel augmente ?

111, § 4

Si l'on ne considère par exemple que les 20 premiers résultats envoyés par un moteur de recherche en réponse à une requête, est-il possible de calculer la précision de ces résultats ? Comment ? Est-il possible de calculer le rappel ?

Searching in a web

111, § 5

De quand date les systèmes de recherche d'information en texte intégral, qui, étant donné un ensemble de documents et une requête, renvoient les documents les plus pertinents ? Les années 2000 ? Les années 1950 ? Les années 1980 ?...

112, § 1

Par rapport aux collections de documents antérieurement indexées, quelle était la principale spécificité du Web lorsqu'il est apparu au cours des années 1990 ?

112, § 2

Quand le Web fut créé, combien de temps fallut-il pour que les informaticiens parviennent à créer les premiers moteurs de recherche ?

112, § 3

Que faisaient les internautes avant l'apparition des moteurs de recherche ?

112, § 4

Qu'est-ce que l'apparition des moteurs de recherche a changé dans la fréquentation et l'usage du World Wide Web ?

112, § 5

De combien de mots se composent la majorité des requêtes formulées ? Un ? Deux ? Trois ?... Quelle est la requête la plus fréquente ?

Dans une collection de documents où un certain mot n'apparaît dans chaque document qu'au plus une fois, selon quels critères sont rangés les résultats affichés par un moteur de recherche s'il utilise l'algorithme du cosinus ?

Lors de leurs travaux pionniers, les créateurs du moteur de recherche Google, Sergey Brin et Larry Page, ont utilisé la requête consistant en l'unique mot *university*. De quoi cela leur fournissait-il un exemple (ou un contre-exemple) ?

Determining What a Page Is About

113, § 1

Est-il possible d'inclure dans une page Web du texte invisible du lecteur humain, et à l'adresse exclusive des moteurs de recherche ? Sous quelle forme ?

Quel peut être l'intérêt de placer sur une page Web du texte invisible du lecteur humain mais indexable par un moteur de recherche ? Cela peut-il donner lieu à des abus ? Lesquels ?

113, § 2

Est-il exact que les moteurs de recherche, quand ils indexent une page Web, indexent aussi le texte du lien par lequel le robot est arrivé sur cette page ? S'il en est ainsi, à quoi cela peut-il servir ?

Étant donné une page Web, qu'est-ce qui paraît *a priori* le plus crédible : les commentaires figurant sur cette page et visibles par les moteurs de recherche, ou les commentaires qui apparaissent dans le texte des liens conduisant à cette page ? Pourquoi ?

113, § 3

Est-il pertinent de considérer que, du point de vue de l'indexation d'une page Web, le texte des liens conduisant à cette page fait partie de la page ? Pour quelles raisons ? Faut-il alors donner à ce texte un « poids » plus petit ? Plus grand ?

Measuring Prestige

113, § 4

Étant donnée une page Web, qu'appelle-t-on son PageRank ? Qui en est à l'origine ? Qui l'utilise ?

Si l'on considère que le PageRank d'une page prétend exprimer son « prestige » numériquement, en quoi nous informe-t-il de l'« autorité » reconnue à son contenu ?

113, § 5

Pour quelles raisons peut-on considérer a priori qu'une page a d'autant plus de prestige que plus de liens pointent vers elle ?

114, § 1

Pour déterminer le prestige d'une page Web, faut-il donner plus de poids à un lien entrant qui vient d'une page envoyant peu de liens vers d'autres pages, ou plus d'importance à un lien émanant d'une page envoyant beaucoup de liens vers d'autres pages ? Pourquoi ?

114, § 2

Quelle différence peut-on voir entre le prestige apporté à une page Web par des liens dirigés vers elle et le prestige qu'une personne détient de naissance (par son appartenance à une famille, une lignée...) ?

115, § 1

Est-il possible pour l'auteur d'une page Web de susciter artificiellement la création de liens pointant vers cette page ? Pour quelles raisons le ferait-il ?

Pour une page Web est-il plus intéressant pour augmenter son prestige de recevoir un lien d'une page ayant un certain prestige et spécialisée dans l'envoi de liens ou d'une page ayant le même prestige mais qui n'envoie de liens qu'en fonction du contenu de la page ?

Les liens sortants d'une page (outlinks) contribuent-ils au prestige de la page ? Comment ?

115, § 2

Selon quels principes, en tenant compte de quels « paramètres », se définit le PageRank d'une page Web ?

Calculating PageRank

115, § 3

Comment peut être calculé le PageRank d'une page A s'il est vrai que celui-ci dépend du PageRank des autres pages du Web, et en particulier de ces pages vers lesquelles la page A envoie des liens ?

115, § 4

La définition du PageRank suppose-t-elle des outils mathématiques puissants ? Si oui, à quelle branche des mathématiques appartiennent-ils ?

Les problèmes mathématiques soulevés par le calcul du PageRank sont-ils inédits ? Si oui, en quoi ?

115, § 5

Pour représenter mathématiquement les pages du Web et les liens entre ces pages, on utilise une matrice de connexion (*connection matrix*). Comment est constituée cette matrice ? Combien de lignes et de colonnes possède-t-elle ? Que représente ce nombre ? Que représente chacune de ses lignes ? De ses colonnes ?

Pourquoi la plupart des coefficients (*entries*) de la matrice de connexion sont-ils nuls ?

116, § 1

Les outils mathématiques utiles pour calculer le PageRank était-il depuis longtemps disponible, ou a-t-il fallu les créer spécialement pour résoudre ce problème ?

A-t-on besoin de mathématiciens dans les sociétés développant des moteurs de recherche ? Si oui, y jouent-ils un rôle important ?

116, § 2

Tous les détails du calcul du PageRank sont-ils publics, ou bien s'agit-il pour une part de secret industriel ?

Lors de leurs premiers essais de calcul de PageRank, Larry Page et Sergey Brin utilisaient un processus de calcul itératif nécessitant combien d'itérations environ ?

Que peut-on penser être le temps nécessaire à Google pour calculer le PageRank pour l'ensemble du Web ? Selon quelle périodicité cette opération est-elle réalisée ?

116, § 3

Lorsqu'on recalcule le PageRank des pages du Web, est-il intéressant de prendre pour point de départ des itérations les valeurs de PageRank déjà disponibles ? Pourquoi ?

Lors du recalcul du PageRank, ce n'est que certaines pages et certains liens qui entraînent des mises à jour (appelées respectivement *node updates* et *link updates*). Est-il possible de procéder à un recalcul du PageRank limité le plus possible à ces pages et à ces liens ? Quels problèmes cela pose-t-il ? Ces problèmes sont-ils actuellement résolus ?

The Random Surfer

116, § 4

Que désigne-t-on en disant que le Web a une structure en nœud papillon (*bow tie*) ?

[Réponse voir WD, chap. 3, pp. 91-94]

Quels problèmes peuvent poser pour la définition et le calcul du PageRank le fait certaines pages n'aient pas de liens entrants (*inlinks*) ou qu'elles n'aient pas de liens sortants (*outlinks*) ?

116, § 5

Afin de préciser de manière satisfaisante le mode de calcul du PageRank, Sergey Brin et Larry Page ont imaginé un surfeur au hasard (*random surfer*). Qu'est ce qui caractérise le comportement de cet internaute dans le cas général, lorsqu'il se trouve sur une page Web possédant des liens sortants (*outlinks*) ?

Est-il exact que la probabilité que le surfeur aléatoire de Brin et Page arrive sur une page A est proportionnel au PageRank de cette page ?

117, § 1

Quels problèmes pose *a priori* au surfeur aléatoire de Brin et Page le fait d'arriver sur une page dépourvue de liens sortants ?

Qu'appelle-t-on un ... (*PageRank sink*) ? Quand un ensemble de pages constitue-t-il un tel piège à PageRank ?

117, § 2

Sous quelles conditions le surfeur aléatoire de Brin et Page est-il dans l'impossibilité d'aller sur une page ou sur un ensemble de pages ?

117, § 3

Quels problèmes posent, à propos du processus itératif de calcul de PageRank, le fait qu'il existe des pages dépourvues de liens entrants ou dépourvues de liens sortants ? Comment Brin et Page ont-ils précisés le comportement de leur surfeur au hasard afin de résoudre ce problème ?

Quelle est la différence dans le comportement du surfeur aléatoire de Brin et Page entre le surf et la téléportation ?

117, § 4

Que se passe-t-il si la probabilité de téléportation du surfeur aléatoire de Brin et Page est prise égale à 1 ? Que se passe-t-il si on lui donne la valeur 0 ? Quelle est la valeur donnée à ce paramètre par Brin et Page au début de leurs travaux (vers 1998) ? Quel intérêt peut-il y avoir à augmenter légèrement cette valeur ?

117, § 5

Si, lorsqu'il se fait téléporter, le choix de la page d'arrivée n'est plus fait au hasard (équiprobabilité) mais selon une autre loi de probabilité (certaines pages ayant une probabilité plus petites que d'autres d'être choisies), que se passe-t-il pour le calcul et les valeurs du PageRank ? Quel usage les gestionnaires d'un moteur de recherche peuvent-ils faire de ce phénomène ? Dans quel but ?

Combining Prestige and Relevance

117, § 6

Dans le cas d'une requête ne contenant qu'un mot, si un moteur de recherche utilisant PageRank classe les résultats qu'il présente à l'utilisateur par ordre décroissant du PageRank des pages correspondantes, quelle anomalie ce classement peut-il comporter ?

118, § 1

Dans la sélection des pages pertinentes par rapport à une requête comportant plusieurs mots quelle règle à propos de la présence de ces mots un moteur de recherche applique-t-il en générale ? En quoi cela diffère-t-il du choix des pages à l'aide de l'algorithme du cosinus ?

Pour estimer la pertinence d'une page par rapport à une requête, les moteurs de recherche ne donnent pas en général la même importance aux différentes occurrences d'un mot selon sa

situation sur la page. Quels aspects de cette situation donnent lieu à une prise en compte plus importante pour déterminer la pertinence ?

Quand on dit que l'occurrence d'un mot compte plus lorsqu'elle se trouve dans une ancre, de quel type d'ancre s'agit-il ?

Lorsqu'une requête se présente comme une suite de mots, quelles propriétés de cet ensemble de mots conduit à donner une plus forte pertinence à une page supposée les contenir tous ?

118, § 2

Que sait-on au juste de ce qui peut influencer la pertinence d'une page Web par rapport à une requête donnée pour un moteur de recherche donné ?

Qu'appelle-t-on « métatag » en HTML ? Que s'est-il passé au cours de la jeune histoire du Web qui expliquerait que, après avoir donné une grande importance aux mots figurant dans un métatag, les moteurs de recherche aient en général changé de point de vue ? Dans quel sens ?

118, § 3

Pour classer les résultats fournis à l'utilisateur, un moteur de recherche utilise une combinaison de critères de pertinence et de critères de prestige. Cette combinaison est-elle fixe ou variable dans le temps ? Pourquoi ? Que peut-on en dire plus précisément ?

Hubs and Authorities

118, § 4

Au moment même où était inventé par Larry Page et Sergey Brin le PageRank (vers 1998), un autre algorithme était mis au point pour classer les pages Web jugées pertinentes. Quel en était le créateur ? Comment s'appelait-il ?

119, § 1

L'algorithme HITS (Hypertext-Induced Topic Selection) se réfère à deux propriétés que peut posséder une page, celle d'être un *hub* (trad.) et celle d'être une *authority* (autorité). Comment se définissent informellement ces propriétés ?

119, § 2

L'algorithme HITS calcule pour chaque page son score de *hub* (*hub score*) et son score d'autorité (*authority score*). Comment se définit informellement le score de *hub* (respectivement, d'autorité) d'une page Web ? À quelle branche des mathématiques faut-il recourir pour en tirer des définitions dénuées de circularité ?

Finding Hubs and Authorities for a Query

119, § 3

Une requête étant soumise à un moteur de recherche qui utiliserait l'algorithme HITS de calcul des scores de *hub* et des scores d'autorité, sur quel ensemble de pages Web cet algorithme est-il mis en fonctionnement ?

119, § 4

Comment appelle-t-on la partie du Web à laquelle est appliqué l'algorithme HITS de calcul des scores de *hub* et des scores d'autorité pour une requête donnée ? [*neighborhood graph of the query*]

Rencontre-t-on dans la mise en œuvre de l'algorithme HITS de calcul des scores de *hub* et des scores d'autorité des problèmes analogues à ce que l'idée de téléportation permet de résoudre pour ce qui concerne l'algorithme PageRank ? Y a-t-il d'autres problèmes ?

Si un moteur de recherche utilise l'algorithme HITS de calcul des scores de *hub* et des scores d'autorité, la mise en œuvre de cet algorithme doit se faire pour chaque requête présentée au moteur (et non pas par exemple une fois par mois pour l'ensemble du Web). Quelles différences y a-t-il à cet égard avec la mise en œuvre de l'algorithme PageRank ?

120, § 1

En quoi un moteur de recherche comme Google bute-t-il sur le phénomène de synonymie (expression différente désignant à peu près la même chose) ? En quoi un moteur de recherche utilisant l'algorithme HITS résout-il ce problème ?

Using Hubs and Authorities

120, § 2

Une requête lui étant soumise, un moteur de recherche utilisant l'algorithme HITS calcule le score de *hub* et le score d'autorité de chacune des pages du graphe de voisinage de la requête. Comment l'utilisateur peut-il tirer profit de ces scores ? Quel est l'intérêt différentiel d'une page ayant un score de *hub* élevé et d'une page ayant un score d'autorité élevé ? Que peut-on dire quant à la présence des mots composant la requête sur une page ayant un score d'autorité élevé ?

120, § 3

Quels sont les principaux avantages et inconvénients du fait que le calcul des scores de *hub* et des scores d'autorité de l'algorithme HITS dépend de la requête proposée ?

Pourquoi peut-on dire que le PageRank ne sert pas seulement à classer les pages proposées par Google ? Pour quelles autres tâches constitue-t-il un indicateur intéressant ?

Lorsqu'un moteur de recherche utilise l'algorithme HITS le calcul des scores de *hub* et des scores d'autorité pratique-t-il la mise en cache de certaines pages Web ? Dans quels cas ?

120, § 4

Comment créer artificiellement un « bon » *hub*, c'est-à-dire une page Web ayant un score de *hub* élevé ?

En supposant qu'on a créé un « bon » *hub*, comment créer alors une page ayant un score d'autorité élevé ?

L'usage de l'algorithme HITS par un moteur de recherche permet-il, à l'instar de PageRank, de diminuer les scores d'une page volontairement lorsque les gestionnaires du moteur de recherche concerné s'aperçoivent que cette page a un ou des scores artificiellement gonflés ? Quel mécanisme disponible avec PageRank ne l'est pas avec HITS ?

Peut-on dire que l'algorithme HITS est plus transparent à l'utilisateur ordinaire que l'algorithme PageRank utilisé par Google ? Pourquoi ?

121, § 1

L'algorithme HITS souffre d'une faiblesse liée au phénomène appelé « dérive du sujet » (*topic drift*), qui est engendrée par la présence éventuelle dans le graphe de voisinage de la requête d'une page dotée d'une très forte autorité, mais qui porte sur un sujet différent, quoique voisin de la requête et qui l'emporte de ce fait sur toutes les pages réellement pertinentes par rapport à la requête. Comment cela est-il possible ?

Lorsqu'un moteur de recherche utilise l'algorithme HITS, la requête proposée joue un rôle évidemment déterminant dans la définition de son graphe de voisinage. Joue-t-elle encore un rôle déterminant, et lequel, lorsque ce graphe de voisinage ayant été déterminé, le calcul des scores est lancé ?

121, § 2

Comment le problème de la dérive du sujet, qui affecte l'algorithme HITS de calcul des scores de *hub* et des scores d'autorité, peut-il être traité ?

Discovering Web Communities

121, § 3

Comment peut-on définir une « communauté Web » ? Quels exemples peut-on en citer *a priori* ?

121, § 4

Qu'appelle-t-on, dans un graphe, un cluster autoréférentiel ? Quels outils mathématiques sont-ils utiles pour déterminer ces parties d'un graphe ? Ces outils sont-ils applicables au graphe particulier que constitue le Web ? Pourquoi ?

Quel nom donne-t-on aussi à la recherche des clusters autoréférentiels d'un graphe ?

121, § 5

Étant donné une requête, à quoi correspondent les communautés Web que l'on peut identifier éventuellement à l'intérieur de son graphe de proximité ? En quoi leur connaissance peut-elle être utile à l'utilisateur ?

121, § 6

En supposant un moteur de recherche utilisant l'algorithme HITS et capable de préciser un certain nombre de communautés Web incluses dans le graphe de voisinage de la requête, que peut-on imaginer comme exemple vraisemblable illustrant l'intérêt de mettre en évidence des sous-communautés Web pour la recherche entreprise ?

122, § 1

Les algorithmes d'analyse de réseau sont-ils assez efficaces pour faire qu'existe effectivement un moteur de recherche décomposant le graphe de voisinage d'une requête en sous-communautés en un temps suffisamment court pour être acceptable par l'utilisateur ?

Pour traiter les problèmes d'analyse de réseaux sociaux soulevés par la volonté de décomposer l'ensemble des résultats en sous-ensembles autoréférentiels, à propos de quels spécialistes la demande a-t-elle augmenté le plus ?

Existe-t-il un moteur de recherche actuel qui propose à l'utilisateur des sous-communautés Web relatives à des sujets plus spécifiques (ces sous-communautés étant définies comme des sous-graphes autoréférentiel, et non à partir des sujets plus spécifiques) ?

Que pourrait être un exemple illustrant la manière dont sont présentés les résultats affichés par un moteur de recherche proposant des sous-communautés de pages ?

Un moteur de recherche a pris le nom de *Teoma*. Que signifie le mot *Teoma* ? En quelle langue ?

122, § 3

Tout moteur de recherche utilisant l'algorithme HITS propose comme résultats de recherche des autorités et des *hubs*. Comment se présentent concrètement ces deux classes de résultats ?

Parmi les résultats fournis par un moteur de recherche utilisant l'algorithme HITS, que peut-on s'attendre à trouver comme *hub* à propos d'une requête donnée ?

122, § 4

Lorsqu'un moteur de recherche utilise l'algorithme HITS, quel type d'anomalie visible par l'utilisateur cela peut-il provoquer ? Comment ce type d'anomalie est-il lié à l'usage de HITS ? Comment peut-on envisager de remédier à cela ?

Bibliometrics

123, § 1

Qu'est ce que la bibliométrie ? Quel est le principe clé pour classer les auteurs ? En quoi cela ressemble-t-il aux principes mis en œuvre dans l'algorithme PageRank ? Pourquoi ces ressemblances ?

Qu'est ce que *l'impact factor* ? Que prétend-on classer par ce moyen ? Quelle institution se charge de son calcul ?

En quoi peut-on dire que l'impact factor est fondé sur un nombre de « liens entrants » (*inlinks*) ?

123, § 3

Quelle critique peut-on adresser à l'impact factor quant à sa prise en compte des citations d'une revue par une autre ?

Qu'appelle-t-on la force de connexion (*connection strength*) d'une revue à une autre ? Qu'appelle-t-on *l'influence weight* d'une revue ? Peut-on dire qu'il existe une ressemblance forte entre *l'influence rate* d'une revue et le PageRank d'une page Web ?

123, § 4

Quelles différences peut-on énoncer en termes de *hubs* et d'autorités, entre le World Wide Web et la littérature scientifique ? Qu'est ce qui peut jouer le rôle de *hub* dans la littérature scientifique ? En quoi la définition des autorités est-elle différente dans la littérature scientifique ? Pourquoi peut-on dire que, alors que le Web a une structure à 2 niveaux (les *hubs* et les autorités), la littérature scientifique a une structure à 1 niveau. A quoi peut-on attribuer ces différences ?

À quels exemples peut-on penser d'autorités du Web qui éviteront de se citer l'une l'autre (en envoyant des liens l'une vers l'autre).

Learning to rank

124, § 1

Quels sont les deux « ingrédients » principaux de la grande activité scientifique et économique en matière de moteurs de recherche ?

Peut-on penser que la grande activité actuelle observable dans les sociétés de moteurs de recherche va continuer à se développer ou va se stabiliser, ou encore va laisser place à d'autres développements jusqu'ici inconnus ?

124, § 2

En quoi les techniques d'apprentissage machine (*machine learning*) et les algorithmes d'apprentissage sont-ils utiles pour fabriquer un algorithme classant les pages Web selon leur pertinence vis-à-vis d'une requête donnée ?

Qu'appelle-t-on données d'apprentissage (*training data*) et en quoi consistent-elles en ce qui concerne l'apprentissage machine mis en œuvre pour fabriquer un algorithme de classement selon la pertinence ?

124, § 3

Qu'attend-on que soient capables de faire les algorithmes de classement fabriqués à partir de certaines données d'apprentissage par la technique de l'apprentissage machine ?

Étant donné un ensemble de caractères (*features*) d'un document et des indicateurs numériques relatifs à ces caractères, un algorithme de classement peut assigner pour valeur numérique de rang à une page donnée une combinaison linéaire de ces indicateurs, dont les coefficients peuvent être regardés comme des poids (*weights*). Comment ces poids sont-ils déterminés *a priori* ? Le jugement humain intervient-il dans cette détermination ? Comment ?

124, § 4

Dans la définition d'un algorithme de classement le système de poids utilisé pour combiner les différents indicateurs numériques relatifs aux caractères pris en compte sur la page (titre, ancre, etc.) dépend-il de la requête ? Pourquoi ?

Quels sont les principaux caractères d'une page Web pris en compte par un algorithme de classement selon la pertinence, dans le cas d'une requête comportant un seul mot clé ? Plusieurs mots clés ?

Combien de caractères d'une page Web un algorithme de classement peut-il prendre en compte ?

Le calcul de la valeur de ces caractères est-il réalisé chaque fois qu'une requête est présentée au moteur de recherche ? Pourquoi ? Comment obtient-on le rang de la page à partir de ces valeurs ?

125, § 1

Lors de la mise au point (ou d'une mise à jour) d'un algorithme de classement selon la pertinence, quel est le critère selon lequel doivent être fixés les poids accordés aux différents caractères pris en compte ? Comment cela est-il réalisé ?

Comment s'appelle l'algorithme de classement du moteur de recherche MSN (de Microsoft) ?

125, § 2

Comment intervient le jugement humain dans la détermination des poids accordés aux différents caractères pris en compte pour mettre au point un algorithme de classement comme celui du moteur de recherche MSN de Microsoft ?

125, § 3

Quel est la partie astucieuse de la procédure de détermination des poids mise en œuvre à propos de l'algorithme de classement du moteur de recherche MSN de Microsoft ?

126, § 1

Quel est l'avantage des techniques d'apprentissage machine, quant à l'exploitation de nouveaux jugements humains relatifs à la pertinence d'un document vis-à-vis d'une requête ?

126, § 2

Comment les gestionnaires d'un moteur de recherche peuvent-ils utiliser le comportement des usagers pour se procurer de nouvelles données d'apprentissage machine afin d'améliorer le système de poids utilisé ?

Quelles propriétés des requêtes des usagers donnent-elles de l'importance, à l'observation de leur *clickstream* pour améliorer l'algorithme de classement selon la pertinence ?

126, § 3

Comment un moteur de recherche peut-il connaître les clics de ses usagers ?

Dans la connaissance qu'un moteur de recherche peut avoir des comportements des usagers, le navigateur utilisé peut-il aussi jouer un rôle ?

Certains moteurs de recherche proposent à l'utilisateur d'installer des additifs (*add-on*). Certains d'entre eux permettent aux moteurs de recherche de mieux observer les comportements de leurs usagers. Quels exemples peut-on donner de tels *add-ons* ?

De quels arguments un usager peut-il tenir compte pour accepter qu'un moteur de recherche retienne des informations concernant son comportement sur le Web ?

126, § 4

Les moteurs de recherche ont utilisé d'abord le contenu des pages (recherche en texte intégral, *full text search*) ; puis est arrivée l'analyse des liens (que ce soit à l'aide de l'algorithme PageRank ou de l'algorithme HITS). Quel est le type d'information qui a pris une place croissante dans les algorithmes des moteurs de recherche au cours de ces dernières années ? En quoi peut-on penser que ce type d'information permet au moteur de recherche de mieux répondre aux attentes des usagers ?

Distributing the Index

126, § 5

Quels problèmes de taille la recherche en texte intégral que pratiquent les moteurs de recherche les plus importants rencontre-t-elle ? De quelle façon sont liés la rapidité de traitement et l'espace occupé en mémoire ?

127, § 1

En quoi la mise au point d'un moteur de recherche compétitif suppose-t-il *et* des mathématiques avancées, *et* une capacité importante de conduire des expérimentations souvent laborieuses jusqu'à leur terme ?

127, § 2

Comment un moteur de recherche utilise-t-il la structure de l'Internet pour garantir sa réactivité et sa fiabilité vis-à-vis de requêtes des usagers ?

Que peut-on dire des différents sites de serveurs d'un moteur de recherche du point de vue de leur distribution dans l'espace ?

127, § 3

Comment est distribué l'index et le contenu du cache d'un moteur de recherche sur l'un de ses sites de serveur ? Combien de serveurs un tel site peut-il comporter ?

Un site de serveurs d'un moteur de recherche est-il constitué de superordinateurs ? Sinon comporte-t-il des micro-ordinateurs de bureau ? Comment ces matériels sont-ils apprêtés et disposés dans les locaux où ils sont stockés ?

127, § 4

Les sociétés de moteurs de recherche utilisent-elles plutôt des logiciels libres (*open source*) ? Si l'on excepte Microsoft, quel est le système d'exploitation le plus fréquemment utilisé par ces sociétés ?

127, § 5

Quand on considère les risques de panne ou de détérioration, pourquoi est-il judicieux pour une société gérant un moteur de recherche de se procurer des matériels à vil prix ?

Developments in Web Search

128, § 1

Que désignent les sigles HTML et XML ?

Quelles sont quelques-unes de principales évolutions qui peuvent affecter la recherche d'information sur Internet à l'avenir ?

Qu'est ce qu'un blog ?

Qu'est ce qu'AJAX ? Quel intérêt présente-t-il ?

Comment d'écrire de façon concise ce qu'est ou ce que serait le Web sémantique (*semantic Web*) ?

Searching Blogs

128, § 2

Qu'appelle-t-on la blogosphère ?

Chercher de l'information dans la blogosphère, est-ce semblable à chercher de l'information sur le Web ?

Quand on cherche de l'information sur un blog, peut-on gérer son accès à l'information en fonction de la « fraîcheur » de celle-ci ?

128, § 3

À un billet publié par l'auteur dans son blog sont souvent attachés, par l'auteur du billet lui-même, des métadonnées. Que décrivent-elles ?

On trouve souvent dans un, à la suite d'un billet publié par l'auteur, la mention *trackbacks*. De quoi s'agit-il ?

On trouve souvent dans un blog, à la suite d'un billet publié par l'auteur, la mention *blogroll*. De quoi s'agit-il ?

En quoi les métadonnées figurant dans les étiquettes (tags) ou dans les *trackbacks* attachés à un billet, ou encore dans le *blogroll* de l'auteur du blog, peuvent-elles être utiles pour la recherche d'information ?

Du point de vue de la recherche d'information, quel est le principal problème qu'engendre la facilité de mettre des annotations (tags) sur un blog ?

Qu'appelle-t-on folksonomie (*folksonomy*) ?

AJAX technology

129, § 1

De quelle expression le mot AJAX (en majuscules) est-il l'acronyme ?

Quel est l'intérêt essentiel de la programmation de pages Web avec AJAX ?

Lorsqu'une page Web est programmée avec AJAX, que fait le serveur Web lorsque l'utilisateur clique sur un lien interne à cette page Web pour obtenir l'affichage d'une certaine information ? Le navigateur utilisé joue-t-il un rôle dans cette opération ?

Quelle impression peut avoir l'utilisateur lorsqu'il clique sur des liens internes à une page Web programmée avec AJAX ? Quelle différence y a-t-il par rapport à des pages Web « ordinaires » ?

129, § 2

Qu'est-ce que le système GMail ?

Quelle évolution du Web en général, du point de vue de l'utilisateur, peut-on attendre de l'utilisation croissante d'AJAX ?

La généralisation de l'emploi d'AJAX va généraliser du même coup l'envoi par les serveurs Web aux navigateurs, non pas de pages Web entières, mais de petits bouts (*nuggets*) de ces pages. Que penser de l'éventualité que les moteurs de recherche indexent ces « petits bouts » individuellement, sans référence à la page Web dans laquelle ils sont prévus pour apparaître ?

The Semantic Web

129, § 3

Une limitation importante des moteurs de recherche est qu'ils ne « comprennent » pas les usagers humains. Compréhendent-ils le contenu des pages Web ?

129, § 4

Vous avez un jour goûté, dans le lointain Tyrol, à un délicieux dessert dont le nom vous échappe et dont vous voudriez avoir la recette. Quelle différence y a-t-il (ou y avait-il), si vous cherchez à retrouver ce nom, entre interroger une personne originaire du Tyrol et interroger un moteur de recherche ?

129, § 5

Dans quelle mesure un moteur de recherche est-il capable de faire une inférence à la place et au profit d'un usager humain ?

Qu'appelle-t-on un service Web (*Web service*) ?

Qu'est ce qu'un agent ? Un agent peut-il se connecter à un moteur de recherche ? Dans quel but ?

130, § 1

Dans le projet de Web sémantique, l'idée est d'enrichir les pages Web pour permettre aux machines de coopérer. Quel est le but assigné à cette coopération ?

Pour quelles raisons est-il difficile pour un ordinateur d'opérer des inférences telles que les réalisent les humains ?

Dans le projet de Web sémantique, on souhaite que des machines puissent traiter le sens. Quelle modification de l'information déposée sur le Web est envisagée pour cela ?

Le Web actuel permet aux humains de communiquer. Dans le Web sémantique on souhaite permettre aux machines de communiquer entre elles. Dans quel but ?

130, § 2

Qu'appelle-t-on URI ? Cela concerne-t-il seulement les pages Web ?

Qu'appelle-t-on *Resource Description Framework* (RDF) ? Comment est-il possible de formuler des énoncés portant sur des ressources du Web ?

Considérons l'énoncé suivant, structuré sur trois lignes successives :

`<http://www.everwonder.com/david/felixthecat/>`

`<http://www.love.example.org/terms/reallyLikes>`

`<http://fishhead.homestead.com/>`

De quelle façon peut-on regarder cet énoncé comme signifiant que Félix aime le poisson ?

130, § 3

Un énoncé RDF a-t-il *a priori*, grâce à sa structure, plus de sens pour une machine que du texte sur une page Web ?

131, § 1

Étant donné une communauté d'agents programmés (sur le Web), qu'appelle-t-on une ontologie (*ontology*) ? Quelle est le but assigné à une ontologie ?

131, § 2

Qui a conçu l'idée du Web sémantique ?

La difficulté de créer des énoncés RDF et des ontologies peut-elle être regardée comme un obstacle actuel au développement du Web sémantique ?

Le développement du Web sémantique est-il susceptible de changer fortement la face du Web ?

131, § 3

Dans la vision d'un Web sémantique, ils s'ajoutent aux documents classiques du Web des fragments de données à propos de ces documents. Quel est le but de cette adjonction ?

En quoi peut-on dire que, s'il parvient à maturité, le Web sémantique permettra de nouveaux partenariats entre hommes et machines ? Peut-on dire alors que le Web et la vraie vie se rapprocheront l'un de l'autre ?

Birth of the Dragons

131, § 4

Qui a utilisé d'abord l'Internet au début (à partir de 1973) ? Sous quelles conditions était-il possible alors de se connecter à un ordinateur distant ? Était-ce difficile ? Pourquoi ?

131, § 5

Quand est apparu le protocole de transfert de fichiers FTP (*File Transfer Protocol*) ? Qu'appelle-t-on FTP anonyme ? Que permettait-il vis-à-vis d'un ordinateur distant ?

132, § 1

Comment l'utilisateur intéressé pouvait-il – avant l'apparition des outils de recherche – être informé de l'existence des fichiers proposés sur l'Internet par les serveurs FTP ?

L'emploi du protocole de transfert de fichiers FTP était-il facile ?

The Womb Is Prepared

132, § 2

Quand et par qui a été écrit le premier logiciel permettant de construire un annuaire des fichiers disponibles sur les sites FTP ? Comment s'appelait ce logiciel ? Pourquoi ? Qu'était-il prévu qu'il fasse ? En quoi l'existence de l'annuaire changeait-elle la vie des internautes d'alors ? Son utilisation était-elle facile ?

132, § 3

Quand est apparu le logiciel Gopher ? Qu'était-il censé faire à l'origine ? Pourquoi porte-t-il ce nom ?

Comment se présentait à l'utilisateur l'utilisation de Gopher ? Quelles différences importantes présentait-il par rapport au logiciel Archie ?

132, § 4

Le logiciel Gopher a d'abord été distribué gratuitement avant de devenir payant en 1993. Cela a-t-il eu des conséquences sur son destin ?

Quel est le statut actuel du logiciel Gopher ?

133, § 1

Quelle était la différence quant à la constitution de leur annuaire de fichiers (ou de dossiers) entre les logiciels Archie et Gopher ?

Comment s'appelait le logiciel permettant de collecter sur les serveurs Gopher les menus afin d'intégrer leurs liens dans l'annuaire des sites Gopher ? Pourquoi ? Où le logiciel a-t-il été créé ?

L'annuaire des sites Gopher était-il facilement accessible ? Pourquoi ? Avec quelles conséquences ?

Quelles étaient la nature et la fonction du logiciel créé sous le nom de Jughead ? Pourquoi ce nom ?

133, § 2

Qui a créé, et dans quel cadre, le World Wide Web ? Quelle distinction remarquable a reçu son créateur ?

Quand le World Wide Web a-t-il été annoncé publiquement ? En quoi le World Wide Web différait-il de Gopher ?

Quel rapport y avait-il, à sa création, entre le World Wide Web et les travaux sur la notion de l'hypertexte ?

133, § 3

Pourquoi le démarrage du World Wide Web fut-il lent ? À quoi attribue-t-on son décollage à partir du milieu des années 1990.

The Dragons Hatch

133, § 4

En 1993 est apparu le premier *crawler* du Web sous le nom de *World Wide Web Wanderer*. Quelle était sa fonction ? Quels problèmes a-t-il très rapidement provoqués ?

Pour résoudre les problèmes engendrés par le *World Wide Web Wanderer* a été lancé le logiciel Aliweb. Quel était l'origine de ce nom ? Que permettait de faire Aliweb ? Le permettait-il encore ?

Qu'était JumpStation ? Comment s'appelait le crawler correspondant ? En quoi peut-il être dit un pionnier ? De quelles faiblesses souffrait-il ?

134, § 1

Quand a commencé l'indexation en texte intégral (*full-text indexing*) de pages du Web ?

Quel a été le premier logiciel à indexer intégralement des pages du Web ?

Quelle innovation proposait le moteur de recherche Lycos ? Était-ce une production universitaire ou un produit commercial ? À la fin de l'année 1996, combien de pages environ Lycos indexait-il ?

Quels autres moteurs de recherche furent lancés dans la période 1994-1997 ?

134, § 2

Le moteur de recherche OpenText a disparu très vite. Pour quelles raisons ? Quel enseignement a-t-on pu tirer de son échec ?

Est-il vrai que des maisons d'édition paient la librairie en ligne Amazon pour que leurs livres figurent en tête de liste des réponses aux requêtes appropriées ?

134, § 3

Contrairement aux premiers outils de recherche, le moteur de recherche AltaVista n'a pas été le fruit d'un travail universitaire. Qui en était à l'origine ? Pourquoi ?

Pourquoi très vite, à partir de sa mise à disposition du public, AltaVista est devenu le leader des moteurs de recherche ? Combien de pages indexait-il à son lancement ?

Quelle a été l'histoire d'AltaVista de sa création à nos jours ?

Sous quel nom est connu le service de traduction automatique qu'AltaVista a proposé en 2002 ?

135, § 1

En 1998 Google innove radicalement par rapport aux moteurs de recherche antérieurs du point de vue des informations indexées à propos d'une page Web. En quoi cette innovation a-t-elle consisté ?

The Big Five

135, § 2

À quelle époque les sociétés de moteur de recherche devinrent-elles les vedettes du marché de l'investissement et de la bourse ?

Que sont devenues les sociétés proposant des moteurs de recherche au public et qui abandonné cette activité ? Quand elles n'ont pas disparu, qu'est devenue leur activité ? Quelles étaient en 2005 les cinq plus importantes sociétés de moteur de recherche aux États-Unis (du point de vue du nombre de requêtes) ?

En 2005, quel était le moteur de recherche le plus utilisé aux États-Unis ? Dans le monde ? Ce moteur recevait-il moins ou plus de 50 % des requêtes ?

135, § 3

Quelle est l'origine du nom *Google* ?

Est-il vrai que, selon une tradition légendaire, les créateurs de Google aient commencé à construire leur moteur de recherche dans un garage ?

La réussite technologique de Google fut-elle rapide ? Quels sont les aspects secondaires qui ont, semble-t-il, fidélisé les usagers ?

Google fut le premier moteur de recherche à indexer un milliard de pages Web. Quand ce chiffre a-t-il été atteint ?

Sur la page de recherche de Google, on trouve un bouton "Search" (« Rechercher »). On trouve aussi un second bouton de recherche. De quel bouton s'agit-il ?

Quand Google a-t-il fait son entrée en bourse ? Cette opération fut-elle entourée d'une importante publicité ou passa-t-elle inaperçue ?

Dès 2004 Google fournissait leur service de recherche à d'autres sociétés. Lesquelles ?

136, § 1

Comme Larry Page et Sergey Brin pour Google, les créateurs de Yahoo! furent deux étudiants en thèse de l'université de Stanford. Quel est leur nom ? Sont-ils toujours en thèse à Stanford ? Qu'en est-il à cet égard de Page et Brin ?

Quel a été le point de départ, en 1994, de ce qui est devenu l'annuaire Yahoo! ?

Que signifie le nom Yahoo! ?

Yahoo! est devenu une société anonyme (*incorporated*) en 1995. Quand est-elle entrée en bourse ?

Combien les « documentalistes » (*Website reviewers*) travaillant pour Yahoo! visitaient-ils de sites par jour vers 1996 ?

L'annuaire Yahoo! possédait-il dès le départ un moteur de recherche interne ?

Quelles furent les principales acquisitions de Yahoo! en 2002 et 2003 dans le monde des outils de recherche ?

Lorsque Yahoo! décida d'adjoindre aux résultats de son annuaire des résultats du Web non contrôlés par ses « documentalistes », vers quel moteur de recherche se tourna-t-il ? Quand lança-t-il son propre moteur de recherche sur le Web ? Avec quelle technologie ?

Yahoo! est-il aujourd'hui plutôt un annuaire ou plutôt un moteur de recherche ?

136, § 2

Quel moteur de recherche utilisait Microsoft les services au début ?

Qu'est ce qu'Inktomi ? Quelle a été son histoire au cours des années 2000 ?

Quand Microsoft a-t-il lancé son moteur de recherche *MSN Search* ? À quels types d'objets ce moteur étendait-il ses recherches ? Permettait-il de faire des recherches sur l'ordinateur local ? Comment ?

Le service MSN Search de Microsoft comporte-t-il de la publicité ?

Qu'est-ce que l'encyclopédie *Encarta* ? Quel rapport y a-t-il entre Microsoft, MSN Search et Encarta ?

136, § 3

Qu'est-ce qu'AOL ?

Qu'est-ce qu'Excite ? En 1997, la société Excite a acquis le premier moteur de recherche en texte intégral ; comment s'appelait ce moteur ?

AOL a d'abord utilisé Excite pour son service de recherche NetFind. En 2002, AOL a commencé à utiliser les services d'un autre moteur de recherche ; lequel ?

Quelles différences y a-t-il entre les pages de résultats affichées d'une part par AOL Search, d'autre part par Google ? Quelles différences y a-t-il entre les services de recherche fournis par AOL d'une part, et Google d'autre part ?

136, § 4

Le moteur de recherche Ask a été lancé en 1997 sous le nom de Ask Jeeves. De qui Jeeves était-il le nom ? Quelle était alors l'ambition de Ask Jeeves du point de vue de l'interaction avec les usagers ?

Après son entrée en bourse en 1999, que devint la société Ask Jeeves ?

En 2001, la société Ask Jeeves acquit un moteur de recherche et proposa des services de recherche classiques (par mots clés). Quel était ce moteur de recherche ?

En 2005 Ask Jeeves a été acheté par un grand groupe de médias, InterActiveCorp, qui a supprimé "Jeeves" de son nom et a rebaptisé la technologie de recherche utilisée. Quel nom lui a-t-il donné ?

137, § 1

Qu'est-ce qu'une méta-recherche ? Qu'est-ce qu'un métamoteur ? Quels exemples peut-on citer de métamoteurs ?

Les moteurs de recherche les plus performants essayent sans cesse d'améliorer leurs algorithmes. Quel est l'enjeu économique de cette course-poursuite ?

137, § 2

Comment est-il possible de comparer de manière non biaisée la qualité des résultats fournis par divers moteurs de recherche, pour une requête donnée ? La pertinence étant estimée par un score de 1 à 5 quelle est l'allure de la distribution des scores de pertinence pour les résultats d'un moteur donné ? Au début de l'année 2006, quel moteur avait les meilleurs scores parmi Google, Yahoo!, Ask Jeeves, MSN Search ?

137, § 3

Peut-on penser que la structure économique du monde des moteurs de recherche subira une évolution importante dans les années à venir ? Pour quelles raisons ? Qu'en est-il de l'évolution des technologies de recherche ?

Inside the Dragon's Lair

137, § 4

L'*Encyclopedia Britannica* (2004) contient environ 44 millions de mots. Vers le milieu de l'année 2006, le moteur de recherche Google traitait chaque jour des requêtes représentant en tout un certain nombre de mots. Quel est approximativement le rapport de ce nombre au nombre mots contenus dans l'*Encyclopedia Britannica* ?

138, § 1

Il y a quelques années, Google annonçait que sur une période d'un mois il recevait environ 200 000 requêtes constituées par le nom de la chanteuse Britney Spears. Il indiquait par ailleurs que, sur une période de trois mois, il s'était vu adresser 488 941 fois la requête **britney spears**. Pourquoi ce nombre n'est-il pas plus proche des 600 000 attendus ?

139, § 1

Derrière leur activité visible (répondre aux requêtes des usagers), les moteurs de recherche mémorisent toutes les requêtes qui leur sont adressées par chacun des usagers. Pourquoi ces données peuvent-elles être regardées comme constituant un trésor ?

140, § 1

Google offre un service appelé *Zeitgeist*. Que signifie ce mot ? Quel est l'objet de ce service ?

Quelle fut la requête la plus souvent soumise à Google durant l'année 2005 ? Peut-on disposer de données analogues pour un mois ou une semaine ?

140, § 2

Que propose le service Google Trends ?

Sur son site, Google précise que dans les informations qu'il donne (par catégories) à propos des mots clés les plus utilisés ou des tendances de tel ou tel mot clé, ne figure aucune information relative à tel usager particulier. Cela signifie-t-il que Google ne communique pas de telles informations ou simplement qu'il n'en dispose pas lui-même ?

140, § 3

Comment se répartissaient les requêtes adressées à Google, vers 2006-2007, entre requêtes en anglais, d'une part, et requêtes dans une autre langue, d'autre part ?

Les services Google *Zeitgeist* et Google Trends existent-ils pour d'autres pays du monde que les États-Unis ? Peut-on savoir par exemple quel était le mot clé le plus fréquemment utilisé en Allemagne en mai 2006 ?

140, § 4

Les requêtes présentées à Google dans le monde entier donnent-elles des informations sur les événements du monde ? Quels exemples d'ampleur peut-on citer pour les années 2001-2004 ?

141, § 1

Un jour, aux États-Unis, des moteurs de recherche ont reçu en très grand nombre la requête suivante : **carol brady maiden name**. Ce pic de demandes pour la requête en question s'est reproduit au cours des heures qui ont suivi, en provenance de régions du pays qui changeaient en fonction de leur fuseau horaire. Comment peut-on expliquer ce type de phénomène ?

En février 2001 Google a reçu un grand très grand nombre de fois l'étrange requête suivante : **all your base are belong to us**. Comment cela peut-il s'expliquer ?

141, § 2

Quelle différence peut-on voir entre les informations accumulées par les moteurs de recherche à propos des requêtes qui leur sont adressées et les informations publiées dans la presse et les ouvrages savants ? Sur quels aspects du fonctionnement de nos sociétés les moteurs de recherche ont-ils des informations uniques ?

141, § 3

Sait-on ce que font les sociétés de moteurs de recherche du trésor d'informations qu'elles accumulent à travers l'enregistrement des requêtes ?

141, § 4

On peut penser que la connaissance des requêtes les plus fréquentes permet à certains moteurs de recherche de faire d'importantes économies pour répondre à ces requêtes. De quelle façon ? Pourquoi cela est-il particulièrement important pour les moteurs utilisant l'algorithme HITS ?

En 2001, certains internautes jouèrent à un jeu malicieux, le "Google-whacking", consistant à trouver des requêtes faites de deux mots qui ne soient contenus que dans une unique page du Web. Pourquoi cela fut-il préjudiciable à la société Google ? De quel point de vue ?

141, § 5

Lorsqu'on leur soumet une requête, certains moteurs de recherche proposent, en même temps que les résultats qu'ils ont trouvés, une requête « corrigée » sur laquelle l'utilisateur peut ou non cliquer. En quoi cela permet-il aux gestionnaires du moteur d'améliorer le système de correction automatique des requêtes ?

141, § 6

Qu'appelle-t-on *text mining* (« fouille ou forage de textes ») ?

Comment à partir d'un ensemble d'expressions (telles que des requêtes) peut-on extraire des associations de mots (qui se retrouvent souvent ensemble) ?

Quels problèmes pose l'extraction d'information des enregistrements de requêtes possédés par les moteurs ? À quels niveaux ?

ANNEXE 5

Des questions de RII extraites de Notess, 2006

1. Comment ouvrir un lien (présent sur une page Web) dans une nouvelle fenêtre ?
2. Comment faire une recherche portant spécifiquement sur les groupes de discussion (*newsgroups*) ?
3. Qu'est-ce que le « Web invisible » et comment y effectuer des recherches ?
4. On dit que, lorsqu'on lui présente une requête, un moteur de recherche comme Google ne recherche pas dans tout le Web les pages qu'il va afficher ; si c'est vrai, que fait-il alors ?
5. Pourquoi arrive-t-il que, lorsqu'on présente à un moteur de recherche comme Google une même requête, au même instant, mais à partir de deux ordinateurs différents (situés dans une même salle), puisse afficher des résultats non identiques ?
6. Lorsqu'on regarde une vidéo en *streaming* (flux continu), on dit qu'il n'y aurait pas téléchargement de la vidéo sur l'ordinateur utilisé. Mais alors que se passe-t-il ? Et comment conserver la vidéo si on le souhaite ?
7. On dit que, bien avant le Web, existait Usenet, avec ses groupes de discussion. Le contenu des discussions de ces groupes est-il encore consultable aujourd'hui ? Si oui, comment ?
8. Comment effectuer une recherche dans les groupes de discussion de Yahoo! ou de Google ?
9. Existe-t-il des moteurs de recherche qui soient spécialisés dans la recherche d'images ?
10. Existe-t-il des bases de données publiques en matière de sciences de l'éducation ? Comment y accéder ?
11. Existe-t-il sur le Web des données consultables mais auxquelles les moteurs de recherche ne permettent pas d'avoir accès ? Pourquoi ?

12. Quand, en réponse à une requête donnée, un moteur de recherche propose un lien conduisant à une page d'un site Web, comment trouver les autres pages de ce site (s'il en existe) qui, selon ce moteur, répondraient à cette même requête ?
13. Qu'est-ce que « le cache de Google » et à quoi cela peut-il être utile ?
14. Peut-on demander à un moteur de recherche de n'afficher que des résultats satisfaisant a priori certains critères (par exemple que des fichiers .doc ou .pdf, ou que des documents en telle langue ou encore figurant sur tel ou tel site, etc. ? Comment le faire ?
15. Quand on connaît l'URL d'une page Web que l'on veut visiter, est-on obligé de la « taper » dans la barre d'adresse du navigateur que l'on utilise ?
16. Quand, par erreur, on tape une requête dans la barre d'adresse d'un navigateur, on obtient souvent des résultats comme s'il s'agissait d'un moteur de recherche (et non d'un navigateur) ? Pourquoi ?
17. Avec un moteur de recherche, peut-on chercher les pages Web qui contiennent tel mot mais pas tel autre ?
18. Qu'est-ce qu'un navigateur ? Est-ce cela qu'on appelle en anglais un *browser* ?
19. Comment disposer d'un navigateur sur son ordinateur personnel ?
20. Qu'est-ce qu'un moteur de recherche ? Et un « métamoteur » ?
21. Le moteur de recherche Google présente deux boutons, « Recherche Google » et « J'ai de la chance ». Si l'on clique sur le bouton « J'ai de la chance », on arrive directement sur une page Web. Comment cette page est-elle choisie ?
22. Il arrive que, lorsqu'on tape l'URL d'un site Web dans la barre d'adresse d'un navigateur, on arrive parfois sur un autre site. Pourquoi cela ? Peut-on éviter cela ?
23. Comment peut-on copier tout ou partie du contenu d'une page Web apparaissant à l'écran afin de coller une telle image dans un fichier .doc ou .ppt ?
24. Qu'est-ce qu'un annuaire ? Quand est-ce utile ?
25. Peut-on faire porter une recherche de façon spécifique sur les forums de discussion ? Comment ? Avec quels outils de recherche ?
26. À propos du Web et de la recherche d'information, on rencontre souvent des termes en anglais. Existe-t-il pour ces termes une traduction « officielle » en français et, si oui, où peut-on la trouver ?

27. Pourquoi une adresse Web commence-t-elle toujours par http:// ?
28. Pourquoi dans une adresse Web trouve-t-on généralement, après http://, les « trois w » (www), et parfois non ?
29. Qu'appelle-t-on « nom de domaine » ? En quoi cela intervient-il dans les adresses Web usuelles ?
30. Quand on utilise un moteur de recherche, y a-t-il autre chose à savoir au-delà du fait de taper des mots clés et de lancer la recherche ?
31. Quand on s'intéresse à un certain domaine, par exemple la généalogie, peut-on n'utiliser que des outils de recherche « généralistes » ou existe-t-il des outils spécialisés ? Comment les connaître et en disposer ?
32. Les fils RSS sont-ils utiles quand on s'intéresse à un certain domaine ou un certain thème ? Comment en tirer profit ?
33. Quand on formule une requête dans la barre de recherche d'un moteur, doit-on se limiter à deux ou trois mots clés ? Pourquoi ?
34. Avec un moteur de recherche, peut-on rechercher des pages Web qui contiennent tel mot mais pas tel autre ?
35. Avec un moteur de recherche, peut-on rechercher des pages Web qui contiennent exactement telle expression ou même telle phrase ?
36. Existe-t-il des différences importantes entre les moteurs de recherche existants ? Est-il souhaitable d'en utiliser plusieurs lors d'une recherche ? Pourquoi ?
37. Lorsqu'un moteur de recherche propose un résultat, que peut-on tirer de ce qu'il indique sur la page de résultats ? Que signifient les différentes indications qu'il affiche ?
38. On parle parfois d'employer des « opérateurs booléens » lors d'une recherche. De quoi s'agit-il, à quoi cela sert-il et comment s'en sert-on ?
39. Dans un travail universitaire (mémoire, présentation, etc.), comment citer une page Web ?
40. On dit qu'il existe sur le Web des données auxquelles les moteurs de recherche ne peuvent donner accès. De quoi s'agit-il et pourquoi cela ?
41. Qu'est-ce qu'évaluer une page Web ? Et comment le faire ?
42. Qu'est-ce exactement que le Web ? Est-ce différent d'Internet ? En quoi ?

43. On dit qu'un moteur de recherche utilise un « crawler ». Qu'est-ce que c'est ? En quoi cela importe-t-il à l'utilisateur de base ?
44. On dit qu'un moteur de recherche « indexe » les pages du Web. En quoi cela consiste-t-il et à quoi cela sert-il ?
45. Qu'est-ce que « l'achat de mots clés » ? En quoi cela influe-t-il sur les résultats affichés par un moteur de recherche ?
46. Comment identifier une base de données « invisible » ?
47. Comment, au cours d'une recherche, utiliser les possibilités offertes par Windows (onglets, fenêtre...) pour « naviguer » parmi les pages déjà ouvertes ?
48. Lors d'une recherche, comment prépare-t-on (et analyse-t-on) une requête à faire à un moteur de recherche ?
49. Qu'appelle-t-on « recherche avancée » ?
50. Qu'est-ce qu'un portail ? Est-ce utile pour la recherche d'information ?
51. Y a-t-il une différence entre une base de données et un index ? Si oui, laquelle ?
52. Windows est-il un moteur de recherche ? Sinon, quel est le moteur de recherche « associé » ? Est-ce Internet Explorer ?
53. Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ?
54. Parmi les résultats affichés par un moteur de recherche, peut-il se glisser de la publicité ? Si oui, comment la reconnaître ?
55. Comment faire pour utiliser *Google News* sans être chaque fois renvoyé vers *Google Actualités* ?
56. Pourquoi arrive-t-il qu'un moteur de recherche ne propose pas un document on ne peut plus pertinent par rapport à la requête présentée, et dont on sait de façon sûre qu'il a été récemment mis en ligne sur un certain site ?
57. Qu'est-ce qu'un moteur de recherche de site ? Est-ce différent d'un moteur de recherche « général » ? Quels sites en proposent-ils un ? Sinon, comment faire pour effectuer une recherche sur le site auquel on s'intéresse ?
58. Quelle différence y a-t-il entre un « site Web » et une « page Web » ?
59. Un site Web a-t-il une URL ?

60. Quand on demande à un moteur de rechercher les documents ayant un certain mot dans leur titre (*title*), de quel titre s'agit-il ? Celui du document ? Ou d'autre chose ?
61. Quand on cherche des documents contenant des mots d'une certaine famille, on peut utiliser l'opérateur de troncature, habituellement noté par un astérisque (exemple : amour*, math*, éduc*). Mais cela semble ne pas marcher avec Google. Existe-t-il des moteurs de recherche pour lesquels ça marche vraiment ?
62. Pour apprécier la « fiabilité » des contenus d'une page Web, on peut penser à examiner les pages Web qui ont un lien vers cette page. Mais comment connaître ces pages ?
63. Peut-on demander à un moteur de recherche de n'afficher que les résultats correspondants à des documents publiés en telle ou telle période (avant 2001, ou le mois dernier, etc.) ? Si oui, comment le faire ?
64. Quand le navigateur affiche une page Web, peut-on savoir d'où provient cette page, qui en est l'auteur, sur quel site elle se trouve, etc. ? Comment ?
65. On dit que, avant le Web, il existait un réseau international de sites d'information appelé « Gopher ». Ce réseau existe-t-il toujours ? Peut-on encore l'utiliser ? Comment ?
66. Combien existe-t-il de moteurs de recherche ? Une dizaine ? Une centaine ? Plus encore ? Comment les trouver ?
67. Qu'appelle-t-on « liens sponsorisés » ? Est-ce utile ou est-ce nuisible pour la recherche d'information ?

ANNEXE 6

Présentation du questionnaire en ligne

On trouvera reproduit ci-après le texte introductif à l'enquête en ligne dont les résultats sont commentés dans le dernier chapitre de cette partie.

À propos de la recherche d'informations sur Internet

Nous serions heureux que vous répondiez au questionnaire qui suit, sur cette page. Mais avant cela, nous devons vous donner quelques explications simples mais indispensables.

NOUS VOUS DEMANDONS DONC INSTAMMENT DE LIRE CE QUI SUIT AVEC ATTENTION.

Vous trouverez ci-après l'énoncé de vingt questions successives, numérotées de 1 à 20. (L'ordre que nous avons choisi est tout à fait arbitraire.)

Imaginons que vous vous trouviez face à la question suivante (qui ne figure pas parmi les vingt questions : c'est pour cela que nous la prenons pour exemple ici...) : « Quelle est la différence entre une bactérie et un virus ? » Ou encore face à celle-ci : « L'anxiété accroît-elle la capacité d'apprendre ? » Ou encore devant celle-là : « Le nombre 0,99999999... (avec une infinité de 9) est-il égal à 1 ? »

Nous ne vous demandons pas de répondre à ces questions, pas plus que de dire si vous estimeriez pouvoir le faire !

Nous vous demandons en revanche de préciser si, au cours des temps passés, vous avez été *en « contact »* avec une réponse – quelle qu'elle ait été – à la question considérée, par exemple dans le cadre d'un cours, à l'occasion d'une lecture personnelle, d'une conférence ou même tout simplement d'une discussion informelle à laquelle vous auriez assisté, et cela même si vous ne vous souvenez plus du tout du contenu de cette réponse !

Plus exactement, nous vous demandons d'indiquer « l'importance » de ce contact, sur une échelle qui va de 0 (pas de contact du tout dont vous vous souveniez) à 4 (contact approfondi : par exemple vous avez suivi un cours de plusieurs heures ou vous avez lu tout un livre sur la question...).

En outre, quel qu'ait été votre contact avec une réponse à la question considérée, nous vous demandons aussi, chaque fois, d'indiquer votre *désir actuel* de disposer éventuellement d'une réponse « solide » à cette question, toujours sur une échelle de 0 à 4.

Ainsi, sur la question des bactéries et des virus, vous pourriez estimer avoir eu autrefois un contact d'ordre 2 avec une réponse (que vous auriez par exemple oubliée), et éprouver aujourd'hui un vif désir, disons d'ordre 3 au moins, de disposer d'une réponse « sûre ». Etc.

Voilà. À vous de faire ! Et merci...

Caroline Ladage & Yves Chevallard (Université de Provence).

¹ La notation C18 employée par l'auteur désigne le XVIII^e siècle ; la notation mC18, elle, désigne le milieu ou, plus largement, le tiers médian du XVIII^e siècle.

² Sur Internet, la notion de forum (ou groupe de discussion) est ancienne mais, selon l'article "Internet forum" de l'encyclopédie Wikipedia, le mot n'apparaît que vers 1996.

³ À propos de questions de référencement et de moteurs de recherche, nous avons par ailleurs étudié la vie de divers forums Internet (Ladage, 2007b). Notons que ceux-ci conjoignent en principe (quoique inégalement) les trois fonctions de *forum*, d'*atelier* et de *boutique*.

⁴ Cet épisode célèbre se trouve dans les *Confessions* (*Confessionum libri tredecim*), livre VIII, chapitre XII.

⁵ L'URL http://www.googleguide.com/synonym_operator.html conduit en effet directement à la réponse attendue. Le 28 août 2008, elle figurait en troisième position sur la *première page* de résultats affichés par Google en réponse à la requête indiquée.

⁶ Voir <http://www.ihelpyou.com/search-engine-chart.html>.

⁷ Voir <http://www.dmoz.org/help/become.html>.

⁸ Voir <http://www.emploi.com/emploi/metiers/cyberdocumentaliste.html>.

⁹ Pour Google, voir <http://www.google.fr/addurl/?hl=fr&continue=/addurl>. Pour le moteur de recherche Exalead, voir <http://www.exalead.fr/search/submitYourSitePage>. (Pour Yahoo! la page proposée concerne son annuaire : <http://fr.docs.yahoo.com/info/ajouter.html>.)

¹⁰ L'obstacle, déjà signalé (dans la deuxième partie de ce travail), tient au fait que les divers moteurs consultés par le métamoteur n'ont pas en général le même langage booléen. On notera ce paradoxe : la pertinence pratique du lieu commun évoqué suppose un usage extensif des requêtes booléennes ; or celles-ci, nous l'avons vu, constituent une frontière praxéologique peu fréquentée.

¹¹ Voir http://www.adbs.fr/outils-de-recherche-web-828.htm?RH=1214314880901&RF=OUTILS_RECHWEB.

¹² Les adresses de Looksmart et de Snap sont <http://www.looksmart.com/> et <http://www.snap.com/>.

¹³ L'interface de recherche avancée de Google semble indiquer que la « date » dont il est question serait la date « de découverte » (par Googlebot ?) de la page Web.

¹⁴ Notons ici que le substantif *popularity* et l'adjectif *popular* sont dépourvus en anglais de la connotation péjorative qui peut affecter, en français, l'emploi des mots « popularité » et « populaire ». L'écart est sans doute maximal sur le mot *populace*, qui appartient à l'anglais comme au français, mais qui, en anglais, n'a rien de la valeur fortement péjorative seule connue en français.

¹⁵ C'est l'unique note de bas de page de l'exposé que nous examinons. L'ouvrage, de quelque 80 pages, ne comporte en tout que trois notes infrapaginales : les deux autres concernent respectivement le caractère @ (p. 16) et la notion de *smiley* (p. 23).

¹⁶ On ne trouve de commentaire de Google touchant les liens commerciaux situés au-dessus des résultats naturels que dans la version anglophone de son site : voir <http://www.google.com/ads/overview.html>.

¹⁷ L'assertion « ... pour annoncer à l'internaute qu'il se trouve sur le World Wide Web » est doublement discutable. Tout d'abord, la formulation « ... qu'il a affaire à un serveur (de pages) Web » serait plus juste. Ensuite et surtout, l'intention naïvement formulée (« pour annoncer à l'internaute ») personnalise une habitude indurée, certes, mais faiblement motivée : on a vu dans la troisième partie de ce travail que la présence du préfixe *www* ne correspondait à aucune exigence technique et n'était due qu'à la “long-standing practice of naming internet host (servers) according to the services they provide”.

¹⁸ Le lecteur de ces lignes peut, quel qu'il soit, obtenir aujourd'hui un nom de domaine en .org, .info, .com, etc. Quant à l'affirmation relative au « pays d'origine », elle pourra déconcerter, par son imprécision : qui, ainsi, ne connaît Google.fr par exemple ?

¹⁹ Se référant à l'adresse <http://www.cleml.org/spe/pedagogie/spedp.html>, les auteures écrivent en effet : « **spe** est le répertoire, **pedagogie/spedp** sont les noms d'un fichier et d'un sous-fichier : ces éléments désignent l'emplacement de cette page sur le serveur du Cleml. »

²⁰ Voir <http://www.educnet.education.fr/dossier/rechercher/evaluation1.htm>.

²¹ La section précédente s'achève avec une sous-section intitulée « Faut-il se passer du copier-coller ? » en laquelle on peut voir une timide ébauche du processus qui conduit, dans le schéma herbartien, des réponses R^\diamond à la réponse R^\heartsuit cherchée : nous ne la commenterons pas davantage.

²² La section (et le chapitre) se termine par une sous-section intitulée « Deviendrons-nous tous journalistes ? », que nous ne commenterons pas ici.

²³ L'acronyme IEEE désigne à l'origine l'*Institute of Electrical and Electronic Engineers* ; cette référence originelle ne couvre plus aujourd'hui qu'une partie des champs d'intérêt de l'association, qui est donc désormais désignée simplement par la suite de lettres I-E-E-E (prononcée – en anglais – “Eye-triple-E”). L'acronyme ECCAI désigne l'*European Coordinating Committee for Artificial Intelligence*.

²⁴ Voir <http://www.linkedin.com/in/teresanumerico>. On rencontre ici un autre domaine qui pourrait se rajouter aux treize visités dans la troisième partie de notre travail : la *philosophie* (et l'*éthique*). Nous n'irons pas plus loin sur ce sujet.

²⁵ La référence est ici à Fernand Raynaud (1926-1973) et à son sketch *Le fût du canon* (1968).

²⁶ On pourra penser aux deux exemples d'exposés donnés dans l'*Introduction générale* de notre travail.

²⁷ Le mot de concordance désigne ici ce que le *Dictionnaire culturel en langue française* (Rey et al., 2005) définit ainsi : « Index alphabétique des unités lexicales d'un texte, avec leur contexte. *Éditer une concordance de la Bible. Concordance établie par ordinateur.* » Le même dictionnaire date l'apparition de ce sens du mot à 1564.

²⁸ Ce commentaire bref mais significatif a déjà été cité dans le chapitre 11 de notre travail.

²⁹ La question relative à la requête "university" trouve son origine, semble-t-il, dans la sous-section 5.1 de Page, Brin, Motwani & Winograd, 1999.

³⁰ Cette appréciation laudative se trouve sur la couverture de l'ouvrage de Notess.

³¹ On pourra à cet égard consulter son site Notess.com.

³² Sur ce sujet nous nous permettons de renvoyer à l'article "Screencast" de l'encyclopédie Wikipedia.

³³ Voir <http://www.infopeople.org/workshop/385>.

³⁴ Ptooi est le dieu suprême du pays imaginaire nommé Djelibeybi par l'écrivain anglais Terry Pratchett dans la suite de romans fantastiques *Discworld*.

³⁵ Sur cette notion, voir par exemple Chevallard, 2006.

³⁶ Voir <http://www.forums-abondance.com/recherche/open.php?mess=733>.

³⁷ La référence est ici à la théorie de l'économiste Walt Whitman Rostow (1916-2003).

³⁸ *Net Snippets* était un logiciel de capture du contenu de pages Web permettant de sélectionner la partie d'une page que l'on souhaite archiver. Sa production a été arrêtée officiellement (et en principe définitivement) en mars 2007 : voir <http://www.netsnippets.com/>.

³⁹ Nous reproduisons sans le modifier le contenu des posts.

⁴⁰ Bien que ce forum de discussion trouve place sur le site *Abondance*, la qualité des réponses nous paraît faible, sinon indigente. Notre propre expérience nous conduit à penser que l'on rencontre moins facilement de telles réponses déstructurées, à la rédaction anarchique, sur les forums des sites anglophones d'orientation thématique et de niveau comparables.

⁴¹ La méthode des isotopes radioactifs est due principalement au chimiste hongrois Georges de Hevesy (1885-1966), qui reçut le prix Nobel de chimie en 1943.

⁴² D'après le *Dictionnaire historique de la langue française* (Rey et al., 1993), avant la fin du XII^e siècle, *témoin* désigne notamment une personne qui certifie une chose vue ou entendue ; depuis le XVI^e siècle, l'emploi du mot s'est étendu à une personne simplement spectatrice d'un événement.

⁴³ La présence de la question 14 dans cette liste confirme à nos yeux le fait que, dans la sous-culture de la RII qui prévaut aujourd'hui loin du monde des bibliothèques, l'usage des annuaires a perdu – sans doute à tort – une part notable de son importance.

⁴⁴ Ce tableau doit être lu ainsi : la question 1 (ligne 1 du tableau) a reçu dans le premier groupe examiné des scores dont le total représente 59 % du total maximal possible (ligne 2 du tableau) ; elle a de même reçu dans le groupe de Forcalquier des scores dont le total représente 45 % du total maximal (le même dans les deux groupes) ; la différence entre ces scores, qui est ici de 14 points de pourcentage, est indiquée dans la ligne 3 du tableau.

⁴⁵ Pour le texte de l'enquête en ligne, voir l'annexe 6, à la fin de cette partie de notre travail. Nous avons écarté toutes les réponses incomplètes, en règle générale plus nombreuses en ligne que dans le cadre d'une passation *in situ* sur papier.

⁴⁶ La faiblesse des effectifs (respectivement égaux à 3, 1, 3, 3) du tableau de contingence correspondant ne permet pas d'appliquer le test du χ^2 à l'hypothèse H_0 d'indépendance des variables « Contact » (faible, fort) et « Recherche » (facile, difficile). On peut alors appliquer le test exact de Fisher (*Fisher's exact test*), qui, ici, conduit à *ne pas rejeter* H_0 sur la base des données réunies (voir <http://www.langsrud.com/fisher.htm>), ce qui est parfaitement raisonnable au vu du tableau ! Bien entendu, cela ne signifie nullement que l'on *accepte* l'hypothèse d'indépendance.

⁴⁷ Voir <http://c.asselin.free.fr/french/newsgroups.htm> et <http://www.learnthenet.com/French/html/27srchn.htm> respectivement.

⁴⁸ Voir <http://www.webrankinfo.com/google/cache.php> et Bécalséri, 2007, pp. 73-74. Cette équipe mentionne également la page 144 du livre *Internet pour les nuls* ; mais il s'agit-là d'une référence erronée, où il est question en fait du *cache du navigateur*.

⁴⁹ Voir www.google.fr/support/bin/answer.py?answer=27035&topic=368.

⁵⁰ Voir http://www.uhb.fr/scd/Methodoc_aspects_critiques.htm#liensnaturels.

⁵¹ Voir http://www.ebsi.umontreal.ca/jetrouve/internet/mot_goo2.htm.

⁵² Certains exposés sur la RII ont pu croiser la question 19 sans la remarquer : ainsi en va-t-il, apparemment, dans le cas de Lenormand, 2007, p. 59.

⁵³ Le *Dictionnaire culturel en langue française* (Rey et al., 2005) fait remonter à 1963 l'apparition de cette acception, où *folklorique* est employé pour qualifier ce qui est « pittoresque et amusant, mais dépourvu de sérieux ».

CONCLUSION GÉNÉRALE

VERS UNE DIDACTIQUE DU WEB
RÉSULTATS ET PERSPECTIVES

Vers une didactique du Web

Résultats et perspectives

*Au fond, ce sont plutôt des questions que nous posons [...],
ce sont des objets d'enquêtes que nous proposons
plutôt que nous ne résolvons un problème et ne rendons une réponse définitive.
Il nous suffit pour le moment d'être persuadé que, dans cette direction,
on trouvera de nombreux faits.*

Marcel Mauss, 1923-1924

Le travail qui s'achève ici et auquel nous voudrions apporter une conclusion ouvrant sur les perspectives de recherche qui nous paraissent s'imposer a tenté d'apporter des ébauches de réponse à trois grandes questions : 1) quelles praxéologies peuvent-elles être considérées comme *utiles* à la RII ? 2) quelle est l'offre praxéologique en la matière ? 3) quels sont les obstacles à un accès large à une offre praxéologique adéquate ? La première de ces questions comporte un présupposé, à savoir qu'il y aurait des praxéologies (ou des savoirs) *utiles à la RII*. Objectivement, on peut dire que, aujourd'hui encore, le point de vue du « Il n'y a rien à savoir ! » conserve quelques partisans. Mais c'est là une vision de la RII qui, nous semble-t-il, tend à s'effriter, comme le constate à juste titre Greg Notess dans le passage suivant (2006, p. 6) : “Many Web searchers [...] feel that they can always find what they seek, but a substantial number of people are more humble in evaluating their own skills. These people have questions and want training.” Cette évolution, dont l'étude empirique n'était pas au centre de notre recherche¹, est en revanche constitutive du projet plus vaste dont celle-ci participe, celui de l'étude des conditions de possibilité d'une culture commune, citoyenne, partageable de la RII, qui réponde aux besoins de l'utilisateur ordinaire sans pour autant hypothéquer son accès ultérieur à des cultures de la RII plus « avancées » ou plus spécifiques. Comme nous avons été conduite à évoquer quelques autres situations que d'aucuns prétendent indépassables, nous ferons précéder l'ensemble des développements qui suivent de cette remarque de Nietzsche dans son livre *Le Gai savoir* (1887)².

Lorsqu'on commença en France à combattre les trois unités d'Aristote et par conséquent aussi à les défendre, on put voir de nouveau ce que l'on voit si souvent, mais toujours avec déplaisir : *on inventa des raisons* pour lesquelles pareilles lois devaient subsister, simplement pour ne pas reconnaître que l'on s'était *habitué* à leur contrainte et qu'on ne voulait pas les changer. (§ 29)

Praxéologies utiles à la RII

La première des trois questions précédentes – « quelles praxéologies peuvent-elles être considérées comme utiles à la RII ? » – ne saurait avoir de réponse absolue. Pour en éprouver un tant soit peu la pertinence, encore faut-il pratiquer la RII, ce qui suppose bien sûr l'existence de *questions* auxquelles on veuille apporter des réponses, mais qui suppose aussi, comme nous nous sommes efforcée de le montrer dans notre première partie, un « réflexe » – « aller voir sur le Web » – qui ne s'épuise pas en quelques gestes simplifiés, trop tôt arrêtés soit sur un verdict de réussite, soit sur un verdict d'échec, qui, l'un comme l'autre, en nombre de cas, sont également prématurés, ainsi que le suggère l'épisode autour d'une peinture d'Emil Nolde analysé dans la deuxième partie de notre travail.

Quel que soit le point de départ de l'étude de la question posée, il nous a semblé qu'on était conduit très vite à buter, dans la plupart des exposés relatifs à la RII, sur des indications qui ne portaient pas en elles-mêmes leur propre intelligibilité parce que, si étendues et détaillées soient-elles, elles constituaient une description de la *praxis* de la RII dont tout arrière-plan avait été gommé, et dont une réception lucide supposait donc qu'on rétablisse la profondeur de champ sans laquelle les mots finissent par cacher les choses. Nous en ferons l'expérience une dernière fois ici. Au début de son livre, que nous avons plusieurs fois cité, Ran Hock (2007) est amené à proposer à son lecteur une stratégie de recherche en six points que nous reproduisons ci-après avec quelques coupures (pp. 13-15).

1. Identify your basic ideas (concepts) and *rely on the built-in relevance ranking* provided by search engines...
2. Use simple *narrowing techniques* if your results need narrowing: ...
3. *Examine your first results* and look for, and then use, terms you might not have thought of at first.
4. *If you do not seem to be getting enough relevant items, use the Boolean OR operation* to allow for alternate terms; for example, *electrical OR electricity* would find all items that have either the term *electrical* or the term *electricity*...

5. *Use a combination of Boolean operations* (AND, OR, NOT, or their equivalents) to identify those pages that contain a specific combination of concepts and alternate terms for those concepts (for example, to get all pages that contain either the term *cloth* or the term *fabric* [...]).
6. *Look at what else the finding tools (particularly search engines) can do* to allow you to get as much as you need—and only what you need. Advanced search pages are probably the first place you should look.

L’auteur lui-même se voit obligé de donner un peu de profondeur de champ à sa description. C’est ainsi que, à propos du point 1 de cette stratégie de base, il ajoute ceci.

When you enter terms in the major search engines and many other search sites, only those records (Web pages) that contain all those terms will be retrieved, and the engine will automatically rank the order of output based on various criteria.

Cet arrière-plan en appelle à l’évidence quelques autres, sans lesquels l’usager est contraint de se mouvoir dans un espace étriqué, tant du point de vue de l’intelligence des situations de RII que du point de vue des gestes à y accomplir. Ce type de constats nous a alors conduite à tenter d’esquisser les contours d’une « science de la RII », nécessairement nourrie par de nombreuses disciplines – disciplines que la culture courante tend à séparer, alors que les cultures de la RII (et en particulier ce que nous avons appelé la culture-mère, supposée pouvoir les engendrer toutes) devraient permettre de « faire travailler ensemble ».

L’ébauche d’inventaire des praxéologies utiles a été menée avec des techniques différentes au fil de ce travail : procédant, dans la deuxième partie de ce travail, à un inventaire le plus immédiat possible (inventaire sur l’apport duquel nous reviendrons plus loin), nous avons exploré, dans la troisième partie, le vaste champ des connaissances « allogènes » éventuellement utiles à la RII, en tentant surtout de poser un problème et de nourrir un débat qu’il ne nous appartient pas de trancher, même si nous pouvons y contribuer par un certain nombre de propositions toujours formulées alors comme de simples hypothèses de travail. Alors que, dans la troisième partie, donc, il s’agissait pour nous d’interroger divers domaines de connaissance pour tenter d’y repérer des praxéologies utiles à la RII, dans la quatrième partie nous avons opéré en quelque sorte à l’envers, en questionnant sur leurs tenants et aboutissants des exposés sur la RII organisés par des problématiques différentes – celle des auteurs de *Web Dragons* n’étant à l’évidence pas superposable à celles de Notess ou de Hock³. Tout cela aboutit à un inventaire large, dessinant – souvent en pointillé – les territoires praxéologiques sur lesquels pourraient s’élever une « science de la RII » qu’il

resterait alors à élémenter rigoureusement, pour en faire le socle approprié sur lequel puisse reposer une culture citoyenne de la RII, camp de base pour toute culture possible.

L'offre praxéologique en RII

La culture courante de la RII telle qu'on peut l'observer – par exemple à l'université chez des usagers non spécialisés – nous a semblé relativement pauvre, sorte de négatif dont le panorama d'une « science de la RII » tel que nous l'avons rapidement brossé dans ce travail serait le positif. Pour le dire autrement ⁴, on voit ainsi nombre d'usagers du Web *utiliser en amateur un équipement praxéologique d'amateur* – et, souvent, s'en dire satisfaits. Cela rappelé, cette pauvreté de l'équipement praxéologique de beaucoup d'usagers doit à l'évidence être mise en relation avec *l'offre praxéologique* en la matière, même si ce n'est pas là, bien évidemment, le seul facteur explicatif (nous allons y revenir).

Pour mettre en évidence les contours d'une telle offre, il faut d'abord en dessiner les « pleins », vers lesquels une lecture inventoriante des exposés sur la RII dirige nos regards : c'est cela que nous avons fait dans la deuxième partie de notre travail notamment, et encore dans le premier chapitre de la quatrième partie. Mais il faut aussi en repérer les « manques », en faire apparaître les *silences*, et c'est cela que nous avons tenté de faire dans les deux derniers chapitres de la quatrième partie de notre travail. Dans ce but, nous avons mis en œuvre une « lecture questionnante » permettant d'interroger les exposés sur ce qu'ils sont, volontairement ou non, amenés à *taire*, ou à ne dire que très partiellement, alors même que ce non-dit a toutes les apparences de dissimuler des problèmes praxéologiques pertinents pour la RII.

Mais « problèmes » ne signifie pas « solutions ». Nous avons montré en effet que nombre de questions en puissance laissées ainsi ouvertes par les exposés interrogés ne trouvent en fait que difficilement réponse, aujourd'hui, dans les cultures de la RII les plus communes. De ce point de vue, reprenant une terminologie usitée en TAD, nous avons distingué les *boutiques*, qui proposent concrètement leur offre (un exposé ou un stage de formation sur la RII constitue une telle boutique, par exemple), les *forums*, peu nombreux et souvent chétifs, où des questions peuvent être soulevées et être reçues en tant que *problèmes* à résoudre, enfin les *ateliers*, où, à propos d'une question Q , en usant éventuellement de réponses allogènes R^\diamond , on s'efforce de produire une réponse R^\heartsuit , qui sera peut-être ensuite proposée en tel ou tel type de boutiques.

Pour éprouver l'offre praxéologique, nous avons réuni nous-mêmes (dans le deuxième chapitre de la quatrième partie) un large assortiment de *questions* de RII. Quelques-unes de

ces questions, nous les avons posées en diverses boutiques, pour en jauger l'offre praxéologique. Mais, mobilisant en cela une technique que nous croyons originale, nous avons surtout interrogé des « témoins », d'une part sur leur *rencontre possible* avec des réponses éventuelles, d'autre part sur la *difficulté à rencontrer* de telles réponses lorsqu'on les cherche expressément. Le bilan tiré de telles investigations confirme ce qu'une connaissance clinique informelle permettait de subodorer : une proportion non négligeable de questions – autour de 40 % sur les échantillons que nous avons pu étudier – auxquelles l'utilisateur est en droit (même s'il n'exerce pas souvent son droit) de demander des réponses à la culture de la RII dans laquelle il est plongé, *en fait n'y trouvent pas aisément réponse*.

Un triple problème épistémologique, didactique, éthique

Nous ajouterons d'abord, ici, de brèves considérations sur un aspect que, dans le cadre nécessairement limité de ce travail, nous avons décidé d'écarter pour y revenir en d'autres recherches : il s'agit de la *fabrication* de réponses, de ce qui se passe, donc, dans les *ateliers*. Nous nous arrêterons ici sur un exemple qu'une actualité récente apporte. Soit la question de savoir comment un moteur de recherche classe les pages qu'il affiche en réponse à une requête. Le temps où Google n'utilisait que son algorithme PageRank est aujourd'hui révolu : l'évolution se lit dans les textes mêmes publiés par la firme Google (on se référera, là-dessus, à Sullivan, 2007). Un jour seulement après le lancement officiel, le 27 juillet 2008, du moteur de recherche *Cuil* (dont le nom serait "an old Irish word for knowledge" et se prononcerait approximativement comme le mot anglais *cool*), moteur présenté comme "the biggest search engine on the planet" et en fait comme le rival par excellence de Google, Danny Sullivan (2008) met en question certaines des prétentions des fondateurs de Cuil. Cette société, indique Sullivan, fait notamment la promotion de son moteur de recherche sur le fait "that it ranks pages by content, rather than popularity." Précisant d'abord qu'il est inexact que Google classerait ses résultats *uniquement* selon leur PageRank, Sullivan s'inscrit en faux contre l'affirmation publicitaire de la société Cuil – un "PR pitch", écrit-il – et ne craint pas d'affirmer : "Cuil is indeed using popularity to rank results, as far as I can tell." Comment peut-il affirmer cela ? Voici sa réponse.

For example, in a search for [harry potter], the Harry Potter & The Order Of The Phoenix movie web site comes up first on Cuil. This is out of thousands of possible pages. How on earth can Cuil know just from the content on the page itself that the movie site should be in the top results, especially in a web environment where people can (and will) custom tailor content to mislead search algorithms?

La clé de la petite énigme sur le bon classement du site du film est, toujours selon Sullivan, la suivante : “The answer is link analysis—counting links and effectively seeing who is pointed at the most. The twist is that it is done by measuring the links from pages relevant to what someone search on.” À propos des pages classées, l’auteur ajoute : “Those with many or important links are likely to do better. Since the Harry Potter movie page has a lot of links pointing at it, it comes up higher in the results.”

Tout cela semble raisonnable. Mais on voit ainsi que la construction d’une réponse à la question des critères de classement employés *ne va nullement de soi*, et doit se faire sur une base largement conjecturale, en employant en outre un outillage de raisonnements et d’essais empiriques relativement peu sophistiqué, qui montre *a contrario* la difficulté de développer, en matière de RII, une *dialectique des médias et des milieux* appropriée. Cuil prétendant disposer d’un index trois fois plus volumineux que celui de Google, Sullivan aborde également, dans le même billet, la question de l’influence de la taille de l’index d’un moteur de recherche sur la qualité de ses résultats : nous pourrions faire à cet égard des remarques analogues.

Ce qu’on peut induire de ces exemples, c’est que, aujourd’hui, dans les cultures « profanes » de la RII – pas celle des ingénieurs de Google ou des fondateurs de Cuil –, l’offre praxéologique, dont nous avons dit la faiblesse, voit son développement limité par une configuration épistémologique et sociale dans laquelle, faute notamment d’*espaces publics* de problématisation et de fabrication de solutions, qui puissent accueillir et réguler forums et ateliers, le travail requis pour produire cette offre demeure souvent, et tout à la fois, peu outillé, mal contrôlé et faiblement impulsé.

Pour le dire autrement, et pour prévenir toute méprise quant à la réception des résultats de cette recherche, nous dirons que, là où devrait être posé clairement, explicitement le « problème praxéologique » – quels types de tâches, quelles techniques, quelles technologies, quelles théories ? –, on ne trouve souvent qu’un rabattement sur des « compétences » dont la traduction en termes de praxéologies semble cruellement absente. Dans le B2i par exemple, l’un des cinq « domaines de compétence » a pour titre *S’informer, se documenter*. L’annexe à la circulaire du 7 novembre 2006, publiée au *Bulletin officiel de l’Éducation nationale* du 16 novembre 2006, présente dans les termes suivants les « savoir-faire » que l’élève de collège devra maîtriser⁵.

Je sais rechercher des références de documents à l’aide du logiciel documentaire présent au CDI.

Je sais utiliser les fonctions principales d'un logiciel de navigation sur le web (paramétrage, gestion des favoris, gestion de l'affichage et de l'impression).

Je sais utiliser les fonctions principales d'un outil de recherche sur le web (moteur de recherche, annuaire...).

Je sais relever des éléments me permettant de connaître l'origine de l'information (auteur, date, source...).

Je sais sélectionner des résultats lors d'une recherche (et donner des arguments permettant de justifier mon choix).

On ne peut que constater là une *déproblématisation* remarquable, qui pourrait laisser croire, à nouveau, que le savoir *ne serait pas* problématique, et que le seul problème à résoudre serait celui de l'accès de l'élève à ce supposé savoir. Là encore, tout se passe comme si le savoir à enseigner paraissait *transparent* à ceux qui énoncent de telles prescriptions comme à ceux qui ont en charge d'en encadrer la mise en œuvre par les élèves. Cette fiction scolaire va de pair, apparemment, avec l'absence d'un travail collectif visant à définir de manière précise, à un moment donné, les praxéologies donnant une vie institutionnelle et personnelle aux « types de tâches » vaguement répertoriés dans ce qui précède (comme par exemple le type de tâches « Utiliser les fonctions principales d'un logiciel de navigation sur le web »). Car, en vérité, rien ici n'est explicite : techniques, technologies, théories ne sont que *supposées*. En fait, à force d'être ainsi réputées « bien connues », « évidentes », elles demeurent *introuvables*, en sorte que, du double point de vue *épistémologique* et *didactique*, on ne peut guère débattre de leur pertinence, de leurs qualités et de leurs imperfections, et travailler – collectivement – à les améliorer (mot dont le sens spécifique fait lui-même partie du débat et appelle un travail approprié). Cette situation ouvre alors la voie à un bricolage personnel et institutionnel *forcé* et, sauf exception, à un certain état végétatif de la culture de la RII : on songera ici au tableau que d'Alembert brossait de la culture des artisans parisiens vers le milieu du XVIII^e siècle. Aux problèmes épistémologiques et didactiques que cette gestion engendre s'ajoute, nous semble-t-il, un problème *éthique*, relatif à la responsabilité de la société vis-à-vis tout à la fois des enseignants et des élèves, dont, par exemple, aucun ne peut véritablement savoir, en principe, si tel « apprentissage » prescrit a bien été accompli, puisque l'objet même de cet apprentissage reste indéfiniment incertain. Le régime scolaire imposé aux praxéologies de la RII (et, plus largement, à celles du monde Informatique & Internet), l'état d'« indiscipline » dont il affecte les complexes praxéologiques concernés – la doctrine officielle est au reste qu'il faut se garder de tout ce qui contribuerait à élever « un usage social au rang d'une

discipline scolaire » (Duvauchelle, s. d.) – participent d’un projet de diffusion praxéologique bien particulier, dont, sur la base des résultats que nous avons obtenus, nous pensons qu’il tend non seulement à limiter, mais en fait à *dissoudre* l’offre praxéologique : l’école devient à cet égard une boutique où l’offre se dissimule plutôt qu’elle ne se propose.

Notons en outre que ce paradigme didactique, dont l’effet le plus net est de rendre sinon impossible du moins très incertain ce travail de la société sur elle-même, dont nous avons évoqué le déploiement solidaire entre forums, ateliers et boutiques, n’est pas réservé seulement à la diffusion scolaire des connaissances en matière de RII ⁶. Dans un livre récent, ainsi, un sociologue, Jacques Siracusa, dessine ce qu’il nomme un « modèle négatif d’organisation pédagogique » relatif à l’enseignement universitaire de la sociologie, tableau que nous reproduisons *in extenso* ci-après, même s’il excède en apparence le problème soulevé dans les lignes précédentes (Siracusa, 2008, pp. 62-63).

Dans ce modèle négatif,

chaque enseignant a reçu une formation différente de celle de ses collègues ;

il a sa propre définition de ce que sont les connaissances sociologiques ; le domaine à enseigner est délimité par une étiquette administrative ; chaque collègue a sa propre utilisation de l’étiquette ; enseigner est une affaire personnelle ;

l’enseignant parle chaque semaine et pour un semestre à des étudiants,

puis leur demande un exercice final ;

celui-ci a une forme traditionnelle et exige des compétences sans rapport évident avec celles servant à produire les informations présentées en cours ;

la relation entre le contenu du cours et les manières permettant aux étudiants de le valider est obscure ;

savoir écrire est central, mais ce n’est pas l’objet des cours ;

la correction des copies est négligée par les enseignants comme par les étudiants ;

à la fin du cursus, chaque étudiant doit écrire une sorte de chef-d’œuvre qui nécessite la manipulation de connaissances sociologiques.

Dans ces conditions, le jeune enseignant comme l’étudiant préparant son chef-d’œuvre sont incités à inventer une grande part de ce qu’ils sont censés faire.

Si j’ajoute une dernière propriété : enseignants et étudiants privilégient une autre activité (la recherche, les loisirs, un travail rémunéré, la vie familiale, etc.), alors l’énergie consacrée à cette invention sera fortement réduite.

Ce que nous avons essayé de suggérer (en matière de RII), c'est que cette « invention », laissée au génie de chacun, enseignant et étudiant, professeur et élève, usager ordinaire, ne saurait guère répondre aux besoins praxéologiques d'une société. Par contraste, il incombe à la recherche en didactique telle que nous avons tenté de l'illustrer ici d'étudier les conditions économiques et les contraintes écologiques d'une indispensable invention collective, suscitée et organisée par des arrangements institutionnels appropriés qui, bien entendu, doivent profiter du génie personnel de chacun.

L'accès à l'offre praxéologique en RII

L'étude des exposés sur la RII – lesquels circonscrivent, pour l'essentiel, l'offre praxéologique – a fait apparaître, au-delà du détail de l'inventaire réalisé, l'emprise de dichotomies on ne peut plus classiques dans leur principe séparant connaissances généralistes et connaissances « spécialisées ». Cette contrainte sociale et culturelle crée et recrée constamment des *frontières* souvent dépourvues de toute fonctionnalité entre, d'un côté, des praxéologies gardant (au moins provisoirement) un statut « ésotérique » (par rapport à tel ou tel champ de spécialité : « informatique », documentation, etc.) et, d'un autre côté, des praxéologies plus ouvertes, à statut réputé « exotérique ». S'agissant des moteurs de recherche, par exemple, il n'est pas immédiat de trouver dans les exposés adressés à l'utilisateur ordinaire une réponse à la question : « Où un moteur de recherche va-t-il chercher ce qu'il s'apprête à afficher à l'écran ? » Le critère du caractère supposé trop « technique » de la réponse à une telle question tend à donner vie à une offre praxéologique incapable de projeter sur l'activité de RII une intelligibilité autre que superficielle⁷. Par contraste, on ne voit guère comment cette intelligibilité pourrait naître si l'offre praxéologique n'apporte pas les notions de *crawler*, d'indexation, d'index, voire de ferme de serveurs, sans parler même ni des algorithmes (tel PageRank) permettant le classement des résultats affichés, ni des techniques de *référencement* (*search engine optimization*), qui font couple avec les techniques d'indexation propres aux moteurs de recherche.

Nous avons identifié chemin faisant un bon nombre de frontières qui traversent le champ de la RII ou le séparent de domaines dont, pourtant, une science de la RII est amenée à se nourrir. Ainsi en va-t-il des frontières qui à la fois unissent (substantiellement) et séparent (formellement, mais non moins réellement) la culture de la RII de la connaissance des éléments de mathématiques, de physique, d'informatique, de science de l'Internet et même de science du Web qui devraient y avoir une fonction nouménologique centrale : on est un *searcher* mieux « équipé » quand on n'est pas radicalement coupé de toute notion de

programmation Web, voire de référencement, par exemple, sans parler des notions utiles d'algèbre linéaire, qui, il est vrai, ne nous semblent pas encore avoir fait l'objet d'un travail d'élémentation idoine⁸. D'autres frontières externes ont été rencontrées et, dans une certaine mesure, franchies : ainsi de la frontière avec le droit, avec l'économie, avec, même, les sciences politiques. Une frontière surtout s'est imposée à l'observation : celle qui sépare la pratique française de la RII de l'emploi de la langue anglaise, frontière dont l'existence traditionnelle engendre des comportements de re-nominalisation (de l'anglais vers le français) et, de façon consubstantielle, de re-conceptualisation, ce qui provoque en plusieurs cas des difficultés praxéologiques précises.

Yves Chevallard (2007a) a ainsi fortement souligné les différences écologiques entre le mot skill, très anciennement inscrit dans la langue anglaise, et le mot de compétence par lequel on tend à le rendre en français. La notion de skill, en effet, inclut apparemment depuis toujours cette idée qu'un skill, ça s'apprend, ça s'acquiert par un travail systématique d'apprentissage, fût-il d'assez courte durée. D'une compétence, en revanche, on dira souvent qu'on la possède ou qu'on ne la possède pas, comme si le temps de son acquisition n'était qu'une parenthèse entre deux états dont la différence s'établit, au reste, par ce qu'on nomme significativement un bilan des compétences. De l'un à l'autre concept, on passe ainsi d'une vision dynamique à un point de vue de statique praxéologique.

Au-delà de la frontière avec l'anglais, on rencontre des frontières internes : on a maintes fois mentionné ce qu'on peut appeler « la frontière des booléens », dont nous avons même suivi plus haut le tracé dans son étonnant détail, mais on doit mentionner aussi, par exemple, la frontière qui sépare, d'un côté, un certain minimalisme de l'outillage et de l'emploi de ces outils et, de l'autre côté, le recours quasi « professionnel » à des stratégies de recherche mobilisant tout une panoplie de *finding tools*.

Le domaine praxéologique à investir apparaît ainsi couturé de frontières tracées généralement par la culture commune, dont l'effet le plus clair est d'interdire l'accès à tel ou tel ensemble de ressources praxéologiques, interdiction qu'est censé compenser, dans le fonctionnement social, la création de « spécialités » que leurs acteurs regardent souvent comme leur assujettissement principal : l'un dira ainsi « Je suis informaticien », l'autre « Je suis référenceur », une autre encore « Je suis un économiste » ou « Je suis juriste », etc. Par contraste, ce dont nous avons commencé d'étudier les conditions de possibilité (sous certaines contraintes) est un personnage qui ne se définit pas par une affiliation institutionnelle unique,

qui pourrait se présenter en disant : « Il m'arrive de faire de la RII », et qui, à cause de cela, est conduit à utiliser des connaissances de mathématiques, de droit, de programmation, d'anglais, de physique, d'économie, etc., sans pour autant se dire « mathématicien », « programmeur », « angliciste », « physicien », « économiste », etc. En d'autres termes, il s'agit de se référer ici à un autre paradigme culturel, gouvernant autrement notre rapport aux différents champs disciplinaires. Pour le dire autrement, il s'agit de se référer à une culture de la *codisciplinarité*.

Culture de la RII et codisciplinarité

La codisciplinarité clairement assumée, c'est-à-dire la pratique ouverte consistant à faire travailler ensemble – tel est le sens du mot *synergie* – des connaissances appartenant à différentes disciplines au service d'un même projet de connaissance ou d'action est une idée neuve plutôt qu'une réalité accomplie. C'est elle, bien entendu, qui est au cœur de la notion d'*enquête codisciplinaire*, à laquelle nous allons revenir un peu plus loin. Mais nous voudrions ici préciser en quelques mots ce que peut être une culture codisciplinaire au service d'un projet, alors même que le modèle dominant est, aujourd'hui (car ce ne fut pas toujours le cas), celui des lourdes cultures monodisciplinaires engendrées par la formation scolaire et universitaire. Examinant le problème de la construction d'une culture codisciplinaire (telle la culture de la RII dont nous avons esquissé un portrait-robot), d'aucuns se placent spontanément, sans toujours en être nettement conscients, dans le paradigme des cultures monodisciplinaires. Or, cette situation engendre une apparente antinomie d'autant plus violente qu'une culture monodisciplinaire assume généralement le postulat, non seulement qu'elle est irréductible aux autres cultures disciplinaires, mais surtout qu'elle *se pose en s'y opposant* : dans ce « jeu » monodisciplinaire, être sociologue, ainsi, c'est autant l'être que *ne pas être* psychologue, ou ethnologue, etc. ; être psychologue social, c'est aussi *ne pas être* sociologue, etc.

Contre ces habitus académiques, la notion de codisciplinarité ne se conçoit certes pas sans la notion de discipline. Mais elle suppose en même temps la notion de *synergie entre disciplines*, ce à quoi le « monodisciplinarisme » tend à faire barrage, alors même que, en réalité, la plupart des disciplines se nourrissent d'emprunts codisciplinaires, mais qui ne sont généralement visibles que des *insiders*. Pour le dire autrement, la codisciplinarité existe activement au sein des disciplines, mais elle y est masquée par l'étiquette monodisciplinaire qui les recouvre. Il ne s'agit donc pas tant de se demander si une culture codisciplinaire est possible, mais *comment* elle devient possible. La réponse que nous avons avancée est que le

facteur écologique essentiel, ici, est à chercher dans *la culture de la discipline*, qui fait par exemple qu'un « physicien » sait des mathématiques, tandis qu'à l'inverse nombre de « mathématiciens », qui, sauf exception, ont un jour *appris* de la physique, se sont, si l'on peut dire, empressés de l'oublier. La capacité de connaître est ainsi liée à la capacité qu'a une *culture* de faire apparaître à ceux qui en sont les acteurs comme normale, légitime, allant de soi, la connaissance de tel ou tel domaine praxéologique.

Un exemple illustrera ce constat ⁹. Dans un livre d'entretiens récent, l'anthropologue Françoise Héritier, spécialiste des systèmes de parenté et d'alliance, ancienne professeure au Collège de France, interroge son interlocutrice, Caroline Broué, en ces termes (Héritier, 2008, p. 33).

Supposez que je vous demande, à vous, par quel terme vous désignez le fils de la fille du frère de la mère de votre père, et je suis sûre qu'il vous faudra plus d'un petit temps de réflexion (alors qu'il ne s'agit que d'un cousin issu de cousins germains) !

Un peu plus loin, l'intervieweuse résume ainsi ce propos : « Ce que vous voulez dire, c'est que tout le monde n'est pas capable de faire cette gymnastique intellectuelle. » L'anthropologue le lui confirme (p. 34) : « Non. Tout le monde n'en est pas capable, et beaucoup de collègues n'aiment pas travailler ces sujets. » Cette affirmation, sans doute empiriquement vérifiée, pourrait laisser penser que seuls quelques spécialistes de « l'algèbre » des relations de parenté peuvent y arriver. Or, Françoise Héritier ne cache pas que sa facilité dans cette matière, elle l'a acquise toute petite, bien avant de s'assujettir à la culture de l'anthropologie, lorsque, en vacances avec ses deux grands-mères, elle les entendait commenter à l'infini les liens de parenté au sein des communautés villageoises proches.

Elles disaient par exemple... : « La fille à la Martine qui a épousé le fils du boulanger de La Chenale, qui lui, avait épousé une cousine germaine de la fille à la Martine... » Vous voyez le genre ! Moi, j'enregistrais tout ça et j'essayais d'en tirer la leçon graphique, de me représenter mentalement les cheminements selon les différentes positions de cousinage. C'était extrêmement instructif, parce que cela me disposait à une gymnastique mentale qui est peut-être un des éléments qui, pour la plupart des gens, fait repoussoir à l'encontre des études sur la parenté. (pp. 32-33)

Le noyau praxéologique que la culture des anthropologues ne parvient pas à faire vivre parmi les siens – jongler avec les liens de parenté –, la culture paysanne dans laquelle était plongée, enfant, Françoise Héritier, le rendait banal, à la portée de chacun de ses membres. La culture

prime : c'est dans cette perspective qu'on abordera l'étude didactique de l'enquête codisciplinaire.

Recherches à venir

Du point de vue des besoins praxéologiques, il existe une affinité évidente entre RII et enquête codisciplinaire : dans l'un et l'autre cas, pour maîtriser l'outil (la RII) comme pour le mettre en œuvre dans une enquête et conduire cette enquête à son terme, il convient d'entrer dans une *culture de la codisciplinarité* assumée au plan épistémologique et opérationnelle au plan des enquêtes proprement dites, c'est-à-dire de la construction de réponses R^\heartsuit à des questions Q . La capacité à assumer des *synergies disciplinaires*, fût-ce, en quelques cas, de façon périphérique ou marginale, qu'appelle l'immense majorité des enquêtes de ce type, même lorsqu'il s'agit d'enquêtes à *dominante (mono)disciplinaire* (où la question Q est par exemple regardée, dans la culture de l'institution où l'enquête trouve son habitat, comme une question « de mathématiques », ou « d'histoire », ou « de biologie », etc.), est une donnée essentielle tant en didactique de la RII qu'en didactique de l'enquête codisciplinaire. *Un premier thème de recherche* qui émerge ainsi de notre travail est celui de *l'étude des conditions et contraintes des cultures de la codisciplinarité*, étude que le développement des dispositifs de formation scolaire et universitaire que nous pouvons regrouper sous la dénomination de *parcours d'étude et de recherche* (PER) motive, et à laquelle il offre en même temps des *terrains* d'observation et, le cas échéant, d'expérimentation.

Un deuxième thème de recherche est évidemment celui de l'étude des conditions et contraintes de *l'enquête codisciplinaire* elle-même. En tant que type de dispositif de construction de connaissances, l'enquête codisciplinaire peut être rapprochée d'autres types de dispositifs de construction de connaissances qui illustrent son « genre prochain » mais par rapport auxquels il importe de préciser sa « différence spécifique » : on peut songer par exemple, ici, à la notion de *WebQuest*, que son initiateur, Bernie Dodge (San Diego State University) définissait ainsi (Dodge, 1995) : “A WebQuest is an inquiry-oriented activity in which some or all of the information that learners interact with comes from resources on the internet, optionally supplemented with videoconferencing.” C'est encore ce que nous serons amenée à faire, en collaboration avec Y. Chevallard notamment, vis-à-vis du courant d'origine américaine de *l'inquiry-based learning*, dans le cadre d'un projet de recherche européen proposé par Paolo Boero (université de Gênes), dont le titre original non abrégé est : *Large scale dissemination and use of inquiry-based teaching methods in the perspective of lifelong learning of mathematics: preparing tools; and creating teachers', students' and*

parents' consensus. Dans ce cadre, nous serons conduite en particulier à contribuer à la création et à l'étude d'un dispositif de formation – qui fournira un « terrain » à des recherches indispensables à son bon fonctionnement – dans un collège marseillais autour de la pratique de l'enquête codisciplinaire par des élèves du cycle central, en articulation avec les enseignements disciplinaires classiques (l'accent étant mis, en l'espèce, quoique non exclusivement, sur la mobilisation et l'usage de connaissances mathématiques en de telles enquêtes).

Plus généralement, notre travail en didactique de l'enquête codisciplinaire, en particulier à propos du rôle que peut y jouer la RII, devrait trouver une source privilégiée de questionnements dans certaines des formations en sciences de l'éducation, et cela pas uniquement à l'occasion de la fabrication des « chefs-d'œuvre » que mentionnait plus haut, avec une pointe de morosité, un observateur sociologue.

À titre d'illustration (qui pourrait relever du secteur de l'éducation en santé), mentionnons ici une étude récemment publiée sous le titre “Medical Myths” (Vreeman & Carroll, 2007). Parue dans le British Medical Journal, elle porte sur sept assertions communément reçues (y compris par des médecins) : “People should drink at least eight glasses of water a day” / “We use only 10% of our brains” / “Hair and fingernails continue to grow after death” / “Shaving hair causes it to grow back faster, darker, or coarser” / “Reading in dim light ruins your eyesight” / “Eating turkey makes people especially drowsy” / “Mobile phones create considerable electromagnetic interference in hospitals”. Les auteurs précisent à propos de leur enquête (à cet égard « révolutionnaire ») : “We used Medline and Google to search for evidence to support or refute each of these claims.” S'agissant ainsi de l'affirmation selon laquelle “we use only 10% of our brains”, ils peuvent conclure (entre autres choses) que “some sources attribute this claim to Albert Einstein, but no such reference or statement by Einstein has ever been recorded”, avant de préciser : “This myth arose as early as 1907, propagated by multiple sources advocating the power of self improvement and tapping into each person's unrealised latent abilities.”

Dans la perspective précédente, nous serons amenée à porter notre attention sur les questions liées à l'intégration de la RII dans le travail d'enquête codisciplinaire. Un troisième thème de recherche issu de notre travail est ainsi celui de l'étude des conditions et contraintes du développement local d'une culture de la RII, dont nous examinerons au double plan qualitatif et quantitatif l'adéquation aux besoins du travail d'enquête codisciplinaire.

Corrélativement, nous approfondirons notre étude du problème de la construction d'une « science de la RII » portée par une culture de la RII appropriée, en situant cette étude dans le cadre plus large d'une science et d'une didactique du Web et de ses usages, comme parties de l'étude plus large des usages didactiques des technologies de l'information et de la communication (Bruillard & Baron, 2006), ce qui recoupe partiellement des propositions programmatiques récentes (Serres, 2007). Il s'agit là d'un *quatrième thème de recherche* qui nous conduira, dans le cadre d'un dispositif expérimental encore à créer, conjuguant les fonctions de *forum*, *atelier* et *boutique*, à étudier les besoins praxéologiques liés aux usages du Web¹⁰. Un tel dispositif expérimental aura vocation à permettre d'observer, dans leur rapport au Web, tant les élèves que leurs professeurs (au collège par exemple), les formés que leurs formateurs (à l'université notamment), aussi bien que les membres de communautés virtuelles. Deux grands domaines d'activité liés au Web seront à cet égard au cœur de nos travaux : d'une part, bien entendu, la RII et ses usages ; d'autre part, le *e-learning*, domaine sur lequel nous avons commencé de travailler aussi bien au plan de l'ingénierie des formations qu'au plan de la théorisation didactique (Chevallard & Ladage, 2008). Plusieurs sujets de recherche sont actuellement envisagés ou ont commencé de faire l'objet d'un travail spécifique : du premier domaine relève ainsi l'étude, au croisement de travaux d'orientations diverses (Bruillard, 2007 ; Lakhani & McAfee, 2007), du rapport à l'encyclopédie Wikipédia chez les usagers du Web ; du second domaine, l'étude des praxéologies développées autour de la notion et de l'usage du *portfolio* électronique, sujet sur lequel nous disposons d'ores et déjà d'un ensemble non négligeable de données (Ladage & Chevallard, 2008).

Tout cela constitue sans doute un programme de travail ambitieux. Nous ne le croyons pas démesuré. Si nous osions, nous reprendrions ici à notre compte, très modestement, cet aphorisme rapporté par Larry Page et que citent Vise et Malseed (2006, p. 11) : “There is a phrase I learned in college called, *Having a healthy disregard for the impossible*, Page said. That is a really good phrase...”

¹ On rappellera cependant notre enquête sur le « désir actuel de disposer d'une réponse » à certaines questions de RII, sur laquelle nous nous sommes appuyée dans le dernier chapitre de la quatrième partie de notre travail. En outre, rappelons l'enquête en ligne (que nous avons encadré) auprès de cadres de santé dont 76 % (d'un effectif de 1210) ont dit ressentir un réel besoin de formation dans le domaine de la RII et ont explicité parfois longuement, en réponse à une question ouverte, la nature du déficit de connaissances et de compétences dont ils estimaient pâtir en la matière.

² Nous reprenons ici la traduction proposée par Georges Liébert dans Nietzsche, 2000, p. 20.

³ Le livre de Notess, nous l'avons souligné, examine les problèmes que doivent affronter les formateurs à la RII. Le livre de Hock veut, lui, instruire le *searcher* : il pourra par exemple être utilisé par un formateur comme référence de base pour déterminer le contenu de son enseignement.

⁴ Pour éclairer la formule schématique que nous employons ci-après, on peut prendre l'exemple de Google. Le moteur de recherche Google, quand on le réduit à l'interface standard (avec simplement sa boîte de requête et son bouton « Rechercher ») peut être regardé comme entrant dans un équipement de simple amateur, tandis que le dispositif de « recherche avancée », avec toute la gamme de ses fonctionnalités, participe, lui, d'un équipement de type « professionnel ». Utiliser l'interface de base de Google en n'y mettant jamais qu'un, deux ou trois mots clés, c'est user en amateur d'un dispositif pour amateurs. On imagine aisément ce que peut être une utilisation plus professionnelle de cet équipement d'amateur : emploi des opérateurs booléens (ce qui revient en partie à utiliser l'interface de recherche avancée) ou, plus trivialement, formulation de requêtes « longues », non réduites à un très petit nombre de mots.

⁵ Voir <http://www.education.gouv.fr/bo/2006/42/MENE0602673C.htm>.

⁶ Bien entendu, le système scolaire réagit et des documents sont proposés qui vont au-delà de la simple énumération de quelques types de tâches grossièrement définis (pour un exemple, voir ainsi « Validation des items B2I », 2008).

⁷ Dans *La face cachée de Google*, les auteurs – le collectif italien Ippolita – soulèvent la question suivante, qui reste ignorée de l'offre praxéologique tout-venant (Ippolita, 2008, pp. 11-12) : « ... l'extraordinaire rapidité des résultats de recherche est le fruit d'une sélection minutieuse qui n'a rien de transparent. De fait, comment des millions d'utilisateurs pourraient-ils feuilleter simultanément, à tout instant, la base de données de Google tout entière, s'il n'y avait pas des filtres opportuns pour restreindre le champ de la recherche, par exemple en la limitant aux données rédigées dans leur langue d'origine ? »

⁸ Pour une ébauche, voir Chevallard, 2008.

⁹ Je dois à Yves Chevallard d'avoir attiré mon attention sur cet exemple.

¹⁰ Bien entendu, les usages du Web ne se réduisent pas à la RII : il est éclairant de ce point de vue de comparer le contenu de notre travail avec, par exemple, celui de Greffe et Sonnac, 2008.

Bibliographie

*Notre génération a ce défaut ancré qu'elle refuse d'admettre
tout ce qui vient des modernes. Aussi, quand je trouve une idée personnelle,
si je veux la publier, je l'attribue à quelqu'un d'autre...*

Adélard de Bath, XII^e s.

N. B. Concernant les documents mentionnés dans ce travail, on trouvera ci-après, sauf erreur de notre part, toutes les références qui n'ont pas été précisées dans les notes de fin de partie (ou, parfois, dans le texte lui-même). S'agissant des articles de l'encyclopédie Wikipedia (en anglais) ou Wikipédia (en français), l'URL proposée renvoie à la version courante de l'article ; la date indiquée par nos soins, d'une part, et l'onglet *history* (en anglais) ou *historique* (en français) figurant en haut de l'article, d'autre part, permettent alors de retrouver la version que nous avons citée. Plus généralement, les URL que nous proposons renvoient (ou renvoyaient, début septembre 2008) soit au document lui-même, soit à une page Web d'où l'on peut accéder à ce document.

Abondance. (2007, 22 novembre). J'ai de la chance de perdre 110 millions de dollars par an...

<http://blog.abondance.com/2007/11/jai-de-la-chance-de-perdre-110-millions.html>

Académie des sciences royale suédoise. (2007, 9 octobre). Le Prix Nobel de Physique 2007.

Nobelprize.org.

http://nobelprize.org/nobel_prizes/physics/laureates/2007/press-fr.html

Adler, J. (2006). Epistemological Problems of Testimony. In E. N. Zalta (Éd.). *The Stanford Encyclopedia of Philosophy*.

<http://plato.stanford.edu/archives/sum2007/entries/testimony-episprob/>

Adresse Web. (2008, 17 mars). *Wikipédia, L'encyclopédie libre*.

http://fr.wikipedia.org/wiki/Adresse_web

Afnic. (2007, 21 mai). Initiation. Qu'est-ce qu'un nom de domaine ?

<http://www.afnic.fr/guide/introduction>

Agamben, G. (2007). *Qu'est-ce qu'un dispositif ?* Paris : Payot & Rivages.

Asymmetric Digital Subscriber Line. (2008, 4 septembre). *Wikipédia, L'encyclopédie libre*.

http://fr.wikipedia.org/wiki/Asymmetric_Digital_Subscriber_Line

Alpert, J. & Hajaj, N. (2008, 25 juillet). We knew the web was big... [post] *The Official Google Blog*.

- <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>
- Ameka, B., Béziat, J., Florentin, C., Marty, N. & de Vulpillières, T. (2006). *Dictionnaire des nouvelles technologies en éducation*. Paris : Nathan.
- American Psychological Association. (2001). *Publication Manual*. Washington : American Psychological Association.
- Archimédien. (2008, 28 mai). *Wikipédia, L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/Archim%C3%A9dien>
- Assouline, P. (2005, 15 janvier). Les mots des déportés. *La république des livres*.
http://passouline.blog.lemonde.fr/2005/01/15/2005_01_les_mots_des_dp/
- Atkins, B. T. et al. (1987). *Dictionnaire français-anglais / English-French Dictionary*. Paris : Dictionnaires Le Robert / Londres : Collins.
- Augier, M. (2000). *HTML le langage du Web*. Paris : PUF.
- Ayto, J. (1994). *Dictionary of Word Origins*. Amersham (Royaume-Uni) : Columbia Marketing.
- Bachelard, G. (1934/2003). *Le nouvel esprit scientifique*. Paris : PUF.
- Bachelard, G. (1938/2000). *La formation de l'esprit scientifique*. Paris : Vrin.
- Bachelard, G. (1970). *Études*. Paris : Vrin.
- Bogatin, D. (2006, 9 août). Google CEO Eric Schmidt: 'all world's information includes personal information'. *ZDNet* [post]
<http://blogs.zdnet.com/micro-markets/index.php?p=320>
- Barron, P. (2008). Internet Searching. *The Infography – Fields of Knowledge*.
<http://www.infography.com/content/813746743748.html>
- Battelle, J. (2006). *The Search. How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York : Penguin Group.
- Bauche, G. (1996). *Tout savoir sur Internet*. Paris : Arléa.
- Bécalséri, L. (2007). *Internet. Le guide complet*. Paris : Micro Application.
- Benenson, J. & Juanals, B. (2003). *Guide bilingue de la communication électronique. Electronic Communications Bilingual Guide*. Paris : Pocket.
- Benenson, J. & Juanals, B. (2004). *Dictionnaire bilingue Internet et multimédia. Internet and Multimedia Bilingual Dictionary*. Paris : Pocket.
- Berners-Lee, T., Masinter, L. & McCahill, M. (1994). Uniform Resource Locators (URL). *IETF – Network Working Group*.
<http://www.ietf.org/rfc/rfc1738.txt>
- Berners-Lee, T. avec Fischetti, M. (2000). *Weaving the Web. The original design and ultimate destiny of the World Wide Web*. New York : HarperCollins.

- Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N. & Weitzner, D. J. (2006). Creating a science of the Web. *Science*, 313, 769-771.
- Berners-Lee, T., Hall, W., Hendler, J. A., O'Hara, K., Shadbolt, N. & Weitzner, D. J. (2006). A Framework for Web Science. *Foundations and Trends in Web Science*, 1(1), 1-130.
<http://eprints.ecs.soton.ac.uk/13347/1/1800000001%5B1%5D.pdf>
- Berry, G. (2008). *Pourquoi et comment le monde devient numérique*. Paris : Collège de France/Fayard.
- Bigot, R. (2005). *La diffusion des technologies de l'information dans la société française*.
<http://www.credoc.fr/publications/abstract.php?ref=R237>
- Bigot, R. (2006). *La diffusion des technologies de l'information dans la société française*.
http://www.arcep.fr/uploads/tx_gspublication/etude-credoc2006.pdf
- Bigot, R. & Croutte, P. (2007). *La diffusion des technologies de l'information dans la société française*.
http://www.arcep.fr/uploads/tx_gspublication/etude-credoc-2007.pdf
- Bit error ratio. (2008, 17 juillet). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Bit_error_ratio
- Bootstrapping. (Computing). (2008, 30 janvier). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Bootstrapping_%28computing%29#_note-0
- Bortzmeyer, S. (2007, 12 février). Fiche de lecture : Google-moi.
<http://www.bortzmeyer.org/google-moi.html>
- Borzeix, J.-M. (2006). *Les Carnets d'un francophone*. Montréal : Boréal.
- Bosch, M. (1994). *La dimensión ostensiva en la actividad matemática. El caso de la proporcionalidad*. Thèse de doctorat non publiée, Université autonome de Barcelone – Barcelone.
- Bosch, M. & Gascón, J. (2005): La praxéologie comme unité d'analyse des processus didactiques. In A. Mercier & C. Margolinas, C. (Éds), *Balises en Didactique des Mathématiques* (pp. 107-122). Grenoble : La Pensée Sauvage.
- Boucherie, A. (2007, 23 septembre). Je me suis fait punir à cause du Hotlink... *Le blog d'Antoine*.
<http://www.leblogdantoine.com/hack/je-me-suis-fait-punir-a-cause-du-hotlink-162>
- Boughanem, M. & Savoy, J. (2008). *Recherche d'information. État des lieux et perspectives*. Paris : Lavoisier.
- Boutiques de sciences (Les). (s. d.). *Espace des citoyens / santé et bien-être (Québec)*.
[http://www.espacecsb.com/wiki/index.php?title=Les Boutiques de sciences&printable=yes](http://www.espacecsb.com/wiki/index.php?title=Les_Boutiques_de_sciences&printable=yes)

- Bowker, R. R. (1883). *The work of the nineteenth-century librarian for the librarian of the twentieth*. In American Library Association. (Éd.), *Papers and proceedings of the sixth general Meeting (Buffalo, 14-17 août 1993)* (pp. 85-88). Boston : Rockwell and Churchill.
- http://stevelawson.name/seealso/archives/2007/09/how_1983_wasnt_like_1983.html
- Brandt-Pomares, P. (2003). *Les nouvelles technologies de l'information et de la communication dans les enseignements technologiques. De l'organisation des savoirs aux conditions d'étude didactique de la consultation d'information*. Thèse de doctorat non publiée, Université de Provence – Lambesc.
- http://www.aix-mrs.iufm.fr/formations/filieres/te/data/publications/brandt/these_pbp.pdf
- Bréda, I. & Hourt, C. (2007). *Les bons usages d'Internet*. Paris : Éditions J'ai Lu.
- Brin, S. & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30, 107-117.
- <http://infolab.stanford.edu/~backrub/google.html#ref>
- Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble : La Pensée sauvage.
- Bruillard, É. (2007, mars). L'éducation face à Wikipédia : la rejeter ou la domestiquer ? *Médialog*, 61, 39-45.
- <http://www.ac-creteil.fr/medialog/ARCHIVE61/wikipedia61.pdf>
- Bruillard, É. & Baron, G.-L. (2006). Usages en milieu scolaire : caractérisation, observation et évaluation. In M. Grandbastien & J.-M. Labat (Éds), *Environnements Informatiques pour l'Apprentissage Humain* (pp. 265-280). Paris : Hermès.
- Bush, V. (1945, juillet). "As We May Think". *The Atlantic Monthly*.
- <http://www.theatlantic.com/doc/194507/bush/>
- Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N. & Hullender, G. (2005). Learning to rank using gradient descent. *ACM International Conference Proceeding Series : Vol. 119. Proceedings of the 22nd international conference on Machine learning, Bonn, Allemagne, 7-11 août 2005* (pp. 89-96). New York : ACM.
- <http://doi.acm.org/10.1145/1102351.1102363>
- Buttner, Y., Maurin, A. & Thouveny, B. (2002). *Le droit de la vie scolaire. Écoles, collèges, lycées*. Paris : Dalloz.
- Buttner, Y. & Maurin, A. (2007). *Le droit de la vie scolaire. Écoles, collèges, lycées* (4^e éd.). Paris : Dalloz.
- Cailliau, R. (1995, 2 novembre). A Short History of the Web. *NetValley*.
- http://www.netvalley.com/archives/mirrors/robert_cailliau_speech.htm
- Cambridge Dictionary of American English*. (2002). [dictionnaire en ligne]

- <http://dictionary.cambridge.org/default.asp?dict=A>
- Campbell, J. (2003). The Case for Creating a Scholars Portal to the Web: A White Paper. *Libraries and the Academy* 1(1).
- <http://www.arl.org/bm~doc/portal.pdf>
- Carr, N. (2008a). *The Big Switch: Rewiring the World, from Edison to Google*. New York : W. W. Norton.
- Carr N. (2008b, juillet-août). Is Google Making Us Stupid? *The Atlantic Monthly*, 301(6).
- <http://www.theatlantic.com/doc/200807/google>
- Carrier, A. (2007). *Informatique et internet au collège et au lycée*. Grenoble : CRDP.
- Cassin, B. (1995). *L'effet sophistique*. Paris : Gallimard.
- Cassin, B. (Éd.). (2004). *Vocabulaire européen des philosophies*. Paris : Éditions du Seuil / Dictionnaires Le Robert.
- Cassin, B. (2007). *Google-moi. La deuxième mission de l'Amérique*. Paris : Albin Michel.
- Cassin, B. & Mathieu, M. (2000). *Voir Hélène en toute femme. D'Homère à Lacan*. Paris : Les empêcheurs de penser en rond.
- Cassin, B., Cayla, O. & Salazar, P.-J. (2004). *Vérité, réconciliation, réparation*. Paris : Seuil.
- Cassin, B. avec Briffard, C. (2006). Le vocabulaire européen des philosophes, dictionnaire des intraduisibles. *Texte !* [Revue électronique]
- http://revue-texto.net/1996-2007/Dialogues/Cassin_interview.html#1
- Cauvin, M. (1970). *Le renouveau pédagogique en Allemagne de 1890 à 1933*. Paris : Armand Colin.
- Caveing, M. (1990). Introduction générale. In M. Caveing & B. Vitrac (Éd.), *Euclide. Les Éléments. Livres I-IV* (pp. 13-148). Paris : PUF.
- Cerf, V. avec Alberganti, M. (2008, 5 avril). Vers l'Internet à tout faire. *Le Monde*.
- http://www.litt-and-co.org/c_g/Textes%20CG%20mensuel/textes_Avril%202008.pdf
- Charrassin, J.-B., Hindell, M., Rintoul, S. R., Roquet, F., Sokolov, S., Biuwe, M. et al. (2008). Southern Ocean frontal structure and sea-ice formation rates revealed by elephant seals. *PNAS*.
- <http://www.pnas.org/content/early/2008/08/08/0800790105.full.pdf+html?sid=888b5223-9a21-40cf-955c-41b41ab65c8e>
- Chevallard, Y. (1989). Le concept de rapport au savoir – Rapport personnel, rapport institutionnel, rapport officiel. In A. Bessot (Éd.), *Actes du Séminaire de didactique des mathématiques et de l'informatique* (pp. 211-235). Grenoble : Université Joseph Fourier Grenoble I.

- Chevallard, Y. (1990). Évaluation, vérification, objectivation. In J. Colomb & J. Marsenach (Éds), *L'évaluateur en révolution* (pp. 13-36). Paris : INRP.
- Chevallard, Y. (1991). *La transposition didactique. Du savoir savant au savoir enseigné* (2^e éd.). Grenoble : La Pensée sauvage.
- Chevallard, Y. (1992). Concepts fondamentaux de la didactique : perspectives apportées par une approche anthropologique. *Recherches en didactique des mathématiques*, 12(1), 7-32.
- Chevallard, Y. (1995). Note sur l'individualisation de la formation. *Didaskalia*, 6, 115-131.
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=104
- Chevallard, Y. (1998a). À propos des TICE : transmission et appropriation du savoir, nouveaux rôles de l'enseignant, organisation de l'établissement. Conférence à l'université d'été « *Les TICE : vers une transformation des pratiques pédagogiques et de l'organisation de l'établissement* » (Toulouse, 26-28 août 1998).
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=32
- Chevallard, Y. (1998b). Questions vives, savoirs moribonds : le problème curriculaire aujourd'hui. *Actes du colloque « Défendre et transformer l'école pour tous », Marseille, 3-5 octobre 1997* [CD-ROM].
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=19
- Chevallard, Y. (2001a). *Les TPE comme problème didactique*. In T. Assude & B. Grugeon (Éd.), *Actes du séminaire national de didactique des mathématiques 2001* (pp. 177-188). Paris : IREM et ARDM.
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=14.
- Chevallard, Y. (2001b). Les mathématiques et le monde : dépasser « l'horreur instrumentale », *Quadrature*, 41, 25-40.
http://yves.chevallard.free.fr/spip/spip/rubrique.php3?id_rubrique=4
- Chevallard, Y. (2003). Approche anthropologique du rapport au savoir et didactique des mathématiques. In S. Maury S. & M. Caillot (Éd.), *Rapport au savoir et didactiques* (pp. 81-104). Paris : Éditions Fabert.
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=62
- Chevallard, Y. (2006). La théorie anthropologique des faits didactiques devant l'enseignement de l'altérité culturelle et linguistique. Le point de vue d'un *outsider*. In J. Aden (Éd.), *Création identitaire et altérité en didactique des langues* (pp. 17-38). Paris : Éditions Le Manuscrit.
http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=70
- Chevallard, Y. (2007a). *Journal du Séminaire de formation de formateurs 2006-2007*.

- http://yves.chevallard.free.fr/spip/spip/rubrique.php3?id_rubrique=11
- Chevallard, Y. (2007b). Readjusting didactics to a changing epistemology. *European Educational Research Journal*, 6(2), 9-27.
- http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=109
- Chevallard, Y. (2007c, juillet-août). Les mathématiques à l'école : pour une révolution épistémologique et didactique. *Bulletin de l'APMEP*, 471, 439-461.
- Chevallard Y. (2007d). Passé et présent de la théorie anthropologique du didactique. In L. Ruiz-Higueras, A. Estepa, & F. Javier García (Éd.), *Sociedad, Escuela y Matemáticas. Aportaciones de la Teoría Antropológica de la Didáctica* (pp. 705-746). Universidad de Jaén.
- http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=134
- Chevallard, Y. (2008). *Journal du séminaire TAD/IDD 2007-2008*.
- http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=131
- Chevallard, Y., Bosch, M. & Gascón, J. (1997). *Estudiar matemáticas. El eslabón perdido entre la enseñanza y el aprendizaje*. Barcelone : ICE/Horsori.
- Chevallard, Y. & Matheron, Y. (2002). Travaux Personnels Encadrés : un cadre d'analyse didactique pour un changement majeur dans l'enseignement au Lycée. In M. Bridenne (Éd.), *Actes des journées de Dijon (24-25 mai 2002)* (pp. 141-150). Dijon : IREM de Dijon.
- http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=50
- Chevallard, Y. & Ladage, C. (2008, juin). E-learning as a touchstone for didactic theory, and conversely. *Journal of e-Learning and Knowledge Society*, 4(2), 163-171.
- http://www.je-lks.it/en/08_02/5Met_chev_franc.pdf
- Chu, S. K. & Law, N. (2007). Development of Information Search Expertise: Postgraduates' Knowledge of Searching Skills. *Libraries and the Academy*, 7(3), 295-316.
- http://muse.jhu.edu/journals/portal_libraries_and_the_academy/toc/pla7.3.html
- Cirade, G. (2006). *Devenir professeur de mathématiques : entre problèmes de la profession et formation en IUFM. Les mathématiques comme problème professionnel*. Thèse de doctorat non publiée, Université de Provence – Marseille.
- Colin, J.-P. (1994). *Dictionnaire des difficultés du français*. Paris : Le Robert.
- .com. (2008, 7 septembre). *Wikipédia, L'encyclopédie libre*.
- <http://fr.wikipedia.org/wiki/.com>
- Comer, D. E. (2004). *Computer Networks and Internets* (4^e éd.). Upper Saddle River (NJ): Prentice Hall.
- Communications protocol. (2008, 8 septembre). *Wikipedia, The Free Encyclopaedia*.

- http://en.wikipedia.org/wiki/Communications_protocol
- Compact Oxford English Dictionary. (2004). [dictionnaire en ligne]
http://www.askoxford.com/dictionaries/compact_oed/
- ComScore. (2007, 10 octobre). *61 billion searches conducted worldwide in August. Google ranks as top global search property* [Communiqué de presse].
<http://www.comscore.com/press/release.asp?press=1802>
- Cookie (informatique). (2008, 7 septembre). Wikipédia, L'encyclopédie libre.
[http://fr.wikipedia.org/wiki/Cookie_\(informatique\)](http://fr.wikipedia.org/wiki/Cookie_(informatique))
- Corporate law. (2008, 23 août). Wikipedia, The Free Encyclopaedia.
http://en.wikipedia.org/wiki/Company_law
- Criticism of Wikipedia. (2008, 6 septembre). Wikipedia, The Free Encyclopaedia.
http://en.wikipedia.org/wiki/Criticism_of_Wikipedia
- D'Alembert, J. L. R. (1986a). Discours préliminaire. In A. Pons (Éd.), *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers* (vol. 1). Paris : Flammarion. (Publication originale : 1751)
- D'Alembert, J. L. R. (1986b). Éléments des sciences. In A. Pons (Éd.), *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers* (vol. 1). Paris : Flammarion. (Publication originale : 1751)
- de Clerck, A. (2007). *Les dessous d'Internet. Au fil de l'électricité, histoire de comprendre*. Paris : Ellipses.
- Delaney, K. J. (2006, 25 janvier). Google Launches Service in China. *The Wall Street Journal on line*.
http://online.wsj.com/public/article_print/SB113815053685555392-M_Jbt9XLO46g2XyuSdu2SLiWIGQ_20070125.html
- Delprat, L. & Halpern, C. (2007). *Communication et Internet : pouvoirs et droits*. Paris : Vuibert.
- Descartes, R. (1979). *Méditations métaphysiques* [Textes présentés par F. Khodoss]. Paris : PUF. (Publication originale : 1647)
<http://books.google.fr/books?id=98I9AAAIAAJ&pg=PA45&lpg=PA45&dq=Descartes+Secondes+objections&source>
- Detienne, M. (2003). *Comment être autochtone. Du pur Athénien au Français raciné*. Paris : Éditions du Seuil.
- Dodge, B. (1995). Some Thoughts About WebQuests. *WebQuest.org*.
http://webquest.sdsu.edu/about_webquests.html
- Domain Name System. (2008, 18 mai). Wikipedia, The Free Encyclopaedia.

- http://en.wikipedia.org/wiki/Domain_Name_System
- Douady, R. (1986). Jeux de cadres et dialectique outil/objet. *Recherches en Didactique des Mathématiques*, 7(2), 5-32.
- Duchêne, R. (1990). *Jean de La Fontaine*. Paris : Fayard.
- Ducourtieux, C. (2008, 3-4 février). Microsoft veut s'offrir Yahoo!. *Le Monde*.
<http://www.pubwebmaster.com/uprate/index.php/2008/02/03/836-microsoft-veut-s-offrir-yahoo-pour-contrer-google>
- Durand, D. (2005, 1^{er} juin). Google devient aléatoire via Shmoogole ! (à quand le procès ?).
<http://media-tech.blogspot.com/2005/06/google-devient-alatoire-via-shmoogole.html>
- Durkheim, É. (1990). *L'évolution pédagogique en France*. Paris : PUF. (Publication originale : 1938)
- Duvauchelle, B. (s. d.). Introduction : les enjeux du B2i. *Le Café pédagogique*.
http://www.cafepedagogique.net/lesdossiers/Pages/b2i_index.aspx
- Dystopia. (2008, 20 août). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Dystopia>
- Eco, U. (1965). *L'œuvre ouverte*. Paris : Seuil.
- Edsger Dijkstra. (2007, 29 septembre). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Edsger_Dijkstra
- Encarta World English Dictionary. (2003) [dictionnaire en ligne]
<http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>
- Enochsson, AB. (2005). The development of children's Web searching skills – a non-linear model. *Information Research*, 11(1).
<http://informationr.net/ir/11-1/paper240.html>
- Eolas, Me. (2006, 7 août). La loi DADVSI commentée. *Journal d'un avocat*.
<http://www.maitre-eolas.fr/2006/08/07/411-la-loi-dadvsi-commentee>
- Ertzscheid, O. (2005, 20 septembre). Trois manières de fouiller le web. *Affordance.info*.
http://affordance.typepad.com/mon_weblog/2005/09/trois_manieres_d.html
- Fallows, D. (2005). *Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve*. Pew Internet & American Life Project.
http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf &
http://www.pewinternet.org/pdfs/PIP_Search_Questions.pdf
- Fallows, D. (2008). *Almost half of all internet users now use search engines on a typical day*. Pew Internet & American Life Project.
http://pewinternet.org/pdfs/PIP_Search_Aug08.pdf

- Farmelo, G. (2002). *It Must Be Beautiful: Great Equations of Modern Science*. Londres : Granta Books.
- Foucart, S. (2006, 5 août). Louis Pouzin, l'homme qui n'a pas inventé Internet. *Le Monde*, p. 11.
- Framing (World Wide Web). (). *Wikipedia, The Free Encyclopaedia*.
[http://en.wikipedia.org/wiki/Framing_\(World_Wide_Web\)](http://en.wikipedia.org/wiki/Framing_(World_Wide_Web))
- Freud, S. (2003). *La question de l'analyse profane / Die Frage der Laienanalyse* (Trad. J. Altounian, A. et O. Bourguignon, P. Cotet & A. Rauzy). Paris : Gallimard. (Publication originale : 1926)
- Ganascia, J.-G. (1998). *Dictionnaire de l'informatique et des sciences de l'information*. Paris : Flammarion.
- Gauducheu, N., Cuisinier, F. & Garitte, C. (2008). Utilisation d'hypermédia et recherche d'information chez les enfants. In A. Chevalier & A. Tricot (Éd.), *Ergonomie des documents électroniques* (pp. 103-128). Paris : PUF.
- Gauvin, P. (2006). Droit à l'image et droit de l'image. *Savoirs CDI*.
<http://savoircdi.cndp.fr/rencontrelyon/gauvin/gauvin.htm#>
- Girardeau, A. (2007, 27 novembre). Critiqué, Google modifie ses conditions d'utilisation. *Écrans*.
<http://www.ecrans.fr/Critique-Google-modifie-ses,2678.html>
- Gizmo. (2008, 19 juillet). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Gizmo>
- Globish. (2008, 17 juillet). *Wikipédia, L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/Globish>
- Gobry, I. (2000). *Vocabulaire grec de la philosophie*. Paris : Ellipses.
- Goodfellow, M. & Wade, B. (2006-2007). The Digital Divide and First-Year Students. *Journal of College Student Retention: Research, Theory & Practice*, v8(4), 425-438.
<http://baywood.metapress.com/app/home/contribution.asp?referrer=parent&backto=issue,3,6;journal,7,36;linkingpublicationresults,1:300319,1>
- Google. (2008). Google Traduction – FAQ. *Google Traduction*.
http://www.google.com/intl/fr/help/faq_translation.html
- Google China. (2008, 26 août). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Google_China
- Google Team. (2004, 27 septembre). China, Google News and source inclusion. *The official Google blog*.
<http://googleblog.blogspot.com/2004/09/china-google-news-and-source-inclusion.html>

- Google Announces Plan To Destroy All Information It Can't Index. (2005, 31 août). *The Onion*.
<http://www.theonion.com/content/node/40076>
- Google (2004, 13 août). Amendment no. 7 to form S-1. Registration statement under the securities act of 1933. *Securities and Exchange Commission*.
<http://www.sec.gov/Archives/edgar/data/1288776/000119312504139655/ds1a.htm>
- Gookin, D. (2007). *Le PC pour les nuls*. Paris : First.
- Graham, L. & Metaxas, P. T. (2003). "Of course it's true; I saw it on the Internet!" Critical Thinking in the Internet Era. *Communications of the ACM*, 46(5), 71-75.
<http://www.wellesley.edu/CS/pmetaxas/CriticalThinking.pdf>
- Grangereau, P. & Guenneugues, L. (2006, 15 février). Chine : Google et Yahoo censurent gratis. *Libération*, pp. 2-3.
- Grefte, X. & Sonnac, N. (2008). *Culture Web. Création, contenus, économie numérique*. Paris : Dalloz.
- Hardy, Q. (2005, 14 novembre). Google Thinks Small. *Forbes.com*.
<http://members.forbes.com/global/2005/1114/054A.html>
- Harper, D. (2001). *Online Etymology Dictionary*. (2001). [dictionnaire en ligne]
<http://www.etymonline.com/abbr.php>
- Haski, P. (2008). *Internet et la Chine*. Paris : Éditions du Seuil / Presses de Sciences Po.
- Hassine, T. (s. d.). Shmoogles: No More Google Goggles!
<http://pzwart.wdka.hro.nl/mdma/alumni/2006/thassine/shmoogles.html/view>
- Hatzipolakis, A. P. (2000, 25 mars). [HM] Boyer's Law [Post 940].
<http://mathforum.org/kb/thread.jspa?forumID=149&threadID=382693&messageID=1179004>
- Hauben, M. (s. d.). *History of ARPANET. Part I: The history of ARPA leading up to the ARPANET*.
<http://www.dei.isep.ipp.pt/~acc/docs/arpa--1.html>
- Héritier, F. avec Broué, C. (2008). *L'identique et le différent*. La Tour d'Aigues : Éditions de l'Aube.
- Herrenschmidt, C. (2007). *Les trois écritures. Langue, nombre, code*. Paris : Gallimard.
- Hit (internet). (2008, 19 juillet). *Wikipedia, The Free Encyclopaedia*.
[http://en.wikipedia.org/wiki/Hit_\(internet\)](http://en.wikipedia.org/wiki/Hit_(internet))
- Hock, R. (2007). *The extreme searcher's Internet handbook*. Medford, NJ : CyberAge Books.
- Hondermarck, O. (2007). *JavaScript. Le guide complet*. Paris : Micro Application.

- Hornby, A. S. et al. (1963). *The Advanced Learner's Dictionary of Current English*. Oxford University Press.
- Hornyak, T. (2006a). *Loving the machine. The art and science of Japanese robots*. New York : Kodansha International.
- Hornyak, T. avec Skillings, J. (2006b). In Japan, robots are more than mere gadgetry—they're practically family. *CNET News*.
http://news.cnet.com/In-Japan,-robots-are-people,-too/2008-11386_3-6122761.html
- Hotlinking. (2008, 19 août). *Wikipédia, L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/Hotlinking>
- Howe, D. (1993-2008). *Free On-line Dictionary Of Computing (FOLDOC)*. [dictionnaire en ligne]
<http://foldoc.org/>
- HTTP 404. (2008, 7 septembre). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/404_error
- HTTP cookie. (2008, 8 septembre). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/HTTP_cookie
- Huitema, C. (1996). *Et Dieu créa l'Internet...* Paris : Eyrolles.
- Hyperlien. (2008, 14 mai). *Wikipédia, L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/Hyperlien>
- Hyperlink. (2008, 2 septembre). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Hyperlink>
- Hyppolite, J. (1966). Commentaire parlé sur la *Verneinung* de Freud. In J. Lacan (Éd.), *Écrits*, (pp. 879-887). Paris : Seuil.
<http://pagesperso-orange.fr/espace.freud/topos/psycha/psysem/hyppoli2.htm>
- ICANN. (2007, 1^{er} mars). Root server attack on 6 February 2007.
<http://www.icann.org/en/announcements/factsheet-dns-attack-08mar07.pdf>
- If Google ogles your e-mail, will Ashcroft be far behind? (2004, 16 avril). *San Jose Mercury News*.
<http://media.www.dailyvanguard.com/media/storage/paper941/news/2004/04/20/Opinion/If.Google.Ogles.Your.EMail.Will.Ashcroft.Be.Far.Behind-2612780.shtml>
- Imprimerie nationale. (2002). *Lexique des règles typographiques*. Paris : Imprimerie nationale.
- In God We Trust. (2008, 7 septembre). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/In_God_We_Trust
- Infoplease Dictionary*. (2008). [dictionnaire en ligne]

- <http://dictionary.infoplease.com/>
- Information Today. (2005). *WebSearch Academy. Mastering Internet Research*.
- <http://www.infotoday.com/Paris/WebSearch/>
- Information retrieval. (2008, 30 août). *Wikipedia, The Free Encyclopaedia*.
- http://en.wikipedia.org/wiki/Information_retrieval
- Information superhighway. (2008, 25 août). *Wikipedia, The Free Encyclopaedia*.
- http://en.wikipedia.org/wiki/Information_superhighway
- Informatique. (2008, 5 septembre). *Wikipédia, L'encyclopédie libre*.
- <http://fr.wikipedia.org/wiki/Informatique>
- Internet Archive. (2008, 3 septembre). *Wikipedia, The Free Encyclopaedia*.
- http://en.wikipedia.org/wiki/Internet_Archive
- Internet forum. (2008, 8 septembre). *Wikipedia, The Free Encyclopaedia*.
- http://en.wikipedia.org/wiki/Internet_forum
- Internets. (2008, 5 septembre). *Wikipedia, The Free Encyclopaedia*.
- [http://en.wikipedia.org/wiki/Internets_\(colloquialism\)](http://en.wikipedia.org/wiki/Internets_(colloquialism))
- Ippolita. (2008). *La face cachée de Google*. Paris : Éditions Payot & Rivages.
- Iteanu, O. (2008). *L'identité numérique en question*. Paris : Eyrolles.
- Janvresse, É. & de la Rue, T. (2008). Codage numérique de l'image. In B. Rittaud (Éd.), *Quand les maths se font discrètes* (pp. 53-81). Paris : Le Pommier.
- Jarraud, F. avec Davidenkoff, E. (2008, 5 août). *Apprendre avec les technologies de l'information et de la communication* [Fichier audio]. Site de France Info :
- http://www.france-info.com/spip.php?article165098&theme=81&sous_theme=319
- Jeanneney, J.-N. (2006). *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris : Fayard.
- Jeanneney, J.-N. (2007). *Google and the Myth of Universal Knowledge*. Chicago : The University of Chicago Press.
- Jennotte, A. (2008, 5 janvier). Fin janvier, Belgacom débranchera ses modems. *Le Soir*.
- <http://www.lesoir.be/actualite/economie/telecoms-la-concurrence-2008-01-05-569714.shtml>
- Judet de la Combe, P. & Wismann, H. (2004). *L'avenir des langues. Repenser les humanités*. Paris : Éditions du Cerf.
- Julia, D. (1981). *Les trois couleurs du tableau noir. La Révolution*. Paris : Belin.
- Kauder, S. (2004, 26 février). Droit à l'image et liberté d'expression ou contraintes et liberté. *Legalbiznext*.
- <http://www.legalbiznext.com/droit/Droit-a-l-image-et-liberte-d>

- KaZaA. (2008, 27 août). Wikipédia, *L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/KaZaA>
- Keen, A. (2007). *The cult of the amateur. How today's internet is killing our culture*. New York : Doubleday / Currency.
- Keen, A. (2008, 27 juillet). *Is the Internet killing the American reader?*
http://andrewkeen.typepad.com/the_great_seduction/2008/07/the-big-debate.html
- Kennedy, A.J. (1997). *Internet et le Web facile*. Paris : Éditions Mille et une nuits.
- Kintzler, C. (1996). *La République en questions*. Paris : Minerve.
- Klemperer, V. (1996). *LTI, la langue du III^e Reich. Carnets d'un philologue*. Paris : Albin Michel. (Édition originale en allemand, 1947)
- Koch, P. & Koch, S. (2008, 15 août). Yahoo! drops Boolean support. *Pandia Search Engine News*.
<http://www.pandia.com/sew/699-yahoo-drops-boolean-support.html>
- Lacan, J. (1966). *Écrits*. Paris : Seuil.
- Lacan, J. (2001). L'Étourdit. In J.-A. Miller (Éd.), *Autres écrits* (pp. 449-495). Paris : Seuil.
- Ladage, C. (2003). *Utilisation des nouvelles technologies dans un dispositif de formation d'un savoir labile*. Mémoire de DESS non publié, Université de Provence, Lambesc.
- Ladage, C. (2004). *Approche didactique de l'enseignement d'un savoir labile en technologies de l'information et de la communication*. Mémoire de DEA non publié, Université de Provence, Lambesc.
- Ladage, C. (2007a). Apprendre la recherche d'informations sur Internet à l'école : à la découverte du poids des mots. *Congrès international AREF 2007 (Strasbourg, 28-31 août 2007)*.
http://aref2007.u-strasbg.fr/actes_pdf/AREF2007_Caroline_LADAGE_113.pdf
- Ladage, C. (2007b). Les communautés virtuelles sur Internet : de l'apprentissage à la production collective de savoirs. Actes du colloque *Tice Méditerranée*, Marseille.
<http://isdms.univ-tln.fr/PDF/isdms29/LADAGE.pdf>.
- Ladage, C. (2007c). Internet virtual communities: from learning to writing knowledge. *Forms of democracy in education: open access and distance learning*. 4th International Conference in Open and Distance Learning. Athènes, Grèce.
- Ladage, C. & Ravestein, J. (2005). Apprentissage et enseignement sur des espaces de communication informels sur Internet : l'exemple du « référencement ». Colloque *Environnements Informatiques pour l'Apprentissage Humain*. Montpellier.
- Ladage, C. & Chevallard, Y. (2008, novembre). La place du portfolio dans la conception et l'implémentation d'une organisation didactique : problèmes ouverts. Communication

- présentée au colloque international « Efficacité & Équité en Éducation », Rennes, 19-21 novembre 2008.
- Lakhani, K. R. & McAfee, A. P. (2007). Wikipedia (A). *Harvard Business School*.
<http://courseware.hbs.edu/public/cases/wikipedia/index.html>
- Laloë, F. (2008, 28 janvier). Attention, l'humanité perd la mémoire. *Le Monde*.
- Lambrichs, L. L. (1993). *La vérité médicale. Claude Bernard, Louis Pasteur, Sigmund Freud : légendes et réalités de notre médecine*. Paris : Robert Laffont.
- Langville, A.N. & Meyer, C.D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton : Princeton University Press.
- Larrieu, J. (2005). *Droit de l'Internet*. Paris : Ellipses.
- Lashinsky, A. (2007, 2 février). Yahoo's mission quest. *Fortune*.
http://money.cnn.com/2007/02/01/technology/pluggedin_lashinsky_yahoo.fortune/index.htm
- Le langage HTML/Liens. (2008, 19 avril). *Wikibooks*.
http://fr.wikibooks.org/wiki/Le_langage_HTML/Liens
- Le Roy Ladurie, E. (1982). *Paris-Montpellier. PC-PSU 1945-1963*. Paris : Gallimard
- Lecourt, D. (2008). *Georges Canguilhem*. Paris : PUF.
- Lemainque, F. (2008). *HTML*. Paris : Micro Application.
- Lenormand, P. (2007). *Internet. Techniques de recherche pour les professionnels*. Paris : ENI.
- Leser, É. (2004, 14 mai). La bataille entre Google, Microsoft et Yahoo ! est lancée. *Le Monde*.
- Levine, J.R., Reinhold, A. & Levine Young, M. (2002). *The Internet for Dummies* (8^e éd.). New York : Wiley Publishing.
- Levine, J. R., Baroudi, C. & Levine Young, M. (2006). *Internet pour les nuls*. Paris : First Interactive.
- Library Science. (2008, 31 août). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Library_science
- Lilen, H. (2007). *Windows Vista et Internet facile*. Paris : First.
- Lilen, H. (2008). *Le livre de l'Internet*. Paris : First.
- Littre, É. (1872-1877). *Dictionnaire de la langue française*.
<http://francois.gannaz.free.fr/Litre/accueil.php>
- Liu, B. (2007). *Web data mining. Exploring hyperlinks, contents, and usage data*. New York : Springer.
- Lubrano, P. (2003). Les adresses IPv4. *Afnic*.
http://www.afnic.fr/noncvs/formations/ipv4_court/ipv4.pdf
- Machine translation. (2008, 3 septembre). *Wikipedia, The Free Encyclopaedia*.

- http://en.wikipedia.org/wiki/Machine_translation
- Maigné, C. (2007). *Johann Friedrich Herbart*. Paris : Belin.
- Mansion, J. E. (1980). *Harrap's new standard French and English dictionary*. (4 vols).
Londres : Harrap.
- Marcelle, P. (2008a, 31 mai). Google limites. *Le Monde*.
<http://www.ecrans.fr/Google-limites,4225.html>
- Marcelle, P. (2008b, 7 juin). Google, no limits ?. *Le Monde*.
<http://www.ecrans.fr/Google-no-limits,4293.html>
- Markup language. (2008, 19 février). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Markup_language
- Mathematical Folklore. (2008, 14 juillet). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Mathematical_folklore
- Mathieu, B. (2005, 3 septembre). Un défi pour les dictionnaires Larousse ou Robert. *Le Monde*.
- Mauss, M. (1950a). *Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques*. In G. Gurvitch (Éd.), *Marcel Mauss. Sociologie et anthropologie* (pp. 143-279). Paris : PUF. (Publication originale : 1923-1924)
- Mauss, M. (1930). Les civilisations. Éléments et formes. In Centre International de Synthèse (Éd.), *Civilisation. Le mot et l'idée* (pp. 82-104). Paris : La Renaissance du Livre.
http://classiques.uqac.ca/classiques/febvre_lucien/civilisation/civilisation_idée.pdf
- Mauss, M. (1950b). *Les techniques du corps*. In G. Gurvitch (Éd.), *Marcel Mauss. Sociologie et anthropologie* (pp. 363-386). Paris : PUF. (Publication originale : 1936)
- McGee, V. E. (1975). *Principes de statistiques. Approche traditionnelle et approche bayésienne*. Paris : Vuibert.
- MCI Communications. (2008, 2 septembre). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/MCI_Communications
- Merriam-Webster's Online Dictionary*. (2003). [dictionnaire en ligne]
<http://www.merriam-webster.com/>
- Mesguich, V. & Thomas, A. (2006a). *Net recherche. Le guide pratique pour mieux trouver l'information utile*. Paris : ADBS éditions.
- Mesguich, V. & Thomas, A. avec Filippone, D. (2006b, 1^{er} juin). « Au moins 60 milliards de pages dont un tiers recensées par les moteurs ». *Le Journal du Net*.
<http://www.journaldunet.com/imprimer/solutions/chat/retrans/060601-mesguich-thomas.shtml>

- Miller, J. et al. (2008, 11 août). Earliest Known Uses of Some of the Words of Mathematics (H).
<http://members.aol.com/jeff570/h.html>
- Mission statement. (2007, 19 septembre). *Wikipedia, The Free Encyclopaedia*.
http://en.wikipedia.org/wiki/Mission_statement
- Modem. (2008, 8 janvier). *Wikipédia, L'encyclopédie libre*.
<http://fr.wikipedia.org/wiki/Modem>
- Modem. (2008, 7 septembre). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Modem>
- Modulation du signal. (2008, 14 juin). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Modulation_du_signal
- Montoliu, C. (1998, mai-juin). La traducción automática en *El Periódico de Catalunya*. *Puntoycoma*, 51.
<http://ec.europa.eu/translation/bulletins/puntoycoma/51/periodico.htm>
- Morel, C. (2007). *L'enfer de l'information ordinaire*. Paris : Gallimard.
- Morin, E. (1999). *La tête bien faite. Repenser la réforme. Réformer la pensée*. Paris : Seuil.
- Mottez, D. & Guggémos, A. (2008, 11 août). Google, mode d'emploi. *Femme Actuelle*.fr
<http://www.femmeactuelle.fr/multimedia/decryptage/tout-trouver-sur-google-00648>
- Nelson, T. (1965). A File Structure for the Complex, the Changing and the Indeterminate. *Proceedings of the 1965 20th national conference* (pp. 84-100). New York : ACM Press.
- Nelson, T. (1970, septembre). No More Teachers' Dirty Looks. *Computer Decisions*, 9(8), 16-23.
- Nelson, B. & Economy, P. (2007). *Le management pour les nuls*. Paris : First.
- Nerrière, J.-P., Dufresne, P. & Bourgon, J. (2005). *Découvrez le globish*. Paris : Eyrolles.
- Nietzsche, F. (2000). *Mauvaises pensées choisies* (choix établi par Georges Liébert). Paris : Gallimard.
- Night Watch (painting). (2008, 3 septembre). *Wikipedia, The Free Encyclopaedia*.
[http://en.wikipedia.org/wiki/Night_Watch_\(painting\)](http://en.wikipedia.org/wiki/Night_Watch_(painting))
- Notess, G. R. (2006). *Teaching Web Search Skills. Techniques and Strategies of Top Trainers*. Medford, NJ : Information Today.
- Nourissier, F. (2008). *Vrac*. In J. Chessex, *Le simple préserve l'énigme* (pp. 9-17). Paris : Gallimard.
- OneLook Dictionary Search*. [méta-dictionnaire en ligne]
<http://www.onelook.com/>

- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
<http://www.psych.upenn.edu/history/orne/>
- Owen, D. (s. d.). The Pros & Cons of Using Frames in Web Pages. *MediaCollege*.
<http://www.mediacollege.com/internet/html/frames/pros-cons.html>
- Page, L., Brin, S., Motwani, R. & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*.
<http://dbpubs.stanford.edu:8090/pub/1999-66>
- Page, L. & Brin, S. avec Sheff, D. (2004, septembre). "Google guys". *Playboy*.
<http://www.google-watch.org/playboy.html>
- Page, L. & Brin, S. (2005). *Google Annual Report 2005*.
http://investor.google.com/pdf/2005_Google_AnnualReport.pdf
- Pain, J. avec Roussel, F. (2006, 15 février). « La Chine est de plus en plus efficace à censurer ». *Libération*, p. 3.
- Pan (Mythology). (2008, 5 septembre). *Wikipedia, The Free Encyclopaedia*.
[http://en.wikipedia.org/wiki/Pan_\(mythology\)](http://en.wikipedia.org/wiki/Pan_(mythology))
- Pastor-Satorras, R. & Vespignani, A. (2004). *Internet. Structure et évolution*. Paris : Belin.
- Pena-Ruiz, H. (2005). *Qu'est-ce que l'école ?* Paris : Gallimard.
- Piaget, J. (1968). *Sagesse et illusions de la philosophie*. Paris : PUF.
- Picket, J. P. et al. (2000). *The American Heritage Dictionary of the English Language* (4^e éd.). Boston : Houghton Mifflin.
<http://www.bartleby.com/am/>
- Pillou, J.-F. (2005). *Tout sur le Webmastering*. Paris : Dunod.
- Piquard, A. (2006, 21 avril). Des bureaux d'enregistrement dénoncent le "scandale" des noms de domaine en .eu. *Le Monde*.
- Podcast. (2008, 6 septembre). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Podcast>
- Pomeroy, A. M. (2002). *Mastering search engines*. Placentia, CA : Creative continuum.
- Pouget, R. (2002, 18 mars). La fragilité du témoignage. *Académie des sciences et lettres de Montpellier*.
http://www.biu-montpellier.fr/academie/academie_edition/fichiers_conf/Pouget2002.pdf
- Pouzin, L. avec Bellin, I. (2007, 19 mars). Louis Pouzin, la tête dans les réseaux. *Interstices*.
http://interstices.info/jcms/c_16645/louis-pouzin-la-tete-dans-les-reseaux

- Price, G. & Sherman, C. (2001). *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. Medford, NJ : Information Today.
- Proust, M. (1927). *Le Temps retrouvé*. Paris : Gallimard.
<http://gallica.bnf.fr/proust/>
- Quaternion. (2008, 7 septembre). Wikipédia, L'encyclopédie libre.
<http://fr.wikipedia.org/wiki/Quaternion>
- Rajoson, L. (1988). *L'analyse écologique des conditions et des contraintes dans l'étude des phénomènes de transposition didactique. Trois études de cas*. Thèse de doctorat non publiée, Université d'Aix-Marseille II – Luminy.
- Ravestain, J. & Ladage, C. (2006). Pour une didactique de l'usage d'Internet. In A. Piolat (Éd.). *Lire, Ecrire, Communiquer et Apprendre avec Internet* (pp. 489-507). Marseille : Solal.
- Ravestain, J., Ladage, C. & Johsua, S. (2007). Trouver et utiliser des informations sur Internet à l'école. Problèmes techniques et questions éthiques. *Revue Française de Pédagogie*, 158, 71-81.
- Ray, C. & Poizat, J.-C. (2007). *La physique par les objets quotidiens*. Paris : Belin.
- Raya, F., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J. et al. (1999). A Visit to the Information Mail: Web Searching Behavior of High School Students. *Journal of the American Society for Information Science*, 50(1), 24-37.
- Reber, A. S. (1985). *Dictionary of psychology*. New York : Penguin Books.
- Reliability of Wikipedia. (2008, 7 septembre). Wikipedia, The Free Encyclopaedia.
http://en.wikipedia.org/wiki/Reliability_of_Wikipedia
- Reitz, J. M. (2007). *Online Dictionary for Library and Information Science*.
<http://lu.com/odlis/about.cfm>
- Rey, A. et al. (Éd.). (1993). *Dictionnaire historique de la langue française*. Paris : Dictionnaires Le Robert.
- Rey, A. et al. (Éd.). (2005). *Dictionnaire culturel en langue française*. Paris : Dictionnaires Le Robert – SEJER.
- Ricard, P. (2007, 30 novembre). Berlin se rallie in extremis au plan de relance du système de navigation par satellite Galileo. *Le Monde*, p. 8.
<http://www.lemonde.fr/web/article/0,1-0@2-3214,36-983894,0.html?xtor=RSS-3244>
- Robin, M.-M. (2008a), *Le monde selon Monsanto. De la dioxine aux OGM, une multinationale qui vous veut du bien*. Paris : La Découverte / Arte Éditions.
- Robin, M.-M. (2008b), *Le monde selon Monsanto. De la dioxine aux OGM, une multinationale qui vous veut du bien*. Paris : France Info / Arte Vidéo.

- Roland Moreno. (2008, 4 août). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Roland_Moreno
- Rothkopf, E.Z. (1970). The concept of mathemagenic activities. *Review of educational research*, 40(3), 325-336.
- RSS. (2008, 21 juillet). *Wikipedia, The Free Encyclopaedia*.
[http://en.wikipedia.org/wiki/RSS_\(file_format\)](http://en.wikipedia.org/wiki/RSS_(file_format))
- Rudenstine, N. (1996, 29 mai). The Internet is changing Higher Education. *U.S. Information Agency Electronic Journals*, 1(12).
<http://usinfo.state.gov/journals/itgic/0996/ijge/gjcom2.htm>
- Salton, G., Wong, A. & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.
<http://www.cs.umbc.edu/~nicholas/676/p613-salton.pdf>
- Schraefel, M. C., Carr, L., De Roure, D. & Hall, W. (2004). You've Got Hypertext. *Journal of Digital Information*, 5(1), article n° 253.
<http://jodi.tamu.edu/Articles/v05/i01/schraefel/>
- Screencast. (2008, 28 juillet). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Screencast>
- Serres, A. (2007). Maîtrise de l'information : la question didactique. *Les Dossiers de l'ingénierie éducative*, 67, 58-62.
- Serveur racine du DNS. (2008, 14 août). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Serveurs_DNS_Racine
- Sherriff, L. (2005, 10 octobre). We'll index the world by 2310, says Google. *The Register*.
http://www.theregister.co.uk/2005/10/10/google_index/
- Siracusa, J. (2008). *Vacances sociologiques. Enseigner la sociologie à l'université*. Saint-Denis : PUV.
- Somme de contrôle. (2008, 21 août). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Somme_de_contr%C3%B4le
- Spink, A, Wolfram, D., Jansen, B. J. & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226-234.
<http://www.scils.rutgers.edu/~tefko/JASIST2001.pdf>
- Spinoza, B. (2005, 29 août). *Traité politique. I. Introduction. Wikisource*. (Publication originale : 1677)
[http://fr.wikisource.org/wiki/Trait%C3%A9_politique_\(bilingue\)_I._Introduction](http://fr.wikisource.org/wiki/Trait%C3%A9_politique_(bilingue)_I._Introduction)
- Stephen, M. (2003). *Basic Computer Skills*. Londres : Hodder Education.

- Stevens, M. A. et al. (1998). *Merriam-Webster's Concise Handbook for Writers* (2^e éd.). Springfield, MA : Merriam-Webster.
- Stevenson, A. et al. (2002). *Harrap's Media & Communication Dictionnaire Anglais-Français / Français-Anglais*. Edimbourg : Chambers Harrap.
- Strand, G. (2008, mars). Google's addiction to cheap electricity. *Harper's Magazine*, 1894, 64-65.
- Sullivan, D. (2001a, 8 mars). InFind Meta Search Closes Temporarily, Perhaps Permanently. *SearchEngineWatch.com*.
<http://searchenginewatch.com/showPage.html?page=2163531>
- Sullivan, D. (2001b, 26 octobre). Boolean Searching. *SearchEngineWatch.com*.
<http://searchenginewatch.com/showPage.html?page=2155991>
- Sullivan, D. (2001c, 26 octobre). Search Engine Math. *SearchEngineWatch.com*.
<http://searchenginewatch.com/showPage.html?page=2156021>
- Sullivan, D. (2006, 17 avril). My decade of writing about search engines. *SearchEngineWatch.com*.
<http://blog.searchenginewatch.com/blog/060417-130526>
- Sullivan, D. (2007, 26 avril). What Is Google PageRank? A Guide For Searchers & Webmasters. *Search engine land*.
<http://searchengineland.com/070426-011828.php>
- Sullivan, D. (2008, 28 juillet). Cuil launches—Can this search start-up really best Google? *Search engine land*.
<http://searchengineland.com/080728-000100.php>
- Sublation. (2008, 17 mai). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/Sublation>
- Tarot, C. (1999). *De Durkheim à Mauss, l'invention du symbolique*. Paris : La Découverte/MAUSS.
- Teevan, J. Adar, E., Jones, R. & Potts, M. (2006). History repeats itself: Repeat Queries in Yahoo's query logs. *Proceedings of the 29th Annual ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 703-704). New York : The Association for Computing Machinery.
<http://people.csail.mit.edu/teevan/work/publications/posters/sigir06.pdf>
- Tesnière, V. & Lesquins, N. (2006, mars). La bibliothèque numérique européenne. Une stratégie culturelle de la Toile. *Bulletin des bibliothèques de France*, 51(3), 68-80.
<http://bbf.enssib.fr/sdx/BBF/frontoffice/2006/03/document.xsp?id=bbf-2006-03-0068-012/2006/03/fam-dossier/dossier&statutMaitre=non&statutFils=non>

- Theilheimer, W. (2006, 14 mars). *Mathemagenic Processing and Ernie Rothkopf*.
http://www.willatworklearning.com/2006/03/mathemagenic_pr.html
- Tiffany, P. & Peterson, S. D. (2005). *Business plans pour les nuls*. Paris : First.
- Todorov, T. (2006). *L'Esprit des Lumières*. Paris : Robert Laffont.
- Toobin, J. (2007, février). Google's Moon Shot. The quest for the universal library. *The New Yorker*.
http://www.newyorker.com/reporting/2007/02/05/070205fa_fact_toobin
- Google-moi : la deuxième mission de l'Amérique : un nouveau livre sur Google. (2007, 23 mars). *Trackbusters*.
<http://www.trackbusters.fr/actualites/20070323-livre-sur-google-barbara-cassin-google-moi.html>
- Uniform Resource Locator. (2007, 19 octobre). *Wikipedia, The Free Encyclopaedia*.
<http://en.wikipedia.org/wiki/URL>
- Uniform Resource Locator. (2008, 8 mai). *Wikipédia, L'encyclopédie libre*.
http://fr.wikipedia.org/wiki/Uniform_Resource_Locator
- Validation des items B2I par les documentalistes. (2008, 20 juin). [Questionnaire_B2I.doc]
<http://www.lyc-erea-toulouse-lautrec-vaucresson.ac-versailles.fr/IMG/doc/>
- van Andel, P. (2005). Sérendipité, ou de l'art de faire des trouvailles. *Automates intelligents*, 61.
<http://www.admiroutes.asso.fr/larevue/2005/61/serendipite.htm>
- Van Lancker, L. (2008). *JavaScript. Introduction et notions fondamentales*. Paris : ENI.
- Vernant, J.-P. & Detienne, M. (1974). *Les ruses de l'intelligence. La mètis des Grecs*. Paris : Flammarion.
- Veyne, P. (2008). *Foucault, sa pensée, sa personne*. Paris : Albin Michel.
- Vine, R. (2004). Going Beyond Search Engines for Faster and Smarter Web Searching. Workingfaster.com
http://sp.searchportfolio.com/white_paper_1.pdf
- Virieux, F. (2004). *Comment ça marche Internet ?* Paris : Le Pommier.
- Virieux, F. & Chebret, S. (2006). *Internet : quel drôle de réseau !* Paris : Le Pommier.
- Vise, D. & Malseed, M. (2006). *The Google Story*. New York : Delta Trade Paperbacks.
- Vreeman, R. C. & Carroll, A. E. (2007, décembre). Medical myths. *British Medical Journal*, 335, 1288-1289.
<http://www.bmj.com/cgi/content/full/335/7633/1288>
- Walter, E. et al. (2006). *Cambridge Advanced Learner's Dictionary* (3^e éd.) Cambridge : Cambridge University Press.

<http://dictionary.cambridge.org/define.asp?key=55931&dict=CALD>

Warbesson, K. (2007). *Créez votre site web*. Paris : Micro Application.

Wardrip-Fruin, N. (2004). What Hypertext Is. *Proceedings of the 15th ACM conference on Hypertext & hypermedia*. New York : ACM Press.

<http://www.noahwf.com/texts/whatHypertextIs.pdf>

Watkins, K. (2008). Return of the Google Game: More Fun Ideas to Transform Students into Skilled Researchers. *School Library Journal*, 54(5), 46-48.

Weinberger, D. (2007). *Everything is Miscellaneous. The Power of the New Digital Disorder*. New York : Henry Holt and Company.

Wiegandt, C. avec *Captain Doc* (2005, 14 décembre). « Un métier qui évolue très vite doit prendre le temps de la réflexion sur son positionnement nouveau ». *Captain Doc*, 44.

Williams, G. (2007, juin). Unclear on the Context: Refocusing on Information Literacy's Evaluative Component in the Age of Google. *Library Philosophy and Practice Special Issue: Libraries and Google*.

<http://www.webpages.uidaho.edu/~mbolin/williams.htm>

Williams, R. (1983). *Keywords. A vocabulary of culture and society*. Londres : Fontana Paperbacks.

Williamson, E. (2008, 1^{er} août). *Google nous rend-il cons ?* [post] Voir http://www.lepost.fr/article/2008/07/29/1233121_google-nous-rend-il-cons.html

Weisstein, E. W. (1999-2008). Hypercomplex Number. *MathWorld*.

<http://mathworld.wolfram.com/HypercomplexNumber.html>

Wiktionary (2008). [dictionnaire en ligne]

<http://www.wiktionary.org/>

Witten, I. H., Gori, M. & Numerico, T. (2006). *Web Dragons. Inside the Myths of Search Engine Technology*. San Francisco : Morgan Kaufmann.

World Wide Web. (2008, 24 avril). *Wikipedia, The Free Encyclopaedia*.

http://en.wikipedia.org/wiki/World_Wide_Web

World Wide Web. (2008, 30 avril). *Wikipédia, L'encyclopédie libre*.

http://fr.wikipedia.org/wiki/World_Wide_Web

Yahoo!. (2008, 8 septembre). *Wikipedia, The Free Encyclopaedia*.

<http://en.wikipedia.org/wiki/Yahoo!>

Yahoo! Directory. (2008, 1^{er} juillet). *Wikipedia, The Free Encyclopaedia*.

http://en.wikipedia.org/wiki/Yahoo%21_Directory

Yahoo! Search. (2008, 28 août). *Wikipedia, The Free Encyclopaedia*.

http://en.wikipedia.org/wiki/Yahoo!_Search

Yi-Wen, T., Meilun, S. & Chin-Chung, T. (2008). Eighth Graders' Web Searching Strategies and Outcomes: The Role of Task Types, Web Experiences and Epistemological Beliefs. *Computers & Education*, 51(3), 1142-1153.

