



HAL
open science

Les liens hypertextuels ou Abondance de liens ne nuit pas

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Les liens hypertextuels ou Abondance de liens ne nuit pas. D. Piotrowski. Lexicographie et Informatique. Autour de l'informatisation du Trésor de la Langue Française, Didier Erudition, pp.299-318, 1996. <hal-01575428>

HAL Id: hal-01575428

<https://hal.science/hal-01575428v1>

Submitted on 19 Aug 2017

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.

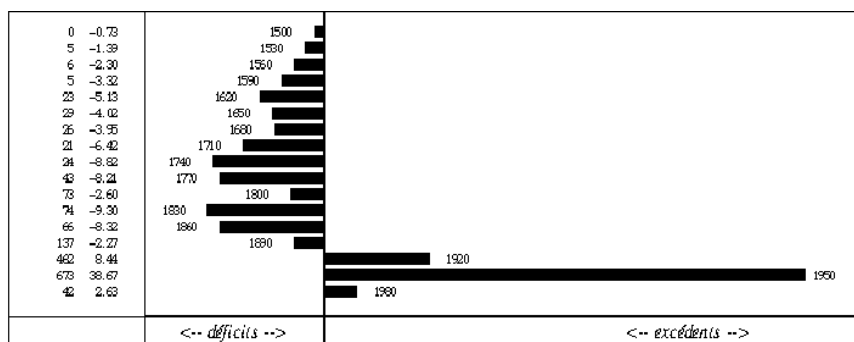


HAL Authorization

Les liens hypertextuels ou Abondance de liens ne nuit pas

Etienne Brunet

Le concept d'hypertexte – ou tout au moins le mot – remonte à 1980. On le doit à un visionnaire, Ted Nelson, dont on ne sait s'il a anticipé la mode ou s'il l'a lancée. Le langage fait une grande consommation des préfixes intensifs et l'on aurait pu penser que l'essor du préfixe HYPER s'essoufflerait à la longue et que le succès irait à quelque autre rival comme SUPRA, EXTRA, ULTRA, META, TRANS ou SUPER. Il n'en est rien. HYPER est au zénith et son succès se confirme dans les dernières tranches de *Frantext*, comme en témoigne l'histogramme ci-dessous.



Préfixe HYPER (1709 occur.). Courbe de l'évolution.
(17 périodes prises en compte. Taille du corpus : 169981066 occurrences)

Graphique 1. L'évolution du préfixe HYPER à travers Frantext

À vrai dire, l'hypertexte n'a aucune influence dans cette distribution, puisqu'on n'en rencontre aucun exemple dans *Frantext*. Au reste, la dernière tranche, qui commence en 1980 et n'a pu aller jusqu'à son terme, contient encore trop peu de textes pour qu'on en tire des conclusions. Encore faudrait-il que ces textes aient un rapport avec le domaine de l'informatique où l'hypertexte a fait son nid et couvé ses petits, français ou anglais : *hypertextuel*, *hypermédia*, *hyperlien*, *hyperdocument*, *hypercard*, *hypertalk*, *hyperbase*, *hyperlink*, *hyperdoc*,

hypertool, etc. Il y a gros à parier qu'on verra bientôt naître un gros *Hypernet* pour désigner l'*Internet* de demain, plus puissant et plus rapide.

La définition de l'hypertexte est simple. On désigne par là un système documentaire doté de liens qui permettent de naviguer d'une information à l'autre. Il n'y a rien là qui soit très nouveau. Depuis toujours, le cerveau humain s'attache à fixer des repères dans le flux des connaissances qu'il acquiert ou des expériences qu'il traverse. Tout cela constitue une forêt de symboles, de rappels, d'échos, un maquis inextricable de lianes et de liens. Et tout l'appareillage que les traditions documentaires ont construit constitue un vaste nœud gordien où s'entrecroisent les notes, les renvois, les commentaires, les index, les bibliographies, les glossaires, les concordances, les dictionnaires, les encyclopédies. Dès qu'on note une référence ou un rapprochement, dès qu'on remplit une fiche, qu'on insère un signet ou qu'on maintient le doigt entre deux pages, on participe à la démarche hypertextuelle. À l'enregistrement linéaire et horizontal de la lecture, la conscience superpose en effet la grille verticale de la mémoire.

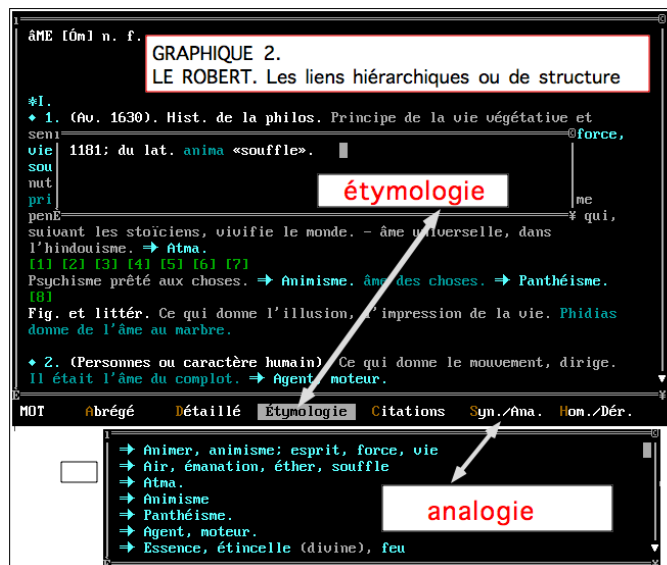
Les fils de liaison sont plus nombreux encore dans la mémoire de l'ordinateur, où l'on ne trouve guère que des adresses, des pointeurs, et mille rebonds où le signal s'agite comme la boule dans l'urne du loto. Nul moyen que l'information échappe à ces synapses tentaculaires. Pendant longtemps, la mémoire centrale étant limitée en capacité, on s'est ingénié à ne jamais y dédoubler les grains d'information et à multiplier par contre les voies ou liaisons d'accès, l'idéal implicite étant celui d'une grande bibliothèque dont tous les livres seraient en exemplaire unique, tous étant pareillement accessibles. Les bases de données relationnelles, les réseaux et jusqu'à *Internet* sont fondés sur cet impératif économique qui proscriit les doublons et le gaspillage de l'espace et spéculé sur la rapidité – parfois illusoire – des transmissions.

-I-

Ces considérations générales ne réduisent pas pourtant l'actualité et la spécificité de l'hypertexte, si en vogue aujourd'hui. Les moyens nouveaux de stockage et de diffusion, qui permettent l'accès direct à n'importe quelle information et son affichage immédiat, offrent à l'utilisateur la liberté interactive de circulation, avec des itinéraires variés et une signalisation précise – ce que recouvre la notion d'hypertexte.

1- Ces avantages sont sensibles dans *Le Robert électronique* qui est pourtant le CD-Rom le moins révolutionnaire qui soit, ayant été conçu à

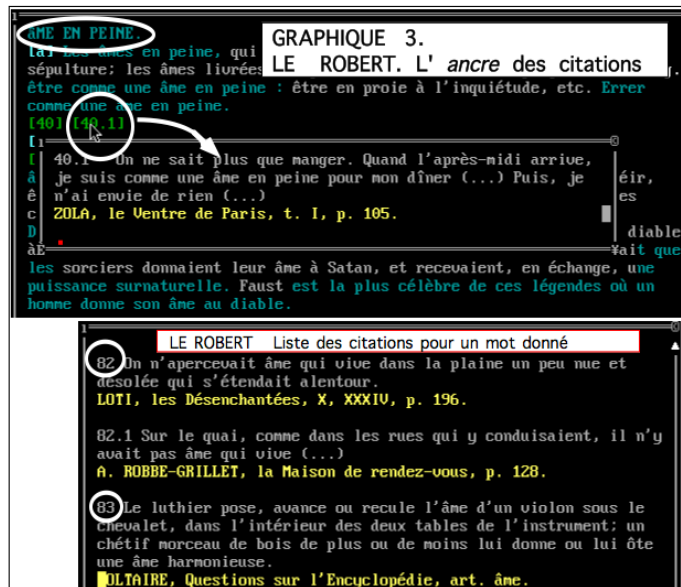
l'image du dictionnaire papier. Le contenu des articles est emprunté sans changement au *Grand Robert*. L'organisation interne en est seulement rendue plus visible non seulement par le jeu des symboles et de la couleur, mais aussi par l'alternative entre présentation abrégée ou détaillée et par le choix de la rubrique affichée (définition, étymologie, citations, analogie ou dérivation). Les liens mis en œuvre sont ici *hiérarchiques* (voir figure 2). Ils définissent la *structure* et renvoient aux différents éléments de l'article. Cela répond à une nécessité ergonomique, vu l'étroitesse de l'écran qui n'offre pas la même surface déployée qu'un livre ouvert de grand format (sans compter une définition très inférieure à celle des **photocomposeuses**). Le lecteur du *Grand Robert* bénéficie d'une vision globale et panoramique qui contraste avec la myopie inhérente aux écrans ordinaires¹. Le CD-Rom ne présente en effet qu'une colonne, qui est d'ailleurs souvent incomplète et qu'on doit faire défiler.



2- Mais quand un dictionnaire préexiste à son traitement informatique (c'est la situation commune au *Robert*, au *TLF* et à bien d'autres), il y a chance que la structure hiérarchique n'y montre pas la cohérence et la constance qu'on exige d'une base de données. En

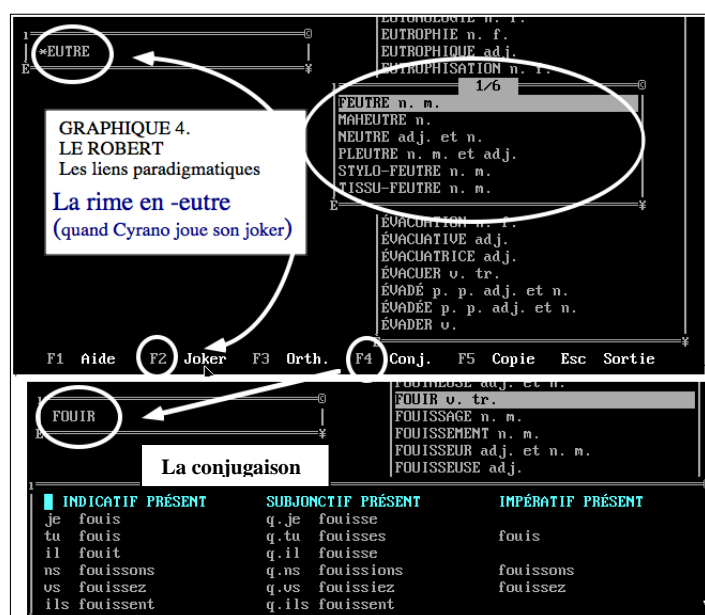
¹ Les écrans d'information ont la même étroitesse dans *Internet*, et plus encore dans le *Minitel*, et l'habitude s'est créée chez les producteurs de fragmenter le contenu en morceaux de la taille d'un écran et d'annoncer l'ensemble dans un sommaire initial, pourvu de flèches hiérarchiques.

particulier, les citations sont disséminées tout au long de l'article, et c'est par un artifice un peu brutal qu'on peut les rassembler dans la même fenêtre en sollicitant le menu *Citations* au bas de l'écran (figure 2). Il est cependant possible de lire la citation *in situ* en sollicitant le renvoi numérique qui apparaît entre crochets dans le texte de la définition. Ce lien est analogue à celui qui rattache traditionnellement la note à l'appel de note. Sa mise en œuvre dans un produit informatique se traduit par une fenêtre superposée à l'endroit exact où la citation a sa place dans le texte d'origine. L'effet est celui d'une sorte de *zoom* qui explicite et développe le signal convenu. Cette technique de l'*ancrage* a été généralisée dans les écrans générés par *Web* et gouvernés par le langage HTML, comme on verra plus loin.



3- Ces liens, internes à l'article, n'ont qu'une portée restreinte. À l'échelle du dictionnaire, il existe un ordonnancement *alphabétique* des entrées, qui relie chacune à la précédente et à la suivante dans la séquence. L'esprit humain est habitué à cet ordre qui remonte à l'origine de l'écriture et que la machine reproduit sans problème dans ses index. La manipulation de tels index permet en outre de neutraliser le début ou la fin des mots, ou telle portion du mot que l'on voudra. Le rôle des jokers est ainsi d'isoler un préfixe, un radical, un suffixe ou une chaîne de caractères quelconque. Si Cyrano improvisant la fameuse ballade du duel avait pu consulter le *Robert électronique* tout en ferrailant, il ne lui aurait

pas manqué la quatrième rime en -EUTRE dont il avait besoin (voir le haut de la figure 4). Dans le cas des verbes, le lien *paradigmatique* est établi entre le radical et les désinences canoniques (voir figure 4, en bas). En réalité, ces ressources de filtrage, de masquage ou de conjugaison sont communes à la plupart des systèmes documentaires et on les retrouve par exemple dans *Frantext*. Sans constituer le moins du monde une innovation, cette facilité offerte par le *Robert électronique* n'en représente pas moins un progrès par rapport à la version papier.



4- Un autre progrès, plus décisif, est apporté par les liens de *croisement* (en anglais *cross reference*), qui permettent, à la faveur d'un mot présent sur l'écran, de se détourner de l'entrée affichée pour en rejoindre une autre. Cette possibilité de bifurcation est généralisée à tous les mots affichés, qui sont tous accessibles au curseur et à la sélection, qu'ils appartiennent à la définition, à la citation ou aux champs analogiques ou dérivationnels. Si on se laisse distraire, le parcours discontinu peut se prolonger à l'infini et, à l'occasion, mettre en lumière la fameuse circularité des définitions lexicographiques. L'exemple architectural de la figure 5 montre le bénéfice qu'on peut attendre de cette circulation rapide à l'intérieur du dictionnaire, le sens d'une entrée se précisant en présence d'une seconde entrée associée. Ainsi la seconde

acception du mot ABAQUE s'éclaire au contact de l'ARCHITRAVE. Cette pratique de la consultation en cascade n'est pas étrangère au dictionnaire papier et combien de voyages imaginaires n'y a-t-on pas faits parmi les mots et les choses, jusqu'à oublier parfois le point de départ. Mais la manipulation de tomes différents est lourde et décourageante, au lieu que l'invite du CD-Rom est alerte et séduisante. Celle du *Robert électronique* est même trop séduisante, en ce sens qu'elle peut égarer l'utilisateur et lui faire perdre sa route. Il manque en effet à la consultation du *Robert* un espion électronique qui repère les lieux parcourus et permette le retour en arrière. Ces liens *historiques*, qui relient les étapes d'un parcours, font partie intégrante de tout voyage sur l'*Internet*. En particulier la consultation du *Web* par *Mosaic* ou *Netscape* imite la démarche du petit Poucet, et dépose des cailloux à chaque détour du chemin. Le saut de l'un à l'autre, en avant ou en arrière, se fait instantanément comme si l'on disposait de bottes de sept lieues.

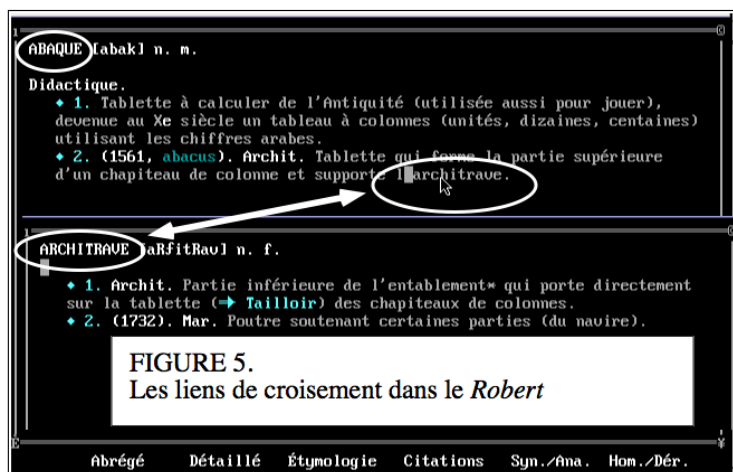


FIGURE 5.
Les liens de croisement dans le *Robert*

-II-

1- Les liaisons historiques² ne font pas défaut par contre à l'*Oxford English Dictionary*, même si l'on aurait pu souhaiter une mise en œuvre

² L'OED offre une aide en ligne dotée de tous les perfectionnements souhaitables. La consultation se fait par mots-clés dont la distribution peut suivre l'ordre alphabétique, ou la séquence structurelle, ou encore les relations analogiques. Les parcours que l'utilisateur entreprend pour combler ses lacunes ou ses incertitudes laissent une trace visible qu'on peut emprunter à rebours.

2- La supériorité éclatante de l'OED ne réside pas dans le simple affichage du dictionnaire papier, ni dans les possibilités d'atteindre sélectivement telle ou telle entrée de la nomenclature. L'avantage décisif de ce produit tient à ses capacités relationnelles, trop timidement développées dans le *Robert*. L'OED constitue une véritable base de données structurée, qui certes donne accès à chacun de ses enregistrements, mais autorise aussi des requêtes générales qui embrassent l'ensemble de la base. Tandis que le *Robert* peut seulement conduire l'utilisateur à l'article VICE (et à ses homonymes), l'OED sait reconnaître les entrées où le VICE s'est introduit sous une forme ou sous une autre. Mieux même, il peut dénoncer le VICE dans l'ensemble du texte, ou seulement dans la définition, ou dans les citations, voire même dans l'étymologie. À vrai dire, le VICE est si répandu, même en Angleterre, que la machine pourrait reculer devant ses débordements. Mais la digue des 8 000 occurrences n'est pas rompue et la machine restitue sans broncher les 1 785 contextes concernés. En limitant la recherche aux citations empruntées à Oscar Wilde, on verra sur l'écran les 117 contextes où cet auteur parle du vice en connaisseur. Si l'on propose un mot moins courant, comme EROTIC, la base entière sera explorée en quelques secondes pour livrer le résultat de la figure 7. Les 120 contextes qui contiennent ce mot sont alors restitués dans la présentation *kwic* devenue très courante. À chacun, une ligne est réservée qui est sensible au clic de la souris, en ouvrant une fenêtre sur le dictionnaire.

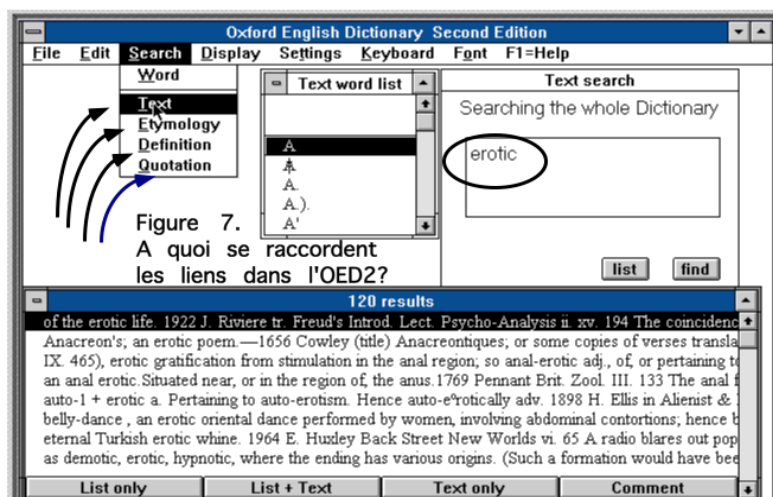


Figure 7

3- On nous pardonnera de ne pas insister sur les vertus (non plus que sur les vices, s'ils existent) de l'OED. Un de ses créateurs figure parmi les intervenants³. Il est mieux qualifié que moi pour en décrire les fonctions. Tout le monde est conscient que l'OED dans sa version 2 représente l'exemple le plus achevé que la lexicographie puisse offrir présentement sur le marché. Tout au plus peut-on observer que ce produit est bon marché si on le compare au *Robert*, et plutôt cher si on le confronte à *Encarta* et aux CD-Rom encyclopédiques. Mais que ne permet-il pas ! Nous en donnerons un ou deux exemples imprévus, qui font appel à la statistique. Toute base de données correctement construite – et l'OED est de ce nombre – peut restituer, mais aussi simplement décompter les enregistrements qui satisfont aux critères d'une requête donnée. On peut se contenter de ces effectifs et construire sans trop d'efforts un tableau à deux dimensions où les lignes seront constituées par les mots différents et les colonnes par les rubriques intéressées. En suivant la mauvaise pente où les mots VICE et EROTIC nous ont entraîné, nous ajouterons les mots SEXUAL et PORNO à notre méchante série. Voici les résultats exploitables :

	texte	étymologie	définition	citations
<i>SEXUAL</i>	1 758	30	770	774
<i>EROTIC</i>	120	6	32	75
<i>PORNO</i>	20	1	1	15

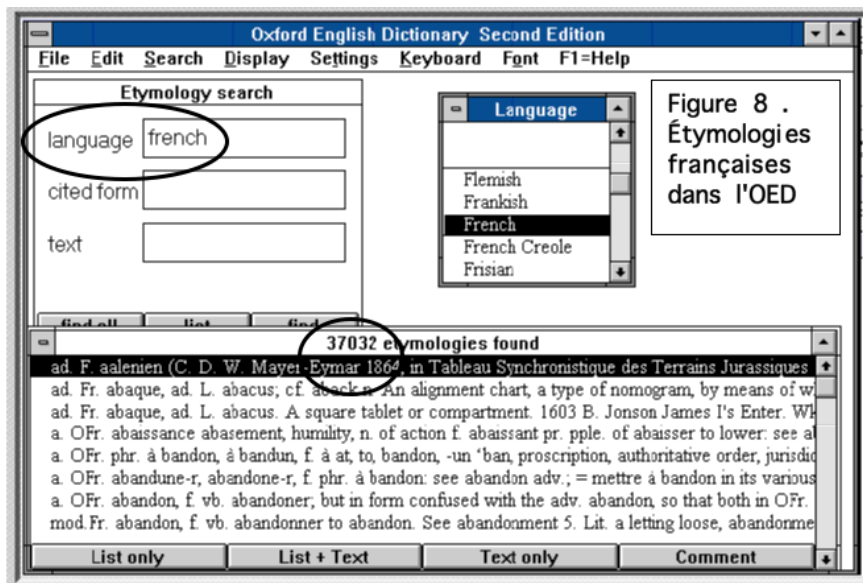
On a écarté le *VICE* pour s'en tenir à la même catégorie grammaticale. Nul besoin de calculer de savants écarts réduits pour rendre compte de la réserve que les rédacteurs opposent aux mots trop violemment marqués. Les auteurs auxquels les citations sont empruntées n'ont pas la même pudeur. Mais comme ce sont les rédacteurs qui choisissent les citations on croit discerner quelque ombre dans les replis de leur conscience.

Plus sérieusement, on peut s'intéresser à l'étymologie pour laquelle l'OED offre une grille particulière, représentée dans le graphique 8. Comme dans les autres champs on a accès à la graphie (dans différents jeux de caractères) ou, grâce aux jokers, à un modèle de production, mais on dispose aussi d'une rubrique propre qui mentionne la langue en

³ En la personne de Lou Burnard, Université d'Oxford.

question. C'est l'occasion de vérifier si de l'autre côté de la Manche on parle aussi franglais. L'effectif relevé pour le français et fourni par la figure 8 (37 032 étymologies) prend sens si on le rapproche de ceux qu'on obtient pour les autres langues⁴ :

	<i>French</i>	<i>Latin</i>	<i>Greek</i>	<i>German</i>	<i>American</i>
<i>language</i>	37 022	50 725	18 675	12 322	—
<i>text</i>	10 634	4 565	4 087	4 502	8 005
<i>definition</i>	1 755	1 330	1 261	940	2 393
<i>citation</i>	6 708	1 797	2 234	2 670	4 694



Le résultat est flatteur pour le sentiment national : avec un effectif de 37 022, les sources françaises ont un débit trois fois supérieur aux germaniques. Le mot FRENCH l'emporte aussi sur tous les autres, pour sa fréquence dans le texte des articles, et surtout dans les citations – ce qui est un phénomène de culture et de civilisation plutôt que de langue

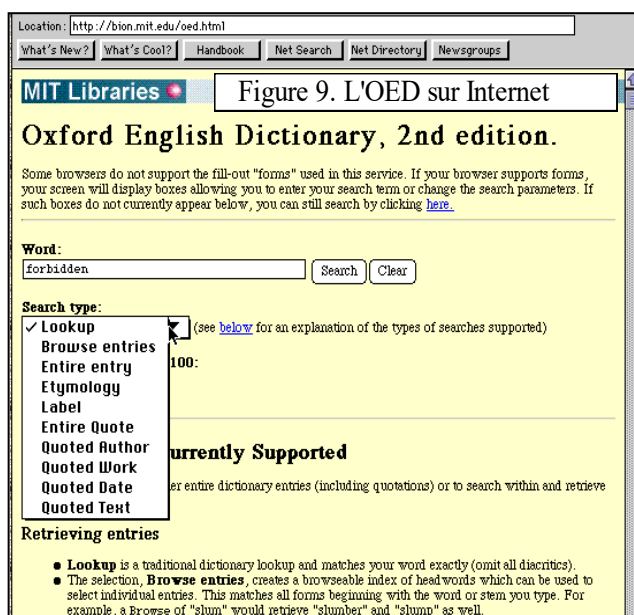
⁴ Voici le classement obtenu :

1	50 725	Latin	7	6 286	Deutch	13	2 824	Teutonic
2	37 022	French	8	5 795	Spanish	14	2 294	Provençal
3	18 675	Greek	9	4 430	Norse	15	2 120	Frisian
4	14 119	English	10	3 438	Swedish	16	1 889	Saxon
5	12 322	German	11	3 130	Portuguese	17	1 744	Anglo-French
6	7 893	Italian	12	3 046	Danish	18	1 480	Gothic

proprement dite. Reste à savoir si le mot est pris en bonne ou mauvaise part, question qu'il est imprudent d'approfondir.

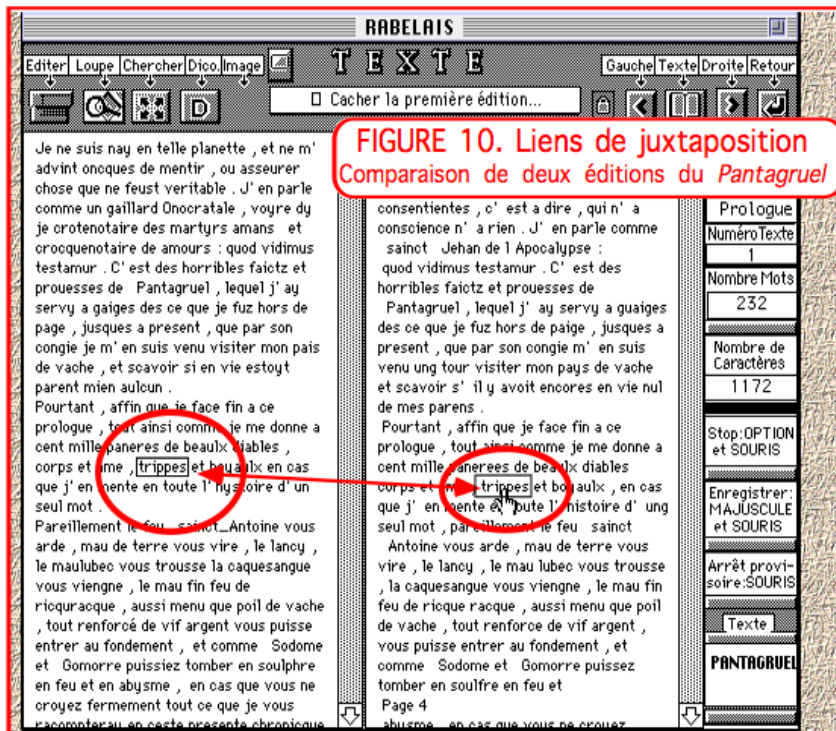
-III-

On pourrait s'en tenir là : prendre l'OED pour modèle et le transposer en français. Ce serait oublier que l'OED est en mouvement, qu'il en est à sa seconde version et que déjà une troisième est à l'épreuve sur *Internet*. Nous n'avons pu l'expérimenter, faute de posséder la clé dont jouissent certaines universités américaines ou canadiennes. La tentative a avorté que nous avons tentée sur le serveur du MIT. Méfiant sur la validité de notre passeport, nous avons proposé le mot FORBIDDEN à l'interrogation, comme précisé dans la figure 9. Nous avons obtenu le message FORBIDDEN avec le refus du mot demandé – dont nous ignorerons à jamais la signification.



1- Prenant appui sur notre expérience personnelle et persuadé que les techniques du texte intégral ne sont pas étrangères à celles des bases structurées, nous explorerons certains domaines où de nouveaux liens peuvent être établis. La liaison à laquelle on songe d'abord n'est qu'une variante de la démarche courante qui produit les concordances ou les contextes. En s'appuyant pareillement sur les adresses de l'indexation, on

peut projeter le lecteur directement sur le passage repéré, puis de proche en proche sur les passages suivants où les mêmes critères sont satisfaits. Ce processus a le mérite de ne pas brusquer le rythme de son attention, ni pour le faire languir, puisque le premier passage lui est donné immédiatement sans attendre que la série soit complète, ni pour lui faire presser le pas, puisqu'on attend son signal pour livrer la séquence suivante. Ces sauts hypertextuels dans le texte même de la base donnent une image assez proche de la recherche manuelle qu'on accomplit en tournant les pages. Encore faut-il permettre à l'utilisateur d'épingler un passage qui suscite son intérêt et l'autoriser à abandonner la partie quand son intérêt s'est émoussé.



2- Les liens de juxtaposition n'ont pas lieu de s'exercer lorsque le corpus qu'on exploite est taillé d'une pièce, ce qui est le cas du *TLF*, même s'il a fallu trente ans pour en assurer la réalisation. Mais il arrive souvent qu'un dictionnaire vieillisse par l'effet de l'évolution de la langue, puis rajeunisse par l'effet des révisions. Ce sera peut-être le cas du *TLF* dans un siècle ou deux. Qu'on songe que le *Dictionnaire de l'Académie*

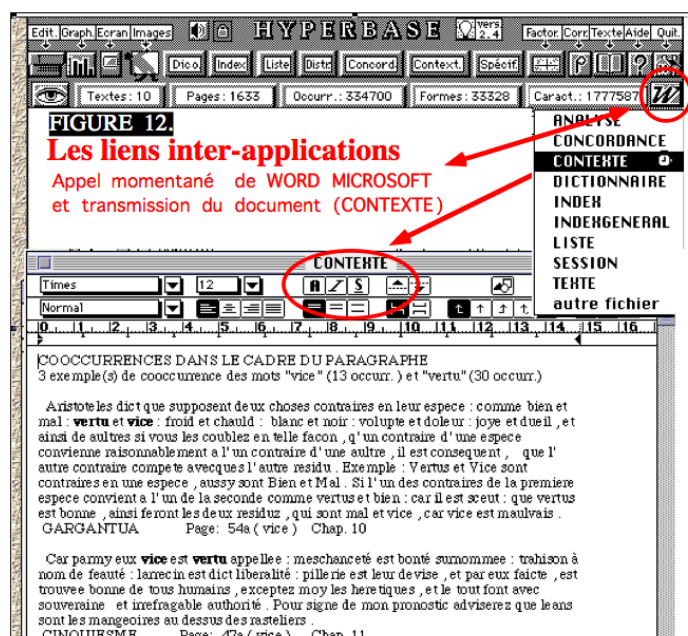
française en est à sa neuvième édition depuis la première en 1694. Cette préoccupation est partagée par tous ceux qui ont à comparer des éditions, et la *Bibliothèque Nationale de France* se soucie de fournir aux chercheurs les outils de la juxtaposition, dont une illustration est fournie par la figure 10, à partir d'un CD-ROM consacré à Rabelais et son temps⁵.

Un clic sur un mot de l'une des éditions juxtaposées déclenche la recherche de la forme correspondante dans l'édition parallèle. La fragmentation, assez délicate lorsqu'il s'agit de texte intégral et que les versions s'éloignent l'une de l'autre, devrait poser moins de problèmes lorsqu'on a affaire à une nomenclature, dont les entrées sont largement communes, sinon la structure et le contenu⁶.

3- Dans l'avenir proche, il est une liaison légitime à laquelle on ne peut pas ne pas penser. C'est celle qui mettrait *en relation le TLF et Frantext*. Juste retour des choses : le TLF dont les exemples sont puisés dans le corpus de *Frantext* s'honorerait en avouant ses dettes et en renvoyant l'utilisateur à la source. Et cette jonction est d'autant plus réalisable que le même demiurge qui a conçu pour *Frantext* le très remarquable logiciel *Stella* (et le moteur de recherche du CD-ROM *Discotext 1*) est aussi chargé de l'informatisation du TLF. Dans le premier prototype qu'il a mis en œuvre et que j'ai pu admirer récemment, certaines procédures de traitement qui ont prouvé leur efficacité dans *Frantext* sont reprises et adaptées au nouveau produit. On imagine aisément de quelle amplification bénéficierait la recherche d'exemples, si le relais était assuré par *Frantext* et qu'au faible pourcentage qui a été retenu par les rédacteurs on ajoutait ce qu'ils ont dû sacrifier. Et inversement est-il un meilleur moyen de pénétrer dans la nef de *Frantext* que de passer d'abord par le porche du TLF ? Et ce porche pourrait être assez large pour donner accès aussi à un édifice plus grand encore : la *Bibliothèque Nationale de France*.

⁵ La publication de ce CD-ROM est prévue pour Juin 1995, aux Éditions *Les temps qui courent*, 118-130 avenue Jean Jaurès, 75019 Paris. Précisons que le projet a été conduit par M.L. Demonet et que nous en avons assuré la réalisation technique, à partir de notre logiciel *HYPERBASE*.

⁶ Il s'agit d'ici d'une situation simple, qui trouve sa solution dans un affichage juxtalinéaire. Mais dans les cas plus complexes, la stratégie devra se diversifier. Sur ce sujet qui touche à la génétique et à la critique des textes, on tirera profit de l'expérience de R. Laufer (*Texte, Hypertexte, Hypermédia*, PUF, 1992), de J.L. Lebrave (« Hypertextes - Mémoires - Écriture », in *Genesis*, n° 5, 1994, p. 9-24 et de B. Stiegler (« Machines à écrire et matière à penser », in *Genesis*, n°5, 1994, p.25-49).

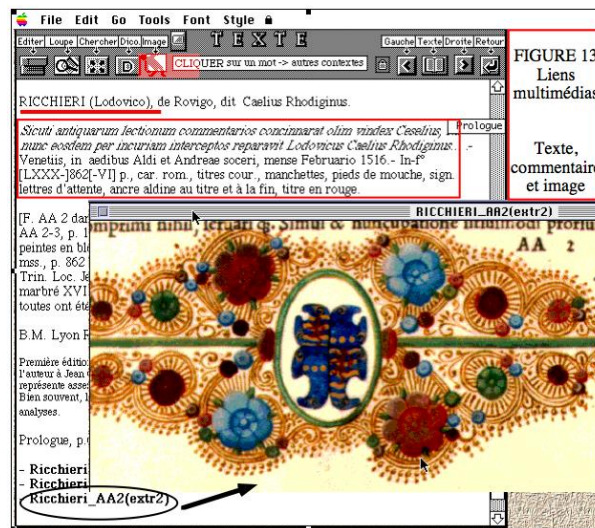


L'utilisateur a nécessairement la pratique d'un traitement de texte, dont le choix ne peut être imposé. Et il se trouve tôt ou tard désireux de transmettre à son traitement de texte familier les résultats générés par l'exploitation de la base. Facilitons-lui la tâche en installant un pont entre les applications sans qu'il soit nécessaire d'abandonner l'une pour rejoindre l'autre⁷. Certes il est toujours possible de créer des fichiers ASCII que l'utilisateur retrouvera en fin de séance. Mais il vaut mieux lui permettre de disposer sans attendre de tels fichiers (pour qu'il puisse par exemple les intégrer dans l'article qu'il rédige). Dans l'exemple représenté ci-dessus, un bouton avec menu *pop-up* permet de donner la main à Word (ou au logiciel de son choix) en ouvrant le document où les résultats de l'exploitation ont été enregistrés (il y a là aussi une grande variété de choix)⁸.

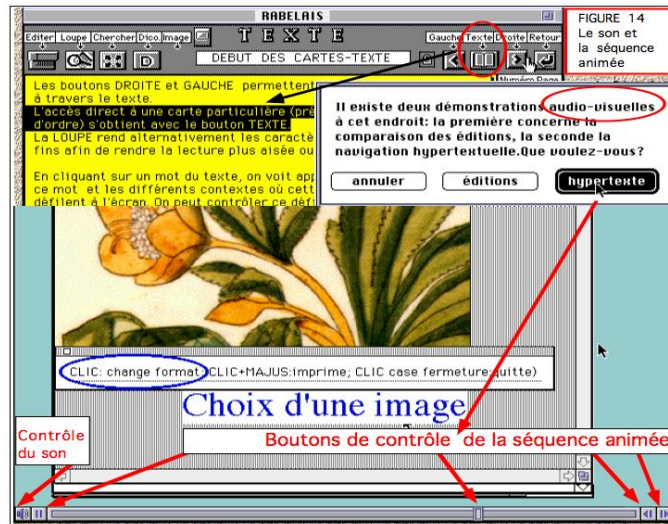
⁷ De toute façon, les mouvements qui se dessinent dans l'informatique tendent actuellement vers l'intégration des applications, qu'il s'agisse de la technologie *OpenDoc* ou *OLE* ou de quelque autre.

⁸ Ni *Discotext 1* ni l'OED2 ne sont des modèles à cet égard. Des raisons commerciales liées au copyright ont empêché que les résultats apparaissent en clair, sinon sur l'imprimante, où il est bien difficile de les repêcher.

4- Reste à étendre le champ des liens en s'ouvrant aux autres canaux de la communication : *l'image, le son, l'animation et la vidéo*. La tradition du livre intègre depuis longtemps l'image au texte et, à l'origine, avant le parchemin et le papier, on peut penser que la relation a été inverse : c'est l'écriture qui s'est mêlée progressivement à l'image. On assiste actuellement à une situation curieuse et presque rétrograde, où le texte redescend au statut de simple image. Car il est bien plus facile au scanner de transformer un texte imprimé en livre d'images que de décrypter les caractères graphiques pour en faire des codes ASCII et retrouver un texte exploitable. Il est des cas cependant où la reproduction de l'image est nécessaire, lorsqu'il s'agit de documents rares ou peu accessibles, dont la forme matérielle importe autant que le contenu. Il en est ainsi du CD-Rom Rabelais (figure 13), qui donne accès à des centaines d'illustrations et où l'on peut examiner les documents de l'époque liés au texte de Rabelais.



5- Nous donnerons pour mémoire la figure 14 qui n'est malheureusement guère parlante, la parole manquant au papier. Elle tend à expliciter les *relations multimédia* qu'un CD-ROM peut inclure dans ses microcuvettes. Les techniques liées au laser – et plus généralement à tout support matériel de la mémoire – enregistrent en effet sous forme binaire n'importe quelle source d'information, y compris les séquences animées ou sonores. Quel intérêt, dira-t-on, pour le TLFi ?



On peut comprendre qu'une encyclopédie ait recours aux moyens audiovisuels pour expliquer les choses. Mais pour expliquer les mots a-t-on besoin de telles ressources ? N'oublions pas qu'un dictionnaire rend compte du langage et que les mots sont dits avant d'être écrits. Quelles acrobaties le code écrit ne doit-il pas exécuter pour rendre compte de la prononciation ? Alors que l'accord international n'a été obtenu que pour les 128 premiers caractères du code ASCII, comment utiliser l'alphabet phonétique international quand on ne sait pas sur quelle plateforme l'information sera affichée ? L'OED a rencontré cette difficulté sans la résoudre pleinement car si on lit sur un Macintosh le CD-Rom prévu pour le standard Windows, on obtient pour la prononciation des codes aberrants. On peut donc imaginer que la rubrique *prononciation* du dictionnaire s'adresserait à l'oreille. Et c'est d'ailleurs le cas de certaines bases de données de dialectologie, comme celle qui est réalisée actuellement sur la Corse.

Mais plus généralement tout logiciel a besoin d'explication et s'accompagne d'un manuel de référence ou d'une aide en ligne, laquelle habituellement reprend le manuel en le dotant d'index et de mots-clés. Mais le plus souvent ni le manuel, ni l'aide en ligne ne suscitent l'intérêt de l'utilisateur, parce que celui-ci manque de temps et de courage et parce que les documents fournis manquent d'attrait pédagogique. Une pratique tend à se répandre qui est de lui proposer une démonstration audiovisuelle, pour les actions délicates à comprendre. Évitions là encore de tomber dans le travers d'une « Visite guidée » qui trop longue et trop

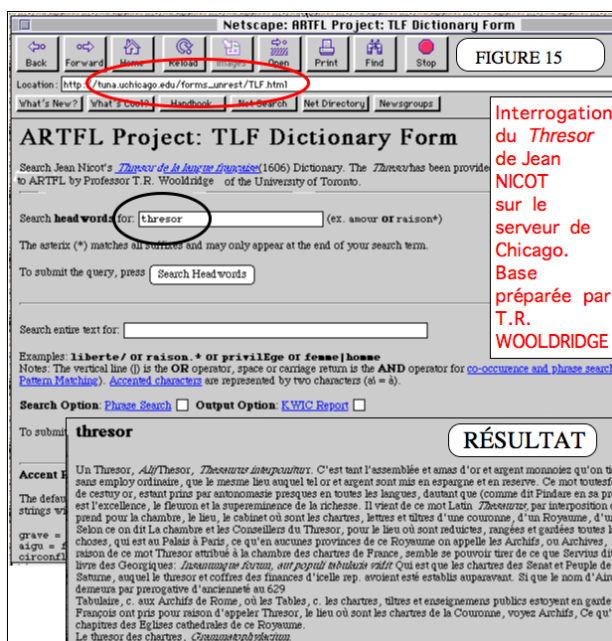
générale découragera les bonnes volontés. C'est lorsqu'il sollicite une fonctionnalité inconnue que l'utilisateur doit être pris en charge, s'il le désire. L'exemple de la figure 14 est relatif à un symbole en forme de livre ouvert que l'utilisateur a le droit de trouver opaque. Si l'on s'interroge sur sa fonction, un court message écrit est d'abord proposé, et, dans un second temps, une séquence *Quicktime* qu'on peut écouter et visionner à sa guise, tous les contrôles étant à portée de main, par exemple pour reprendre un passage et recaler le son.

-IV-

Mais, dira-t-on, comment faire tenir tout cela sur l'étroite surface d'un CD-ROM ? Comment y engranger tant de liens entrecroisés, tant de fichiers associés, tant de fonctions cumulées ?

1- La réponse est négative dans le temps présent. L'OED a eu besoin de tous les octets disponibles sur la surface optique et aucun des 632 Mo n'a été gaspillé. Mais une mutation technique se prépare qui met aux prises deux clans pour une même norme. On ne connaît pas encore le vainqueur, mais qu'il s'agisse du DVD (Digital Video-Disc) de Philips et Sony, ou du SD (Super Density) de Toshiba, la capacité du futur CD-Rom sera multipliée ainsi que le débit de transmission. On parle de 3 à 10 milliards d'octets. C'est assez pour contenir en même temps le TLF et *Frantext* et bien d'autres choses encore. Et déjà bleuit l'horizon où l'on devine les premières lueurs du laser bleu. Ici comme en d'autres circonstances la technique informatique comblera son retard avant que les données soient prêtes et les outils fabriqués.

2- Mais par d'autres voies qui font appel au réseau télématique la technique actuelle ouvre des perspectives que le TLF pourrait prospecter. Ici on ne parlera plus de liens locaux, mais de liaisons extérieures qui s'étendent aux limites du monde civilisé. Déjà *Frantext* est accessible sur le réseau *Internet* (par l'intermédiaire du logiciel de communication *Telnet*). Et l'on a vu que certains privilégiés avaient le droit d'exploiter une version WWW de l'OED. Bien d'autres dictionnaires sont déjà en ligne (nous en avons compté plus d'une vingtaine), même si tous ne sont pas gratuitement offerts. On choisira pour exemple la réalisation de Wooldridge, parce que c'est la plus généreuse (aucun droit n'est à payer) et parce qu'elle n'est pas sans relation avec le *Trésor de la langue française*. Il s'agit pareillement en effet du *Thresor*, écrit à l'ancienne, puisque c'est celui de Jean Nicot.



Comme notre ami Wooldridge doit prendre la parole sur ce sujet nous nous en voudrions d'empiéter sur son domaine, fût-ce pour en dire du bien.

3- Contrairement à ce qui se passe dans une situation locale, où le concepteur d'un produit a les mains libres pour établir les liens qu'il souhaite et les traiter à sa façon (on attend simplement de son programme qu'il soit efficace), les réalisateurs qui abordent *Internet* doivent se soumettre à une normalisation bénéfique mais contraignante. Celle qui a le plus de succès présentement et qui régit plus de 3 millions de machines connectées est celle que le CERN a établie et que respectent les serveurs WEB (ou W3 ou WWW). Cette norme a pour nom HTML (HyperText Markup Language). Elle est la cousine de la norme SGML à laquelle les données du TLF sont soumises. Comme un rapprochement entre les spécifications des deux langages est en train de s'opérer présentement qui prendra effet avec la version HTML 3.0, on est en droit d'espérer que les données du TLF pourront sans dommage et presque sans effort être adaptées au format requis et accessibles à travers le plus grand réseau mondial. Pour donner une idée de la facilité avec laquelle des liens hypertextuels peuvent être noués dans le langage HTML, on présente ci-dessous un exemple synoptique qui met en présence un document HTML et l'effet obtenu. Quoique simple, cet exemple n'en montre pas moins

comment s'établissent les liens qui mènent aux images, aux sons, aux animations et – ce qui nous importe plus encore ici – aux bases de données en interactif.

Le langage HTML (source)

```
<HTML><HEAD><TITLE>Les objets
HTML</TITLE></HEAD><BODY> <H3>Un
&eacute;chantillon des diff&eacute;rents objets
HTML</H3><H3>1. Le texte</H3>
Dans les documents HTML diffus&eacute;s par W.W.W. et
visualis&eacute;s par MOSAIC ou NETSCAPE l'accent est
mis sur le contenu et non sur la pr&eacute;sentation
laquelle peut varier selon que le client dispose de telle ou
telle machine &agrave; la r&eacute;ception.
Cependant plusieurs attributs sont param&eacute;trables ,
notamment la police ou plus exactement le "style":
<UL> <LI><EM>par exemple, italique...</EM>
<LI><STRONG>... ou bien encore,
gras.</STRONG></UL>
<H3>2. Les images</H3><IMG SRC="image.gif"> ( GIF )
<p>
Deux formats d'images peuvent &eacute;tre directement
inclus dans un document HTML : il s'agit des images GIF et
XBM. Les autres types d'images (e.g. JPG, PICT, TIFF,
etc...) ne peuvent &eacute;tre consult&eacute;s qu'au travers
de <A HREF="image2.jpg">liens externes</A> <H3>3. Le
son</H3>
Comme la plupart des images, un son doit &eacute;tre
associ&eacute; &agrave; un <A
HREF="bienvenue.au">lien
externe</A>,g&eacute;n&eacute;ralement au format
<i>.au</i>.
<H3>4. Les films</H3>
La remarque pr&eacute;c&eacute;dente s'applique aux <A
HREF="sequence.mov">films</A>, qui peuvent &eacute;tre
au format MPG ou QUICKTIME. En fait, pour &eacute;tre
tr&eacute;g&eacute;ral, un lien peut pointer sur
un document d'un type quelconque, d&eacute; l'instant que
le client poss&eacute;de un moyen de visualiser ce
document (e.g. du <A
HREF="/misc/eclips.ps">postscript</A>)
<H3>5. L'interactivit&eacute;</H3>
Enfin, on peut installer un dialogue entre le serveur et le
client, gr&eacute;ce &agrave; des "widgets" Motif,
directement disponibles dans un document HTML (sur un
client XMosaic v2.x). Citons : <FORM ACTION="/cgi-
bin/echantillon">
<UL>
<LI> Un champ de saisie <input NAME="champ"
VALUE="Default">
<LI> Un menu <SELECT NAME="menu">
<OPTION>Item1<OPTION>Item2<OPTION>Item3
</SELECT>
<LI> Un bouton <input TYPE="submit"
VALUE="Bouton">
<LI> Une alternative <input TYPE="radio"
NAME="radio" VALUE="Choix1" CHECKED>Choix1
<input TYPE="radio" NAME="radio"
VALUE="Choix2">Choix2
</UL></FORM>
</BODY></HTML>
```

Effet obtenu

Netscape: Les objets HTML


Un échantillon des différents objets HTML

1. Le texte

Dans les documents HTML diffusés par W.W.W. et visualisés par MOSAIC ou NETSCAPE l'accent est mis sur le contenu et non sur la présentation laquelle peut varier selon que le client dispose de telle ou telle machine à la réception. Cependant plusieurs attributs sont paramétrables , notamment la police ou plus exactement le "style":

- *par exemple, italique...*
- ... ou bien encore, **gras**.

2. Les images

 (GIF) Deux formats d'images peuvent être directement inclus dans un document HTML : il s'agit des images GIF et XBM. Les autres types d'images (e.g. JPG, PICT, TIFF, etc...) ne peuvent être consultées qu'au travers de [liens externes](#)

3. Le son

Comme la plupart des images, un son doit être associé à un [lien externe](#), généralement au format *.au*

4. Les films

La remarque précédente s'applique aux [films](#), qui peuvent être au format MPG ou QUICKTIME. En fait, pour être très général, un lien peut pointer sur un document d'un type quelconque, dès l'instant que le client possède un moyen de visualiser ce document (e.g. du [postscript](#))

5. L'interactivité

Enfin, on peut installer un dialogue entre le serveur et le client, grâce à des "widgets" Motif, directement disponibles dans un document HTML (sur un client XMosaic v2.x). Citons :

- Un champ de saisie
- Un menu
- Un bouton
- Plusieurs boutons Choix1 Choix2

Au terme de notre parcours, nous nous sentons comme Gulliver au pays des nains : ligoté de toutes part et ne pouvant faire un geste sans tirer sur un lien. Du mince fil du téléphone part une grappe infinie de fils invisibles qui enserrant le monde entier. Et l'image de l'araignée qui tisse sa toile est le symbole du WEB.

Abondance de liens ne nuit pas, est-on tenté de dire. Certes on ne peut faire l'impasse sur les autoroutes de l'information où théoriquement circulent des millions de transferts à la seconde (bientôt 150 Mégabits sur ATM). Et il faut se préparer à cette mutation des échanges. Mais il faut surtout en mesurer les dangers : vieillissement des informations par absence de mise à jour, précarité des liens pour cause de changement d'adresse, futilité des messages dont beaucoup stationnent inutilement sur la voie publique, insécurité des communications, risque de mêler les fils d'Ariane et de se perdre dans le labyrinthe du Minotaure électronique et surtout embouteillage grandissant du trafic : il est devenu difficile à certaines heures de joindre sur *Internet* un correspondant américain et l'on songe à telle autoroute des abords de New-York qui est si encombrée et si lente qu'elle passe pour être le plus grand parking du monde.

Dans un ouvrage récent⁹, le directeur du CNRS Guy Aubert invite à se méfier du « tout électronique » et de cette tour de Babel aveugle et bavarde qu'un plaisantin a appelée « BabelWeb ». Il continue, pour les échanges humains et les communications scientifiques, à faire confiance aux bibliothèques, aux cafétérias, aux discussions *in praesentia*. Et cela justifie le présent colloque, où les vrais liens peuvent se nouer.

⁹*L'Internet professionnel*, 1er trimestre 1995, Éditions du CNRS, p.233-234. Ouvrage collectif édité par Alain Simeray (CNRS/SOSI).