



La production journalistique et Google : chercher à ce que l'information soit trouvée

Guillaume Sire

► To cite this version:

Guillaume Sire. La production journalistique et Google : chercher à ce que l'information soit trouvée. Sciences de l'information et de la communication. other, 2013. Français. NNT : . tel-02433761

HAL Id: tel-02433761

<https://hal.science/tel-02433761v1>

Submitted on 9 Jan 2020

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 Panthéon-Assas

**Ecole doctorale de Sciences économiques, de Gestion et
d'Information et Communication (ED07)**

Thèse de doctorat en Sciences de l'Information
et de la Communication
soutenue le 18 novembre 2013

La production journalistique et Google : chercher à ce que l'information soit trouvée



Université Panthéon-Assas

Guillaume Sire

Sous la direction de Mme le Professeur Nathalie Sonnac

Membres du jury :

Mme Josiane JOUËT, Professeur à l'Université Paris 2 Panthéon-Assas, Présidente du Jury

Mme Nathalie SONNAC, Professeur à l'Université Paris 2 Panthéon-Assas, Directrice de Thèse

Mme Cécile MÉADEL, Professeur à l'École des Mines de Paris

M. Arnaud MERCIER, Professeur à Université de Lorraine

Mme Laurence MONNOYER-SMITH, Professeur des Universités, Vice-Présidente Commission Nationale du Débat Public, Rapporteur

M. Franck REBILLARD, Professeur à l'Université Paris 3 Sorbonne-Nouvelle, Rapporteur

Avertissement

La Faculté n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse ; ces opinions doivent être considérées comme propres à leur auteur.

« "Montrez-moi patte blanche, ou je n'ouvrirai point",
S'écria-t-il d'abord. (Patte blanche est un point
Chez les loups, comme on sait, rarement en usage.)
Celui-ci, fort surpris d'entendre ce langage,
Comme il était venu s'en retourna chez soi. »

Jean de La Fontaine — *Le Loup, la Chèvre et le Chevreau* — 1668

« [L]e journalisme est dans l'enfance, il grandira. Tout, dans dix ans d'ici, sera soumis
à la publicité. La pensée éclairera tout, elle... »

Balzac — *Les Illusions perdues* — 1843

« Évaluer c'est créer : écoutez donc, vous qui êtes créateurs ! C'est leur évaluation
qui fait des trésors et des bijoux de toutes choses évaluées.
C'est par l'évaluation que se fixe la valeur : sans l'évaluation, la noix de l'existence
serait creuse. Écoutez donc vous qui êtes créateurs !
Les valeurs changent lorsque le créateur se transforme. Celui qui doit créer détruit
toujours. »

Nietzsche — *Ainsi parlait Zarathoustra* — 1885

Avant-propos

Avant de commencer mon parcours de thèse, je travaillais en qualité de prestataire pour le compte d'une entreprise du CAC 40 où j'œuvrais à la gestion des sites web et de l'intranet. J'y ai réalisé que l'écriture et la mise en ligne de documents pouvaient être contraintes et qu'une relativement bonne connaissance du code HTML permettait de détourner et de réarranger certains cadres d'action préétablis. Il s'agissait à la fois d'un langage, d'outils techniques, d'infrastructures physiques, de logiciels, d'organisation d'entreprise — la mienne (agence) et l'autre (client) — d'informations, de méta-informations et aussi, pensais-je à l'issue de situations conflictuelles, de *pouvoir*.

En mai 2010, je lus une chronique du *New York Times* où l'éditorialiste David Carr s'interrogeait sur la manière dont l'écriture journalistique pouvait être contrainte, et même *formatée*, en raison d'enjeux liés au référencement sur Google (Carr, 2010). Je démissionnai quelques semaines plus tard pour m'inscrire à l'Institut Français de Presse où le Professeur Nathalie Sonnac et l'équipe scientifique du Centre d'Analyse et de Recherche Interdisciplinaires sur les Médias avaient accepté de m'accueillir. Prenant acte des propos de Yochai Benkler selon lesquels « les contours de l'environnement informationnel et la distribution du pouvoir relatif au contrôle des flux d'information depuis et jusqu'aux individus constituent le produit contingent d'une combinaison de technologies, de comportements économiques, d'enjeux sociaux, de structures institutionnelles ou de lois » (Benkler, 2006, p. 147), je me lançai dans l'étude de la combinaison des activités de Google et des éditeurs de presse dans le but de tracer les contours d'un environnement informationnel singulier et d'identifier, dans cet environnement, à qui, sous quelles formes et à quel point, appartenait le *pouvoir* de contrôler les flux d'information politique et générale.

Remerciements

Je tiens à remercier ceux qui ont nourri le parcours dont ce texte rend compte. Je les ai lus, rencontrés, sollicités, questionnés, assaillis, épuisés, tant et si bien que je n'ai eu pour rédiger qu'à distiller l'essence de ce qu'ils m'avaient appris et donné.

C'est au Professeur Nathalie Sonnac que va mon premier remerciement, pour la confiance qu'elle m'a faite, son aide, ses conseils, nos discussions. Elle fut un mentor pour ce travail autant que pour mes premiers pas d'enseignant.

Je remercie l'Université Panthéon-Assas de m'avoir accordé un contrat doctoral et permis ainsi de travailler dans les meilleures conditions. Je remercie en particulier les membres de la Commission à qui il appartenait de prendre cette décision.

Ma gratitude va à la Présidente du jury Josiane Jouët. Je la remercie d'être aussi disponible pour les jeunes chercheurs du CARISM, attentive et affable, tout en faisant preuve d'une exigence scientifique sans faille.

Je remercie chacun des membres du jury : Cécile Méadel, Arnaud Mercier, Laurence Monnoyer-Smith et Franck Rebillard. J'ai beaucoup appris grâce à leurs travaux aussi bien à propos de mon objet de recherche qu'à propos de la démarche scientifique en général, et je suis par conséquent très heureux qu'ils aient accepté d'examiner les miens. Je remercie Laurence Monnoyer-Smith et Franck Rebillard d'avoir accepté d'être les rapporteurs de ce travail.

Je remercie mes collègues du CARISM, en particulier son Président Frédéric Lambert pour les échanges que nous avons eus, Jean-Baptiste Comby pour ses conseils et sa bienveillance, et les jeunes chercheurs Coralie Le Caroff, Virginie Sonnet et Alan Ouakrat pour leur soutien et leur amitié. Je remercie également Hélène Guweddeko et Gianni Gortan, que j'ai parfois enquisinés, et qui, toujours, m'ont répondu et aidé.

Je remercie les chercheurs des laboratoires extérieurs avec lesquels j'ai pu échanger : Robert Boure, Nikos Smyrniaios, Bernhard Rieder, Francesca Musiani, Juliette De Maeyer, Romain Badouard, Jean-Christophe Plantin et Clément Mabi. Je les remercie pour leurs relectures et leurs recommandations. Ils m'ont tous énormément aidé, et j'ai grâce à eux le sentiment d'avoir intégré une communauté

débordant largement le cadre d'un seul laboratoire. Je voudrais remercier en particulier Bernhard Rieder, avec qui j'ai eu le privilège de collaborer, et Clément Mabi, dont ce travail doit beaucoup à l'amitié et aux suggestions.

Je remercie les acteurs qui ont accepté que je les observe et les interroge, en particulier chez Roularta Media France où les employés sont libres de s'exprimer. Etudier une population d'acteurs, c'est être auprès d'eux l'étudiant : les *comprendre*, c'est-à-dire étymologiquement les « prendre avec soi ». J'ai appris auprès d'eux tous, et j'espère finalement avoir *compris* ce qu'ils m'ont montré et expliqué.

Je remercie les membres de ma famille : ma mère qui m'a transmis son enthousiasme et sa curiosité, mon père qui m'a donné le goût de la recherche et m'a encouragé dans cette voie, et ma sœur et mon frère sur le soutien et l'affection desquels j'ai toujours pu compter.

Je remercie également mes amis, parmi lesquels ma gratitude va particulièrement à Pierre-Louis Boyer, qui m'a fait part de son expérience, m'a stimulé et m'a soutenu, ainsi qu'à Albéric, Alexandre, Emmanuel, Jehan, Julien, Maximin, Marc, Marc-Elie, Martin, Pierre C., Pierre J., Rodolphe et Romain.

Enfin, et *infiniment*, je remercie mon épouse Daphné. Je ne serais pas venu à bout de ce projet si elle n'avait pas fait en sorte qu'il soit compatible avec le nôtre, qui sera toujours à mes yeux le plus important.

La production journalistique et Google : chercher à ce que l'information soit trouvée

Résumé

Nous cherchons dans ce travail à détricoter la relation à la fois compétitive et coopérative, indifféremment technique, économique, juridique, sociale, politique et résolument communicationnelle de Google et des éditeurs de presse. Pour cela, après avoir historicisé la rencontre de deux univers singuliers, nous décrivons ce que les éditeurs peuvent faire pour franchir le prisme du moteur de recherche et y optimiser la visibilité de leur production. Nous tâchons ensuite de décrypter ce que la firme Google peut faire faire aux éditeurs en analysant leurs relations de pouvoir, leurs incitations, leurs projets et leurs environnements informationnels respectifs. Enfin, nous rendons compte de ce que les éditeurs français issus de la presse imprimée font effectivement : ce qu'ils communiquent à Google, par quels moyens et à quel prix, pour quels résultats espérés, à l'issue de quelles concessions, quels détours, quelles contestations. Nous expliquons comment les conditions et les modalités de captation du trafic sont susceptibles d'influencer la valorisation du contenu, l'organisation de sa production, la structure du site, les pratiques journalistiques et les lignes éditoriales. Nous montrons qu'un aller-retour performatif se crée entre énoncés et conditions d'énonciation, agissant *par* et *sur* les textes, les architextes et les hypertextes. En somme, c'est à la compréhension de ce que deviennent l'actualité et ceux qui la mettent en récit, dès lors qu'ils cherchent à ce que l'information qu'ils produisent soit trouvée par les utilisateurs de Google, que notre thèse est consacrée.

Descripteurs : Google ; journalisme ; SEO ; Internet ; moteurs de recherche ; hyperliens ; HTML ; marchés multiversants

Google and journalistic production: Making news findable

Abstract

In this thesis, we aim to disentangle the cooperative but also competitive relationship between Google and news publishers, which is at the same time technical, economic, legal, social, political and certainly communicational. In order to do so, we trace the historical development of two singular universes, describing what publishers can do to overcome the search engine and optimize their ranking. We then analyse how Google can influence publishers' conduct, by studying power relations, respective incentives, aims, and informational and socio-economic backgrounds. Finally, we report on actual practices of French traditional news publishers: what they communicate to Google, by which means and at what price, for which expected results, after which concessions, detours and controversies. Thus, we explain how search engine optimization is likely to affect the way content is valued, its production organisation, the website's structure, journalists' practice and editorial policy. We show a back and forth movement between performative utterances and performed circumstances, having an effect *on* and *by* texts, architexts and hypertexts. To sum up, this thesis is dedicated to understanding what happens to news and publishers once they strive for their information to be found by Google's users.

Keywords : Google ; journalism ; SEO ; Internet ; search engines ; hyperlinks ; HTML ; multi-sided markets

Principales abréviations

ACAP : *Automated Content Access Protocol*

AMJ : Association Mondiale des Journaux

ANT : *Actor-Network Theory*

CAA : Centre d'Aide Actualité

CAW : Centre d'Aide aux Webmasters

CDN : *Content Delivery Network*

CE : Commission Européenne

CFC : Centre Français d'exploitation du droit de Copie

CMS : *Content Management System*

CNIL : Commission Nationale de l'Informatique et des Libertés

CNRTL : Centre National de Ressources Textuelles et Lexicales

CPC : Coût-Par-Clic

CPM : Coût-Pour-Mille

EPC : *European Publishers Council*

FAI : Fournisseur d'Accès à Internet

FPE : *Front Page Editor*

FTC : *Federal Trade Commission*

HADOPI : Haute Autorité pour la Diffusion des Œuvres et la Protection des droits sur Internet

IPA : *International Publishers Association*

IPG : Association de la presse d'information politique et générale

SEM : *Search Engine Marketing*

SEO : *Search Engine Optimization*

SIC : Sciences de l'Information et de la Communication

SIPI : Sites Issus de la Presse Imprimée

SNEL : Sites Nés En Ligne

SPIIL : Syndicat de la Presse indépendante d'information en ligne

SPQN : Syndicat de la Presse Quotidienne Nationale

SPQR : Syndicat de la Presse Quotidienne Régionale

WAN : *World Association of Newspapers*

Sommaire

Introduction	19
Partie 1 — D'une configuration l'autre : problématisation et mise en perspective du dispositif	25
Chapitre 1. Penser le moteur de recherche comme un dispositif d'information et de communication	25
1.1 Reconfiguration des voies d'accès	25
1.1.1 Publier et publiciser	25
1.1.2 Enjeux du référencement	26
1.1.3 La question du prix à payer	27
1.2 Etudier l'infomédiation en SIC	28
1.2.1 Attachement à la dimension technique	30
1.2.2 Approche communicationnelle du web	30
1.2.3 Du côté de la production	32
1.2.4 Cadrage interdisciplinaire	34
1.2.5 Capacitation et incitations	35
1.3 Anthropologie d'une configuration médiatique	36
1.3.1 Théorie de l'acteur-réseau	37
1.3.2 Traductions	40
1.3.3 Programmes d'actions	41
1.3.4 Enrôlement et gouvernementalité	42
1.3.5 Performativité	45
1.3.6 Médiation sociotechnique	46
1.4 Le concept de dispositif appliqué à l'infomédiation	49
1.4.1 Foucault, Deleuze, l'ANT et les SIC	50
1.4.2 Dispositif d'infomédiation	53
1.4.3 Analyser où se situe le pouvoir	54
1.4.4 Lignes de fuite	57
1.5 Problématique et hypothèses	59

1.6 Terrains	61
1.6.1 Méthodologie de collecte	61
1.6.2 Approche qualitative	63
1.6.3 De la difficulté d'étudier Google	64
1.6.4 Editeurs et référenceurs	66
1.6.5 Groupe Express Roularta	71
Chapitre 2. Généalogie de la recherche d'information	73
2.1 Musée imaginaire	73
2.1.1 Produire pour être stocké, stocker pour être trouvé	73
2.1.2 Répondre à un besoin d'information	74
2.1.3 A chaque support ses méta-informations	74
2.1.4 A nouveau contenant, nouveaux procédés	75
2.1.5 A nouveau support, nouvelle présentation	76
2.1.6 L'utopie de Paul Otlet	77
2.2 La science de la recherche d'information	79
2.2.1 Genèse	79
2.2.2 Algorithme d'appariement	80
2.2.3 Arrivée du web	82
2.2.4 Analyse des liens	84
2.2.5 Fonctionnement opaque	86
2.3 ADN de Google	86
2.3.1 Université et entrepreneuriat	86
2.3.2 Ingénierie	87
2.3.3 Esprit Montessori	89
2.3.4 La publicité et l'argent	90
2.4 Comment Google est devenue leader sur le marché des moteurs	91
2.4.1 Portails et intégration verticale (1997-2001)	92
2.4.2 Syndication et consolidation (2001-2004)	92
2.4.3 Diversification et différenciation (2004-2009)	93
2.4.4 Monopolisation (2009-...)	95

Chapitre 3. Histoire de la distribution-diffusion des nouvelles	99
3.1 Parler à des statues parlantes	99
3.1.1 Place Pasquino	99
3.1.2 Monopole des maîtres-courrier	101
3.1.3 Contournements	103
3.1.4 Fin du monopole postal	104
3.1.5 La stratégie Hachette	104
3.1.6 Presse autorisée et presse clandestine	105
3.2 De la Libération à nos jours	106
3.2.1 Loi Bichet	106
3.2.2 Distribution et ligne éditoriale	109
3.2.3 Exposition sur le web	110
3.3 L'Express	111
3.3.1 Arbre généalogique	111
3.3.2 Roularta Media France	115
3.3.3 Du site compagnon au site d'information	117
3.3.4 Chiffres et perspectives	121
Conclusion Partie 1	122
<i>Partie 2 – Configuration négociée et négociable : déploiement du dispositif</i>	123
Chapitre 4. Le moteur comme champ de manœuvres	125
4.1 Crawling	126
4.1.1 Fonctionnement, fonction, enjeux	126
4.1.2 Googlebots	128
4.1.3 Affaiblir, rompre ou localiser	130
4.1.4 Fortifier ou spécifier	131
4.1.5 Optimiser, sonder et contrôler	136
4.2 Indexation	138
4.3 Appariement	140
4.3.1 PageRank	141
4.3.2 Taux de clics	142
4.3.3 Fraîcheur des documents	142

4.3.4 Efficacité de l'infrastructure	143
4.3.5 Qualité éditoriale	144
4.3.6 AuthorRank	145
4.3.7 Signaux sociaux	146
4.3.8 Tendances de Recherche	147
4.4 Interface	149
4.4.1 Liens simples	149
4.4.2 Liens enrichis	150
4.4.3 Liens moteurs verticaux	155
4.4.4 Liens de reformulation	156
4.5 Google Actualités	156
4.5.1 Intégrer	157
4.5.2 Communiquer	163
4.5.3 Optimiser	168
4.5.4 Préconisations éditoriales	172
4.5.5 Choix des rédactions	175
Chapitre 5. Le moteur comme lieu d'échange et d'opportunités	179
5.1 Addition des parcours et somme des possibilités	180
5.1.1 Interaction / coproduction	180
5.1.2 Bien public / positions privatives	182
5.1.3 Liste expérientielle	184
5.1.4 Arbre de navigation	185
5.2 Opportunités et incitations	187
5.2.1 Google et la publicité	187
5.2.2 Diversification concentrique	194
5.2.3 Les données : clef de voûte stratégique	197
5.2.4 Marché multiversant	199
5.2.5 Arbre des opportunités	211
5.3 Position des éditeurs	216
5.3.1 Environnement coopétitif	216
5.3.2 Les éditeurs et les effets de réseaux	218

5.3.3 Deux tentatives échouées	221
5.3.4 Dilemme de l'éditeur	225
Chapitre 6 : Le moteur comme terrain d'entente	227
6.1 Quelque part entre la loi et l'exception	227
6.1.1 <i>Fair use</i>	228
6.1.2 <i>Hot news misappropriation</i>	230
6.1.3 Une jurisprudence favorable	232
6.1.4 Triple test et droit d'auteur	233
6.1.5 Particularités françaises	235
6.1.6 Quel droit appliquer ?	237
6.1.7 Négociations en Union européenne	239
6.2 France : entente cordiale	252
6.2.1 Le meilleur ennemi	252
6.2.2 Projet de loi	254
6.2.3 Accord : 1 ^{er} volet	259
6.2.4 Accord : 2 ^{ème} volet	261
6.3 Tacite ou explicite : quel code, pour quel contrat ?	264
6.3.1 L'initiative ACAP	265
6.3.2 Protocole d'inclusion	268
6.3.3 Traduction dans l'oreille d'un sourd	274
6.3.4 Consensus léonin	279
Conclusion Partie 2 : le dispositif mis à plat	281
Champ de manœuvres	282
Lieu d'échange et d'opportunités	286
Terrain de négociations et d'entente	287
Partie 3 – Reconfiguration médiatique : effets et limites du dispositif	289
Chapitre 7 – L'organe de presse comme espace stratégique	291
7.1 Rationalisation du référencement	292
7.1.1 Dépendance	292
7.1.2 Situation d'urgence	297

7.1.3 Choisir d'internaliser	298
7.2 Le référenceur comme centre de traduction	300
7.2.1 Profils hétérogènes et fonction hybride	301
7.2.2 Rémunération	302
7.2.3 Place dans l'organisation	303
7.3 Le SEO comme négociation	305
7.3.1 Chapeaux noirs et chapeaux blancs	306
7.3.2 Les chapeaux gris de la presse en ligne	309
7.3.3 Tours et détours du dispositif	311
7.4 Organisation stratégique de L'Express	315
7.4.1 Réservoir de trafic, vitrine	315
7.4.2 Mutualisation des coûts, du <i>P&L</i> et du management	317
7.4.3 Méthode agile et coordination	319
7.4.4 Interventions hors ligne	321
Chapitre 8 : Le site de presse comme hôte du crawler	322
8.1 Devant et derrière la page	322
8.1.1 Adresses URL	323
8.1.2 Optimisation des titres	323
8.1.3 Disposition de la page	324
8.1.4 Méta-tags	324
8.1.5 Microdonnées	327
8.1.6 Robots.txt	330
8.1.7 Sitemaps XML	331
8.2 PageRank Sculpting : valorisation par le lien	335
8.2.1 Psychanalyse du PageRank	335
8.2.2 Egocentrisme hypertexte	343
8.2.3 Pages d'atterrissage	348
8.2.4 Liens invisibles	353
8.2.5 La stratégie du conjugueur	357
8.2.6 Mise en perspective	362

8.3 Dans les ramifications de l'arbre de navigation	365
8.3.1 Du bon côté du réseau d'incitations	365
8.3.2 Ligatus, marqueur d'indépendance	369
8.3.3 Référencement payant	371
8.3.4 Vente de liens	375
8.4 Sous la Coupole de L'Express	378
8.4.1 Un seul domaine	378
8.4.2 <i>Sitemaps</i> à tiroirs	382
8.4.3 Pages d'atterrissage mutualisées	384
8.4.4 AdWords	385
Chapitre 9 – Les journalistes et le journalisme chez RMF	387
9.1 Le pouvoir faire des journalistes	390
9.1.1 Pyramide inversée	390
9.1.2 Choix des sujets	393
9.1.3 Choix des mots	405
9.1.4 Dédoublage du titre	409
9.1.5 Bâtonnage	414
9.1.6 Du papier au web	416
9.1.7 Liens hypertextes	416
9.1.8 Temporalité de publication	421
9.1.9 Mises à jour	423
9.2 Rapports dialectiques	428
9.2.1 Meilleur ennemi n°1 : Google	428
9.2.2 Meilleur ennemi n°2 : le référenceur	435
9.2.3 Se préserver et se fixer une limite	439
9.3 Deux points d'achoppement	440
9.3.1 Des fissures à la Coupole	441
9.3.2 <i>Express Yourself</i> et les poissons-pilote	444
Conclusion Partie 3	450

Conclusion	455
Déploiement du dispositif	455
Rencontre d'éléments hétérogènes	456
Espaces stratégiques	458
Machine à faire dire (et à ne pas faire dire)	459
Machine à faire voir (et à ne pas faire voir)	462
Rapports de pouvoir	463
Lignes de fuite	464
Réassemblage du dispositif	465
Hypothèse 1	465
Hypothèse 2	467
Hypothèse 3	470
Hypothèse 4	471
Hypothèse 5	473
Problématique & hypothèse générale	474
Limites	475
Perspectives	478
Epilogue	481
Bibliographie	482
Annexes	510
A. Figures, tableaux et cadres	510
B. Carnets d'entretiens	513

Introduction

« This is an important moment in the history of search engines, the internet, and a networked global society. Search engines emerged as the epicenter of the early web, and represent a nexus of control for the future. Those who recognize their power, and are able to exercise control over it, will help to shape our collective future. »

(Halavais, 2009, p. 4)

Le 11 septembre 2001, les attentats du *World Trade Center* furent vécus en direct par plusieurs dizaines de millions d'individus. Une telle synchronisation résultait de l'action simultanée de caméras, micros, journalistes, amateurs, câbles, télévisions, radios, téléphones, ordinateurs disséminés et connectés — « chambre écho » (Derville, 1999) d'un événement tragique, symbolique et, par là même, immédiatement historique (Baudrillard, 2002 ; Derrida et Habermas, 2004). Nous vîmes les Tours Jumelles s'écrouler en direct.

L'infrastructure internet et le média web n'étaient plus, alors, des nouveautés. Les principaux titres de presse avaient mis en ligne des sites compagnons sur lesquels leurs contenus étaient diffusés. Les chaînes de télévision et les stations radio avaient lancé leurs plateformes, tandis que des entreprises de presse étaient nées en ligne, appelées *pure-players*. Au même titre que la télévision, la radio et la presse, le web fit ainsi partie de l'événement du 11 septembre.

Dans les heures qui suivirent les attentats, « où plus rien d'autre sur Terre ne semblait exister » (Mercier, 2009, p. 17), de nombreux internautes utilisèrent le moteur de recherche Google, sur lequel les requêtes les plus effectuées furent : « *cnn* », « *world trade center* », « *pentagon* », « *osama bin laden* » (Wiggins, 2001). Mais alors que le moteur était efficace pour trouver la page d'accueil de *CNN* ou la biographie de Ben Laden, il s'avéra qu'il ne permettait pas de trouver directement des informations à propos des attentats en cours (*ibid.*). L'internaute qui effectuait la requête « *world trade center* » accédait en effet à des sites dont aucun ne mentionnait la catastrophe. C'est pourquoi la firme Google invita les internautes, *via* des liens

hypertextes situés sous la barre du moteur, à se diriger vers d'autres sites. Il s'agissait à la fois de l'aveu d'une carence du dispositif en matière du traitement de l'actualité et de l'indication de ce qui, selon la firme, constituait des sources fiables : sites issus de médias traditionnels (*Washington Post*, *CNN*, *New York Times*, *ABC News*) et un site né en ligne (*Yahoo News*).

Tout en pointant vers ces sites, la page d'accueil de Google indiquait que la télévision ou la radio seraient sans doute plus appropriées pour obtenir de l'information en temps réel. Quelques heures plus tard, Google enleva ce message et multiplia les liens vers les sites issus de médias traditionnels. Les internautes pouvaient accéder au contenu des éditeurs « en cache », c'est-à-dire recopié et hébergé sur les serveurs de Google. Cette mesure permettait aux serveurs des éditeurs de ne pas surchauffer. Les dirigeants de Google avaient ainsi décidé de mettre leur infrastructure au service d'une action visant 1) à répondre au besoin d'information exprimé par les internautes 2) à pallier les carences techniques des sites des éditeurs. C'est un exemple probant de la manière dont « la médiatisation dans l'urgence, engendrée par le caractère inattendu du drame, influence les dispositifs et les processus de circulation sociale du savoir sur les faits » (Lamy, 2006, p. 108). Google ajouta également un lien vers la compagnie d'aviation *American Airlines*, dont les deux avions avaient été détournés. Le site pointait ainsi vers le contenu produit par des communicants, placé sur le même plan que les contenus produits par des journalistes. Le 12 septembre 2001, Google rassembla les liens sur une nouvelle page et laissa sous la barre du moteur un lien pointant vers elle appelé « *new links and support information* » – embryon de Google Actualités.

Ingénieur travaillant chez Google depuis 1999, Krishna Bharat expliqua à propos des attentats du 11 septembre : « Il y avait énormément d'informations et énormément de points de vue : le point de vue américain, le point de vue mondial, le point de vue afghan, le point de vue européen. C'était fascinant. J'ai réalisé que le web n'était pas pratique pour rechercher de l'information à propos d'un sujet que vous voulez comprendre entièrement. Chaque entreprise de presse a fait un sacré travail de récolte et de publication, mais elles n'avaient pas eu le temps, ou l'envie, de recouper leurs sujets avec les sujets des autres » (Vise, 2005, chap. 12).

Selon l'ingénieur, il manquait donc un outil capable de rassembler en un même énoncé plusieurs articles traitant d'un même sujet. Après le 11 septembre, Bharat décida de travailler à la création de cet outil en utilisant les 20 % de temps que Google offre aux employés désireux de mettre au point des projets innovants. Il établit un champ d'indexation restreint à quelques sites jugés fiables. Le moteur serait ainsi capable de scanner les contenus plus rapidement et régulièrement que ne pouvait le faire le moteur généraliste. Bharat utilisa ensuite la technique dite de « *clustering* » afin de regrouper en « grappes » (*clusters*) les contenus consacrés à un même sujet. Finalement, après une phase de test auprès des employés de Google, le PDG Eric Schmidt demanda à Bharat de se consacrer entièrement à son projet, lequel aboutit à la mise en ligne de la première version de Google Actualités mi-2002 aux Etats-Unis.

Google Actualités est caractérisé par un positionnement hybride entre l'attachement au système positiviste, anglo-saxon, selon lequel le but des organes de presse est de transmettre le récit de faits bruts, et la vision constructiviste, latino-européenne, selon laquelle « la réalité sociale, et donc l'information, sont conçues comme étant inévitablement le résultat de “constructions” et non comme de simples données de l'observation » (Delforce, 2010, p. 17). Selon Bharat en effet, « [Google Actualités] est l'opportunité d'avoir un intermédiaire objectif et algorithmique. Si vous avez confiance en l'algorithme pour croiser correctement les points de vue, alors il n'est plus gênant que ces points de vues soient biaisés » (Garber, 2011). Autrement dit, le dispositif permettrait selon son concepteur d'atteindre les visées du système positiviste en croisant des traitements propres au système constructiviste, et ainsi de compléter le travail effectué par les journalistes. Un tel projet s'appuyait sur des caractéristiques propres à l'infrastructure Internet et au média web : la « lecture diffractée » et le croisement de « journaux écrits, sites de stations radio ou de télévisions, dépêches d'agence et archives. Cela sonne le glas de l'unicité des sources d'information, de même que cela remet en cause profondément la linéarité de la lecture » (Mercier, 2009, p. 25). Nous pouvons nous demander dès lors à quel point, et selon quelles modalités, l'activité de Google et celle des journalistes se complètent *effectivement*, et nous interroger quant aux effets de leur cohabitation.

Ces questions sont d'autant plus problématiques que Google génère depuis octobre 2000 des revenus considérables sur le marché de la publicité (Battelle, 2005 ; Levy, 2009 ; Auletta, 2010 ; Pollock, 2010), tandis que les éditeurs de presse, dont la plupart cherchent à financer leur activité sur ce même marché, peinent à mettre en place un modèle économique viable (Smyrnaio, 2005 ; Attias, 2007 ; Sonnac, 2009 ; Ouakrat, 2011 ; Sonnac et Gabszewicz, 2013). Ainsi donc, Google et les éditeurs de presse en ligne sont en compétition sur le marché de la publicité, le premier y réussissant mieux que les seconds, tandis que les journalistes produisent un contenu vers lequel le moteur effectue des liens. C'est à la compréhension de cette situation mi-compétitive mi-coopérative, dite « de coopétition » (Smyrnaio et Rebillard, 2009, 2011¹) que nous avons œuvré.

La première partie de notre compte-rendu sera consacrée à la problématisation et à l'historicisation de notre objet. Dans le chapitre 1, nous présenterons notre ancrage disciplinaire et la construction de notre appareil théorique, après quoi nous introduirons notre problématique, nos hypothèses, les différents niveaux de notre terrain et notre méthodologie de collecte. Nous avons préféré intégrer ces différents éléments dans un chapitre à part entière, ce qui explique que notre introduction soit relativement courte. Notre premier chapitre continuera donc d'introduire notre objet tout en le problématisant, en situant épistémologiquement notre démarche et en décrivant dans le détail notre méthodologie de résolution. Les chapitres 2 et 3 seront quant à eux consacrés à une mise en perspective, cela car nous pensons, à l'instar de Franck Rebillard, qu'internet est le résultat d'une interpénétration d'éléments lui préexistant, lesquels « se marient et s'entremêlent progressivement, contribuant à former à terme un ensemble inédit » (Rebillard, 2007, p. 117). Ainsi, historiciser les mutations qui ont abouti à la situation analysée est selon nous une étape fondamentale de sa problématisation, permettant d'intégrer notre travail à une perspective « d'évolution cumulative de l'internet, fondée sur l'hybridation conflictuelle de logiques multiples de représentations, d'applications, et d'utilisations » (*ibid.*, p. 125). Nous procéderons par conséquent à une généalogie de la recherche

¹ La « coopétition » est un terme largement utilisée en sciences de gestion et en économie, mais « importé » en SIC par Franck Rebillard et Nikos Smyrnaio, lorsqu'ils s'intéressaient en particulier à la relation de Google et des éditeurs de presse en ligne.

d'information (chapitre 2) puis des modalités de production et de diffusion du contenu journalistique (chapitre 3). Après quoi nous aborderons le « dur » de la rencontre de ces deux univers en effectuant l'analyse descriptive de ce que les différents acteurs peuvent faire, de ce qu'ils ont intérêt à faire et de la manière dont les actions de Google et des éditeurs de presse sont coordonnées (Partie 2). Nous expliquerons comment Google Search et Google Actualités fonctionnent et décrirons les champs de manœuvres propres aux acteurs en présence (chapitre 4). Nous analyserons ensuite les opportunités et les incitations susceptibles de motiver leurs comportements (chapitre 5), ainsi que le type d'accord à l'issue duquel, bon gré mal gré, Google et les éditeurs sont à même de collaborer (chapitre 6). Nous aboutirons ainsi à une compréhension fine de la situation de communication et de ses effets potentiels, après quoi nous rendrons compte de nos constatations empiriques quant à ces effets (Partie 3). Nous avons mené pour cela 19 entretiens semi-directifs auprès de dirigeants, de journalistes et de membres des équipes techniques et commerciales de différentes entreprises de presse françaises, ainsi que de certains experts en référencement et d'employés de Google. Nous avons également observé et comparé 252 articles prélevés en octobre 2012 sur 6 sites issus de la presse imprimée, et avons effectué une monographie chez Roularta Media France, propriétaire entre autres de Lexpress.fr, où nous avons mené 21 entretiens semi-directifs supplémentaires². Ces données nous permettront d'analyser les changements organisationnels effectués autour de la question du référencement (chapitre 7). Nous décrirons également les effets des stratégies de captation sur les sites de presse et sur le traitement de l'actualité (chapitre 8). Puis nous expliquerons de quelle manière, à quel point et à quel prix les pratiques des journalistes peuvent permettre d'optimiser le référencement sur Google (chapitre 9). En conclusion, nous dresserons le bilan de nos apports et de leurs limites, avant d'en dégager des pistes qui gagneraient à être investiguées.

² Nous reviendrons dans le détail sur nos terrains et notre méthodologie de collecte dans le chapitre 1 (1.7).

Partie 1 — D'une configuration l'autre : problématisation et mise en perspective du dispositif

« Peut-être qu'en écoutant les choses, les rêves qui les précèdent, les délicats mécanismes qui les animent, les utopies qu'elles entraînent à leur suite, nous nous rapprochons du même coup des êtres qui les produisent, en usent et les échangent, tissant ainsi le collectif mixte, impur, sujet-objet qui forme le milieu et la condition de toute possibilité de communication. [...] Dans les hypertextes, tout fonctionne à la proximité, au voisinage. Le cours des phénomènes y est affaire de topologie, de chemins. Il n'y a pas d'espace universel homogène où les forces de liaison et de déliaison, où les messages pourraient circuler librement. Tout ce qui se déplace doit emprunter le réseau hypertextuel tel qu'il est, ou est obligé de le modifier. Le réseau n'est pas dans l'espace, il *est* l'espace. »

(Lévy, 1991, p. 61-69)

CHAPITRE 1. PENSER LE MOTEUR DE RECHERCHE COMME UN DISPOSITIF D'INFORMATION ET DE COMMUNICATION

« Tout travail de recherche sur la communication s'avère solidaire d'une certaine conception de l'homme et de la société. Il dépend d'un ensemble de présupposés et de prémisses théoriques qui conditionnent à la fois le regard qu'on porte sur celle-ci et la capacité à l'analyser. »

(Olivesi, 2006, p. 181)

1.1 Reconfiguration des voies d'accès

1.1.1 Publier et publiciser

Plusieurs chercheurs ont comparé le rôle de Google à celui des *gatekeepers* traditionnels (Machill *et al.*, 2004 ; Diaz, 2008 ; Röhle, 2009), car le moteur classe, choisit et traite l'information (Hindman *et al.*, 2003) à la place d'acteurs qui auraient perdu au moins partiellement leur pouvoir en la matière (McQuail, 2005 ; Carlson, 2007 ; Bruns, 2008). La comparaison avec les *gatekeepers* doit toutefois être

tempérée par le fait que Google ne soit pas en mesure de contrôler l'accès des informations à l'espace public. N'importe qui peut en effet publier une information sur le web, et l'activité de Google a trait non à la publication mais à la *publicisation*, c'est-à-dire à l'inscription des contenus à l'avant-scène d'un continuum où certains propos sont plus publics que d'autres (Cardon, 2010). Le moteur est ainsi un amplificateur plutôt qu'un gardien, et quand un éditeur cherche à y optimiser son référencement, c'est du succès de sa production dont il s'agit plutôt que de son accès à l'espace public.

1.1.2 Enjeux du référencement

La plupart des éditeurs de presse ont opté pour un modèle gratuit pour les internautes, et se financent grâce à l'affichage de publicités sur leurs pages (Herbert et Thurman, 2007 ; Ouakrat *et al.*, 2010 ; Ouakrat, 2011). Ainsi, pour être rentable, un site doit atteindre un certain seuil d'audience. Dès lors, étant donné la position de carrefour d'audience de Google dans un contexte d'étranglement de la distribution-diffusion (Hesmondhalgh, 2007 ; Smyrniotis et Rebillard, 2009), il existe un enjeu économique fort autour de la question du référencement, presque aussi important que l'enjeu lié à l'effet de marque (Sonnac, 2009).

Le deuxième enjeu du référencement concerne la légitimité des journalistes. En effet, dans un espace où leur production côtoie celles des communicants et des amateurs, les journalistes peuvent chercher à gagner en visibilité dans le but d'être des leaders d'opinion : « écrire pour être lus » (Siméant, 1992, p. 40). Or, puisque Google est un des principaux, sinon le principal, régisseurs de l'échelle de visibilité du web, la mise en lumière de la production journalistique par ses services pourrait constituer un moyen de « labelliser l'information-presse par rapport à toutes les autres » (Wolton, 2003, p. 14).

Cette idée de labellisation, ajoutée aux enjeux d'ordre économique, pose une question fondamentale : le webjournaliste, pour être lu par les utilisateurs des services de Google, pourra-t-il écrire ce qu'il souhaite et comme il le souhaite, ou devra-t-il accepter de faire des concessions quant à la « vision du monde » (Zelizer, 1992, p.12) qu'il voudrait communiquer ? Comment sa *crédibilité* et sa *visibilité*,

ainsi que la *rentabilité* de l'entreprise qui l'emploie, peuvent-elles aller de pair dès lors qu'il cherche à ce que l'information qu'il produit soit trouvée ? Ces questions doivent être considérées dans un contexte où les journalistes sont « de plus en plus contraints d'osciller entre logique industrielle et marchande d'un côté et logique citoyenne de l'autre ; [...] écartelés entre l'obligation de respecter des impératifs de vente, d'audience, et leur souci d'assurer au mieux l'autonomie de leur pensée » (Rieffel, 2008, p.103). La presse en ligne est en effet marquée par une « culture du clic » (Anderson, 2009) susceptible d'influencer la mise en récit de l'actualité.

1.1.3 La question du prix à payer

N'importe quel site peut être indexé sur le moteur généraliste Google Search. En revanche, il existe certains critères à respecter pour qu'un éditeur puisse intégrer le champ d'indexation de l'agrégateur et moteur vertical³ Google Actualités. Par ailleurs, il est possible de mettre en œuvre des méthodes dites de « *Search Engine Optimization* » (SEO) dans le but de maximiser la visibilité des documents. Ces méthodes font l'objet de compétences particulières et demandent en général qu'un spécialiste soit consulté ou embauché. Elles concernent aussi bien les aspects techniques et structurels du site, que certains aspects éditoriaux. Ainsi, les enjeux liés au référencement peuvent potentiellement avoir une influence sur l'organisation du travail, la structure du site et la ligne éditoriale. Il y aurait donc une sorte de « prix à payer » technique et éditorial pour recevoir une audience significative, dès lors qu'il s'agirait pour l'éditeur de chercher à être trouvé.

Certains travaux universitaires se sont intéressés à la manière dont la question du référencement pouvait influencer le processus de production de l'information et, finalement, l'information elle-même. Astrid Mager a par exemple expliqué comment certains éditeurs de contenus d'ordre médical pouvaient utiliser le maillage hypertexte afin d'être mieux référencés (Mager, 2009, 2010). Theo Rhöle a quant à lui présenté la manière dont Google pouvait instaurer vis-à-vis des éditeurs un « régime disciplinaire » et fortifier ainsi une « norme de publication » (Rhöle, 2009). Cependant, dans les deux cas, il s'agit de

³ Le moteur généraliste Google Search indexe l'ensemble du web accessible, tandis qu'un moteur vertical comme Google Actualités, qui est aussi un agrégateur (puisque'il propose des listes de liens sans qu'aucune requête n'ait été effectuée), a un champ d'indexation restreint *a priori*.

rendre compte des effets de l'optimisation sur l'usage des liens hypertextes, et non sur les sujets traités, le choix des mots, les niveaux de navigation, les pratiques de ceux qui rédigent les documents et l'organisation de l'entreprise qui les emploie. Il semble donc qu'il y ait encore à faire de ce côté, et notamment pour ce qui est du cas particulier des entreprises de presse, car même si certains auteurs intéressés par le journalisme l'ont évoqué (Robinson, 2007 ; Carlson, 2007), et si certains journalistes ont eux-mêmes publié à ce sujet (Richmond, 2008 ; Carr, 2010 ; Mettout, 2010), il est rarement question dans la littérature scientifique de la nature et du résultat des stratégies de référencement de la presse en ligne. La seule étude consacrée exclusivement à cette question est à notre connaissance celle que Murray Dick a menée auprès des rédactions de *BBC News*, du *Guardian* et de *Northcliffe* (Dick, 2011). D'après ses conclusions, la culture du référencement a beau ne pas être « dogmatique » dans ces entreprises (*op. cit.*, p. 474), elle y est visible, et des pressions peuvent être exercées sur les pratiques autant que sur l'éthique des journalistes⁴. Il y aurait donc bien un « prix à payer » pour les journalistes. Murray Dick souligne finalement que la visée de son travail est exploratoire et que celui-ci devra être complété, ce que nous avons précisément cherché à faire dans ce travail.

1.2 Etudier l'infomédiation en SIC

Le moteur généraliste Google Search et le moteur/agrégateur Google Actualités récoltent les informations disponibles sur le web et génèrent des méta-informations à leur sujet : leur date de mise en ligne, leur position les uns par rapport aux autres, leur vitesse de chargement... Les informations récoltées et les méta-informations générées sont ensuite ordonnées et communiquées aux internautes désireux de combler un besoin d'information. Compris ainsi, le moteur est une « méta-forme » (Johnson 1997, p. 32) qui « joue en même temps le rôle de guide, de filtre, de pare-choc et de traducteur. [Il] nous aide à classer l'abondance d'éléments dans des cases, nous propose des orientations et des interprétations toutes faites » (Rieder, 2006, p. 83). Franck Rebillard et Nikos Smyrnaiois désignent de telles méta-formes, dans le cas particulier de l'accès à l'information d'actualité, par le vocable

⁴ Nous discuterons des conclusions de Murray Dick à de nombreuses reprises, notamment dans le chapitre 9, où nous pourrions les comparer aux nôtres.

infomédiaire (Rebillard et Smyrnaio, 2010). Les infomédiaires, expliquent-ils, se situent à mi-chemin entre l'édition et la diffusion de la production journalistique : ils mettent en avant un contenu journalistique qu'ils n'ont pas produit, mais se positionnent également à un niveau méta-éditorial, c'est-à-dire que la mise en avant du contenu lié à l'actualité se fait selon des procédés de traitement, de hiérarchisation et de présentation qu'ils ont déterminés en partie ou totalement. C'est cette méta-éditorialisation, ainsi que leur position de carrefours d'audience, qui fait dire à Marc-Olivier Goyette-Côté que « [l]es infomédiaires ont un poids structurant important dans la définition des usages en cours de formation sur la Toile en influençant la manière dont circule l'information d'actualité et les manières dont elle peut être valorisée » (Goyette-Côté, 2012).

Le web peut exister indépendamment de Google, cependant la place qu'y occupe la firme peut provoquer une modification de la manière dont les acteurs y agissent et, finalement, de la nature de ce qu'on y trouve et de la combinaison indifféremment technique et sociale des processus dont il est le média. De la même façon, si un topographe dessine une carte d'une zone qui n'avait jusqu'à lui jamais été cartographiée, celle-ci pourra conduire les habitants de cette zone à changer leur façon de se déplacer ainsi qu'à tracer de nouveaux chemins ou à bâtir de nouveaux bâtiments. Un énoncé produit dans le but de se repérer peut ainsi aboutir à ce que la réalité désignée soit modifiée, réinterprétée, déplacée, traduite — auquel cas un nouvel énoncé devra être produit pour rendre compte des transformations que le premier aura entraînées. En pragmatique du langage, on dit que l'énoncé n'est pas *constatif* mais *performatif*⁵, c'est-à-dire qu'il contribue à instaurer ce dont il rend compte (Austin, 1970 ; Denis, 2006). Une « partie de ping-pong » se joue alors entre traductions et interprétations, énoncés et réalités, réalités et énoncés, habitants et topographe, dont l'information et son objet constituent indifféremment les stimuli, les causes et les résultats. C'est cette partie de ping-pong et sa *performativité* que nous proposons de jauger dans le cas de la firme Google et du processus d'infomédiation induit par ses outils. Nous présenterons ci-après l'ancrage d'une telle démarche dans le champ francophone des Sciences de l'Information et de la Communication (SIC).

⁵ Nous reviendrons dans le détail sur la définition de ce terme en 1.3.5.

1.2.1 Attachement à la dimension technique

Jean Davallon définit la principale particularité du point de vue du chercheur en SIC comme l'attachement de la pratique scientifique à la dimension technique de son objet (Davallon, 2004). En d'autres termes, il s'agit pour le chercheur de s'intéresser aux moyens et aux supports, infrastructures, interfaces, contraintes techniques, pratiques d'écriture, sans lesquels l'information et la communication ne seraient pas. L'adoption d'un tel point de vue consistera dans notre cas à étudier les moyens et les supports *par* lesquels, *avec* lesquels et parfois *malgré* lesquels l'information produite par les journalistes peut être communiquée aux utilisateurs des services de Google, et ce qu'une telle situation induit en matière de traitement de l'actualité d'une part, et de structuration du champ informationnel d'autre part.

Tandis que les sciences sociales se sont assez tardivement intéressées aux techniques de communication (Flichy, 1997 ; Jouët, 2000), les SIC ont largement contribué à combler ce manque, attachées à décrypter ce qui permet et ce que soutient « l'action communicationnelle » (Miège, 2004). Inscrire nos travaux dans ce champ reviendra par conséquent à révéler dans le cas du moteur « l'architecture d'un objet technique » (Proulx, 2005) ainsi qu'à analyser chez les éditeurs « la rationalisation et même l'industrialisation de la phase de conception » (Miège, 2006, p. 169). Ce faisant, nous veillerons à étudier l'information et la communication « en tant qu'elles sont matérialisées, institutionnalisées et opérationnalisées dans la société » (Davallon, 2004, p. 35). Autrement dit, plutôt que de nous contenter de décrire une technique de communication, nous essaierons d'analyser comment elle s'insère dans les rapports entre entreprises et entre individus, comment (et à quel point) elle contribue à les façonner, et ce que, finalement, elle *signifie*.

1.2.2 Approche communicationnelle du web

L'attachement à la dimension technique passera par l'étude du fonctionnement du moteur de recherche en particulier mais aussi, plus généralement, par l'étude des techniques qui permettent au web de fonctionner. Le fait de ne pas s'intéresser à la seule technique du moteur est justifié dans la mesure où celle-ci a précisément pour

but de prendre connaissance et d'interpréter, de *traduire*, l'usage effectué par une multitude d'acteurs des techniques propres au web en général.

Schématiquement, le web⁶ s'appuie sur une couche physique, nommée internet, permettant de transmettre et de recevoir de l'information, une couche logicielle permettant de la traiter, et une couche éditoriale composée par l'information elle-même. Nous considérerons ces trois couches dans notre étude, et la manière dont le moteur peut les traduire mais aussi, et autant, la manière dont ceux qui se saisissent des techniques propres à chacune des couches souhaitent que le moteur traduise leurs actions. Nous nous intéresserons bien sûr aux techniques d'écriture des journalistes : le choix des mots, de la syntaxe, la disposition des paragraphes, des images et des intertitres, la hiérarchisation, le rubriquage. Par ailleurs, une des originalités de notre approche consistera à prêter une attention particulière aux balises HTML attachées aux contenus comme autant d'« instruments de communication » (Lyons, 1978, p. 33), c'est-à-dire que nous nous intéresserons aux « outils qui permettent l'existence de l'écrit à l'écran et qui, non contents de représenter la structure du texte, en commandent l'exécution et la réalisation » (Jeanneret et Souchier 1999, p. 103). Yves Jeanneret, Emmanuel Souchier et Jean Davallon nomment de tels outils des « architextes », expliquant que « le texte naît de l'architexte qui en balise l'écriture » (*ibid.*, p. 103) et qui assure « la greffe du texte et du réseau » (Davallon et Jeanneret, 2004, p. 53).

Comme tout langage, le code informatique « ne se réduit pas à un regard sur le monde. Il est aussi action » (Lochard et Soulages, 2006, p. 238). En effet, ce que *dit* le code a précisément pour but de *faire* quelque chose : présenter un site et son contenu d'une certaine manière, communiquer avec le navigateur, le moteur de recherche, les logiciels, l'infrastructure. En cela, le code est *performatif* par excellence, puisque son objectif n'est pas de constater une situation mais de susciter une réalité. Dès lors, le contexte d'énonciation (codification) et l'énoncé (code) sont inséparables dans l'analyse de *ce qui performe* et de *ce qui est performé*. Nous verrons dans notre cas que les balises HTML implémentées par un éditeur sont susceptibles de *faire* quelque-chose au moteur de recherche, mais aussi, et pas moins,

⁶ Ce que nous appelons le « web » fait référence à l'environnement informationnel s'appuyant sur le protocole Http et nécessitant, pour l'individu qui souhaiterait y accéder, d'utiliser un navigateur.

à son usager, à la firme Google, à l'éditeur et à l'information elle-même. Ainsi, en plus d'analyser la codification comme un langage, nous l'étudierons comme une action et une intention. Car en effet, comme n'importe quel « acte de parler », l'acte de coder « *opère* dans le champ d'un système linguistique ; il met en jeu une *appropriation*, ou une réappropriation, de la langue par les locuteurs ; il instaure un *présent* relatif à un moment et à un lieu ; et il pose un *contrat avec l'autre* dans un réseau de places et de relations » (De Certeau, 1980, p. 38). Nous irons donc voir ce qu'il y a « derrière » les sites web autant que ce qu'il y a « devant », dans le but de comprendre pourquoi et comment les éditeurs « s'adressent » au moteur, et comment le moteur « comprend » ce qu'ils cherchent à *provoquer*. Nous nous intéresserons à la fois à ce qui est codé, ce qui code, ce qui décode, dans quel contexte, ainsi qu'au lien social créé, modifié, suscité, rompu ou renforcé par ces actions/intentions de communication. Nous réfléchirons par exemple à la façon dont les balises peuvent être utilisées comme des outils de *contractualisation*, et de quelle manière, et à quel point, le code peut avoir valeur de loi (Lessig, 1999) tout en décryptant ce que le fonctionnement du moteur *fait* de cette loi, à cette loi, par cette loi. Une telle démarche est conforme au projet des SIC, en cela qu'elle consiste à « prendre au sérieux l'épaisseur des différentes médiations dont la communication peut faire l'objet » (Davallon et Jeanneret, 2006) et à considérer le média comme étant « à la fois produit et producteur de langage et de lien social » (Davallon, 1992, p. 103). En somme, le code est une langue *vivante* dont nous étudions l'usage effectué par les éditeurs, l'interprétation opérée par Google, le résultat auquel chacun de ces interlocuteurs souhaiterait parvenir et le résultat auquel, effectivement, ils parviennent. Chercheur en Sciences de l'Information et de la Communication, ce dont nous rendrons compte, au final, c'est d'une *conversation*.

1.2.3 Du côté de la production

Notre travail de thèse est consacré aux logiques de production de l'information, et non de sa réception. Il aurait été intéressant de problématiser la réception, c'est-à-dire de nous intéresser aux internautes qui consultent Google pour s'informer à propos de l'actualité, cependant cela n'aurait pas été possible dans le

temps imparti. Par ailleurs, à l'instar d'Eric George, nous croyons qu'« il y a presque toujours antériorité de l'offre du moins entre les producteurs et les concepteurs d'une part et le grand public d'autre part. Quels que soient les raffinements en matière de prise en compte des usagers du côté de l'offre, et ce même en amont des filières, il n'en demeure pas moins que dans le contexte d'appropriation des dispositifs techniques, matériels et logiciels, les usagers grand public sont toujours en situation de réponse par rapport à une offre, même modifiée après des rencontres entre concepteurs, producteurs et usagers » (George, 2004, p. 7-8).

Dès lors, nous rejoignons un courant des SIC appelé « théorie des industries culturelles et médiatiques »⁷ (Augey et Rebillard, 2009) dont les chercheurs justifient leur intérêt pour l'offre en cela qu'ils considèrent, pour leurs objets en particulier, que « l'influence de l'émetteur est en général déterminante » (De la Haye et Miège, 1984, p. 33) et s'efforcent par conséquent de « penser la tension qui, continuellement, alimente la dynamique industrielle de la culture et de l'information » (Mœglin, 2012).

Même si les internautes jouent un rôle certain (que nous mentionnerons à plusieurs reprises), il se trouve que l'essentiel du processus d'infomédiation est déterminé par les négociations et les interactions ayant cours entre Google et les éditeurs. Il y a, pour ainsi dire, « prééminence de l'offre » (Wolton, 2000, p. 68). Les internautes pourront ensuite ne pas agir comme cela avait été prévu et influencer le résultat du processus d'infomédiation, mais ils ne se déplaceront pas moins dans un complexe communicationnel où « ce qui fait système entre les composants du complexe » (Davallon, 2004, p. 34) ne dépend que très peu de leurs actions comparativement à celles de Google et des éditeurs. C'est pourquoi nous croyons, à l'instar de Nikos Smyrnaio, que dans le cas de la presse en ligne et de l'infomédiation de ses contenus, « si l'émetteur-producteur n'est pas omnipotent, il influe néanmoins nettement sur les comportements, et il modèle, ou du moins tente de modeler, la demande » (Smyrnaio, 2005, p. 23).

⁷ Il est possible de dire également « industries de la culture et de la communication » (Bouquillion, 2008).

1.2.4 Cadrage interdisciplinaire

Les SIC ont la particularité d'être une « interdiscipline » (Bougnoux, 1995 ; Mucchielli 2001 ; LeCoadic 2004 ; Miège, 2004). Sur le site web de la 71^{ème} section du Centre National Universitaire, elles sont d'ailleurs présentées comme « un champ résolument interdisciplinaire ». Elles « récupèrent des dimensions d'analyse nées dans les disciplines déjà existantes. Mais elles les retravaillent en liaison avec d'autres dimensions, en veillant toujours à articuler les différents niveaux du phénomène » (Perret, 2004, p. 125). Cette particularité nous permettra d'aborder sans *a priori* l'analyse de ce que nous avons présenté dans notre avant-propos, à partir des travaux de Yochai Benkler (2006, p. 17), comme le produit contingent d'une *combinaison* d'éléments hétérogènes. En effet, c'est pour se saisir de la multiplicité des facettes de son objet que le chercheur en SIC articule plusieurs disciplines. Bien sûr, nous pensons, à l'instar de Robert Boure, qu'il serait « absurde de s'intéresser à tout – sous peine de dilution du savoir » (Boure, 2000, p. 9). Cependant, nous verrons que l'environnement communicationnel de l'infomédiation ne peut pas être compris sans faire appel à plusieurs disciplines, notamment à la sociologie des techniques, l'économie, le droit, la sociologie du journalisme et la sociologie des organisations. Ainsi, nous tenterons de relier autour de notre axe des méthodologies provenant de différents champs dans le but de les faire « interagir » (Miège, 2000, p. 558). Notre problème sera alors de « trouver la voie difficile de l'entre-articulation entre des sciences qui ont chacune, non seulement leur langage propre, mais des concepts fondamentaux qui ne peuvent pas passer d'un langage à l'autre » (Morin, 1990, *cit. in*. Rieder, 2006, p. 21). Etant donné une telle contrainte, il est certain que nous ne pourrions pas adopter une posture interdisciplinaire sans « accepter les limites de [notre] propre épistémologie » et « renoncer à la tentation explicative générale » (Monnoyer-Smith, 2007, p. 17). Nous n'expliquerons donc pas *tout* ce qui concerne la relation entre Google et les éditeurs de presse, mais ce que cette relation implique en matière de communication.

1.2.5 Capacitation et incitations

C'est à la seule condition d'étudier ce que l'environnement informationnel induit en matière de capacité d'action pour chacun des acteurs concernés que nous pourrions comprendre leurs déplacements et, finalement, rendre compte de la manière dont l'information est produite par les journalistes et mise en lumière (ou non) par Google. Bernhard Rieder explique dans sa thèse, de façon parallèle à la notion d'espace de possibilités d'action constitué par un logiciel (Yoneyama, 1997, p. 114), qu'il est possible de « parler de l'architexte comme l'instance productrice d'un espace de possibilités d'écriture qui ouvre certaines pistes et en ferme d'autres » (Rieder, 2006, p. 209). C'est précisément à ces possibilités d'écriture que nous nous intéressons dans le cas des éditeurs de presse en ligne : Que peuvent-ils faire ? Que savent-ils faire ? Qu'ont-ils intérêt à faire ? Que font-ils ? A quel prix ? Pour quels résultats espérés et effectifs ? En d'autres termes, nous nous intéressons à la capacitation (*empowerment*) des éditeurs et à ses effets. Il s'agira en particulier de décrire la marge de manœuvres des journalistes en matière d'optimisation, et la manière dont ce champ d'action peut être concilié, ou au contraire entrer en conflit, avec ce qu'ils considèrent comme étant un traitement adéquat de l'actualité.

Nous prêterons également une attention particulière aux « incitations » concernant les différents acteurs jouant un rôle dans le processus d'infomédiation, et la manière dont ces incitations peuvent s'accorder ou être contradictoires, ainsi que la manière dont elles peuvent, étant donné les marges de manœuvres des acteurs en présence, influencer le procès de communication et son résultat. La théorie économique sera particulièrement intéressante pour développer ces questions, grâce aux notions de « jeu », d'« arbre de décision », d'« effets de réseaux croisés » et d'« utilité ». Nous verrons ainsi en quoi les phénomènes de l'information-communication « apparaissent aujourd'hui comme fortement dépendants de décisions, de stratégies ou d'influences relevant de la sphère économique » (Miège, 2006, p. 163). D'autre part, nous décrirons de quelle manière les acteurs peuvent agir selon ce qu'ils savent des incitations des autres, en plus des leurs propres. Nous pourrions ainsi d'autant mieux décrypter les relations de *pouvoir*, dès lors que nous aurons identifié ce qu'il est théoriquement possible de faire pour les uns et les autres et ce

qu'il est souhaitable pour chacun de faire faire aux uns et aux autres, étant donné la géographie des intérêts, opportunités et capacités d'action.

*

Nous devons à présent construire un appareil théorique qui nous permettra d'équiper et d'opérationnaliser le concept, vague sinon, de champ interdisciplinaire. C'est ce cadre qui définira la méthode de collecte et fournira les clefs analytiques de notre compte-rendu. A ce stade, notre problématisation est assez avancée pour savoir que nous devons grâce à notre appareil théorique :

- attacher la pratique scientifique à la dimension technique de notre objet, tout en nous permettant de dépasser la seule considération technique ;
- étudier un complexe hétérogène fait d'humains, d'infrastructures, de code HTML, de représentations, d'intérêts, d'opportunités et de moyens ;
- articuler l'apport de plusieurs disciplines ;
- nous situer du côté de la production pour analyser l'infomédiation « en train de se faire » ;
- analyser la capacitation des éditeurs au sens large, et des journalistes en particulier, ainsi que la manière dont leur environnement est balisé ;
- analyser les modifications induites par la stratégie de référencement mises en place par les entreprises de presse ;
- analyser le pouvoir des uns et des autres, et ses effets ;
- identifier ce qui peut inciter les acteurs à agir de telle ou telle façon, impactant alors plus ou moins directement la nature et le résultat de leurs actions.

1.3 Anthropologie d'une configuration médiatique

Les ingénieurs de Google créent et mettent en ligne des outils qui constituent, comme toute technique, une mise en tension entre la volonté d'élargir le champ des possibles et la tentation de réduire l'environnement à n'être qu'un milieu conditionné (Bachimont, 2010, p. 175). Le processus de production journalistique est lui aussi l'objet d'innovations techniques, et donc de la même mise en tension. D'autre part, la

nature de son produit fini, c'est-à-dire de la mise en récit et en circulation du fait d'actualité, est influencée par les acteurs dont ces innovations sont le fait (Boczkowski, 2004).

Etant donné ces caractéristiques, nous nous sommes tournés vers les outils théoriques et les méthodes de la sociologie des techniques. En particulier, nous avons utilisé les clefs fournies par la « sociologie de la traduction » ou « théorie de l'acteur-réseau »⁸ (ANT ci-après, pour *Actor Network Theory*). Lorsque nous avons opté pour cet équipement théorique, souvent mobilisé en SIC, il avait déjà été choisi pour des travaux concernant aussi bien le processus de production journalistique (Plesner, 2009 ; Anderson, 2009 ; Schmitz Weiss et Domingo, 2010) que la mise en œuvre et les effets des moteurs de recherche (Rieder, 2006 ; Röhle, 2009 ; Mager, 2010), ce qui nous a encouragé dans cette voie.

1.3.1 Théorie de l'acteur-réseau

Les principaux théoriciens français de l'ANT sont issus du Centre de Sociologie de l'Innovation de l'Ecole des Mines et ont proposé, à partir du début des années 80, une nouvelle approche selon laquelle la production scientifique ne devait plus être considérée comme étant nécessairement l'explication de plus en plus juste d'une réalité absolue, mais comme une pratique faite d'incertitudes, de sentiments, de lieux communs, de réflexes, de hasards, de négociations et d'arts de faire, parmi différents individus plus ou moins proches dans l'espace et le temps (Latour, 1999a, 2005). Puis ils ont étendu ce cadre à l'étude des techniques (Woolgar, 1991), considérant que le succès d'une innovation ne pouvait pas être expliqué par le fait que celle-ci soit tout simplement « plus efficace que ce qui existait avant elle », mais

⁸ Nous allons nous référer tout au long de notre thèse à des notions forgées « par », « dans » et « pour » ce qu'on nomme aujourd'hui, selon certaines nuances, « théorie de l'acteur-réseau », « sociologie de l'acteur-réseau », « sociologie actantielle », « sociologie de la traduction », « ontologie de l'actant-rhizome » ou « sociologie des associations ». Toutes ces appellations font référence au même cadre théorique, lequel a été renommé selon les désirs et les besoins de ceux qui s'en saisissaient. L'un des pères-fondateurs de la théorie, Bruno Latour, a lui-même écrit à propos du terme « acteur-réseau » qu'il était devenu inapproprié (Latour, 1999b). De notre côté cependant, nous croyons que ces changements de nom, quand bien même ils seraient justifiés, sont contre-heuristiques. Le nom n'est pas la théorie qu'il désigne : ainsi qu'importe sa *précision* du moment que l'on sait *avec précision* ce à quoi il fait référence. Les débats à son propos constituent à notre avis une perte de temps dans la mesure où ils tournent autour de questions d'ordre épistémologique sans y répondre jamais, ni même clairement les formuler. Par ailleurs, l'existence de plusieurs noms rend difficile la comparaison, l'archivage et la cumulativité de travaux se prévalant d'un même positionnement. C'est pourquoi nous avons choisi de prendre systématiquement le terme qui semblait être le plus utilisé en France et à l'étranger : théorie de l'acteur-réseaux, dont le diminutif est ANT.

comme le résultat de négociations, de détournements, de hasards, de rencontres et de méthodes de qualification et d'évaluation. L'étude des sciences et des techniques a ainsi adopté une position qui consiste à étudier les scientifiques et les ingénieurs « en action » — ce qu'ils font, ce qu'ils prévoient, ce qu'ils pensent et les objets qu'ils mobilisent. Ce faisant, l'ANT propose de révéler les stratégies qui ne seront plus visibles une fois l'action achevée et son résultat stabilisé sous forme de « fait scientifique » ou d'« artefact », de « mobile immuable », jusqu'à ce qu'une innovation ou une controverse ne rouvre la « boîte noire » constitutive d'une telle stabilisation.

La racine étymologique du mot « information » sous-entend que quelque-chose a été « mis en forme », et l'objet des SIC est précisément d'étudier d'une part ce qui a permis cette mise en forme, comment, qui, pourquoi, pour qui, et, d'autre part, ce que devient cette mise en forme, comment elle est mobilisée, parfois remise en forme, par qui, à quelles fins. En nous plaçant du côté de la production, nous proposons d'étudier la mise en forme de l'actualité « en action », de la même manière que l'ANT étudie la production scientifique ou l'ingénierie « en action ». C'est la raison pour laquelle nous reprenons les méthodes de l'anthropologie des sciences : non pas dans le but d'étudier la production du fait scientifique, mais pour étudier la production du fait d'actualité. En effet, les journalistes, comme les scientifiques, doivent produire certaines *inscriptions* qui sont *mises en circulation* et n'agissent pas indépendamment des événements dont elles rendent compte, mais constituent au contraire « la condition même de leur existence » (Nora, 1972, p. 162). « Les journalistes ne font pas advenir une réalité inexistante, mais leur mise en scène de l'information contribue à doter un fait d'une valeur relative, en lui donnant une plus grande visibilité, au point parfois d'écraser tous les autres faits » (Mercier, 2009, p. 17). Ainsi, nous aborderons la mise en énoncé d'un événement non pas en décrétant qu'il s'agit d'un événement nécessairement plus important que les autres, mais en considérant que pour « faire l'actualité » l'énonciation de cet événement a dû faire l'objet de négociations, de détournements, de hasards, de rencontres, d'arts de faire et de méthodes d'évaluation.

Du point de vue épistémologique, l'ANT a la particularité de ne pas s'intéresser aux seuls individus mais également aux objets en tant qu'*actants*⁹, remplaçant ainsi les notions de « Société » et de « Contexte » par celle de « collectif produit par des humains et des non-humains » (Callon, 2006, p. 272). C'est la force ou la faiblesse des associations entre les êtres et les choses, leur *efficacité*, que le chercheur analyse plutôt que de distinguer *a priori* deux groupes, l'un qui serait social et déterminant, l'autre qui serait technique et déterminé (Callon et Latour, 2006). Dès lors, l'ANT est assez proche de ce qu'en philosophie on nomme la « métaphysique orientée objet » (Harman, 2002 ; Garcia, 2010). Selon Bruno Latour en effet, l'ANT est une « sociologie orientée-objet à l'usage d'humains eux-mêmes orientés-objet » (Latour 2007, p. 106). Elle nous permettra de considérer la dimension technique de notre problème en préconisant « dans l'analyse des processus communicationnels et informationnels, la prise en compte simultanée des réseaux sociaux, des acteurs et des artefacts » (Serres, 2002).

« S'il est possible, écrit Bruno Latour à propos de l'activité scientifique, de décrire la même activité comme étant si puissante et pourtant si marginale, si concentrée et si diluée, cela signifie qu'elle a les caractéristiques d'un réseau » (Latour, 2005, p. 431). A partir de cette observation, les théoriciens de l'ANT ont défini le concept de *réseau* comme un *liquide interstitiel vivant*¹⁰ qui lierait entre eux des actants hétérogènes, davantage qu'une « chose qui se trouverait là et qui aurait vaguement la forme d'un ensemble de points interconnectés » (Latour, 2007, p. 189). C'est parce qu'il est hybride que le réseau est qualifié de « sociotechnique » (Callon, 2006), et parce que les acteurs, quand ils agissent, mobilisent tout ou partie d'un réseau, qu'on parle d'*acteur-réseau*¹¹. A partir d'un tel postulat, il nous faudra

⁹ Ainsi, selon Bruno Latour, « aux éternelles passions, trahisures et autres stupidités des hommes et des femmes, il nous faut ajouter l'entêtement, la ruse et la force des électrons, des microbes, des atomes, des calculateurs et des missiles » (Latour, 2006, p. 88).

¹⁰ Bruno Latour parle de « plasma » (Latour, 2007, p. 348-354).

¹¹ L'ANT est fortement influencée par le travail de Michel Serres et du « diagramme en réseau » avec lequel celui-ci a introduit la série des *Hermès* (Serres, 1969, p.11-20). Ce diagramme, formé d'une pluralité de points reliés entre eux par une pluralité de ramifications, permet en effet de briser « définitivement la linéarité des concepts traditionnels : la complexité n'est plus un obstacle à la connaissance ou, pis, un jugement descriptif, elle est le meilleur des adjutants du savoir et de l'expérience » (*op. cit.*, p.20). Dans d'autres ouvrages, Michel Serres parle de « contrat », dont nous pouvons également rapprocher la notion de réseau comme elle est définie par l'ANT : « Contrat signifie qu'un collectif quelconque tire ensemble quelque chose : une charrue, un fardeau. Pour le faire, il faut des liens qui associent les tracteurs et ceux-ci à la chose tractée » (Serres, 1994, p.132). Ainsi l'acteur-réseau selon Michel Callon et Bruno Latour et le diagramme/contrat selon Michel Serres sont des concepts poursuivant un même but : faire de la complexité un moteur pour le chercheur plutôt qu'une

considérer qu'« il n'y a pas de déplacement possible sans l'établissement de toute une connectique coûteuse et fragile qui ne vaudra qu'à la condition de l'entretenir régulièrement et dont la force ne sera jamais plus grande que celle de son plus petit maillon » (Latour, 2012, p.73), et que c'est l'établissement de cette connectique qui permettra que la mise en récit d'un événement par un journaliste soit classée en haut des listes générées par les services de Google et y « fasse » l'actualité.

1.3.2 Traductions

L'ANT nomme *traduction* « l'ensemble des négociations, des intrigues, des actes de persuasion, des calculs, des violences grâce à quoi un acteur ou une force se permet ou se fait attribuer l'autorité de parler ou d'agir au nom d'un autre acteur ou d'une autre force » (Callon et Latour, 2006, p.13). Elle est un « processus avant d'être un résultat » (Callon, 1986, p. 205), à l'issue duquel des *porte-parole* sont désignés. Etudier la traduction, dans notre cas, reviendra à étudier *tout ce qui* permet que la firme Google puisse être chargée de la publicisation de la production journalistique, et à étudier les phénomènes sociaux que son moteur de recherche *traduit* en termes techniques et sous forme de liste hypertexte. De toute évidence, cela ne peut avoir lieu sans embûches : des négociations sont menées, des confrontations ont lieu, des désaccords perdurent. Google et les éditeurs font certaines concessions et procèdent à certains détours et détournements dans le but de collaborer et de satisfaire leurs intérêts propres.

En plus de son sens linguistique, l'ANT donne à la traduction « le sens géométrique de translation. Parler de traduction d'intérêts signifie à la fois que l'on propose de nouvelles interprétations et que l'on déplace des ensembles » (Latour, 2005, p.284)¹². L'ANT nous encourage donc à identifier ce qui est *interprété* et *déplacé* par les concepteurs du moteur, les journalistes et le moteur lui-même.

entrave, et tracer grâce à elle les associations permettant à un collectif d'agir tout en conditionnant la nature et le résultat de son action.

¹² Nous retrouvons cette double signification en anglais où le verbe traduire (*translate*) désigne à la fois une forme d'interprétation (*translation*) et de déplacement (*translation*).

1.3.3 Programmes d'actions

Aucun concepteur n'imagine un objet sans lui imaginer une utilité et sans penser quelque-chose à propos de celui qui utilisera cet objet. On parle alors de « sociologie implicite » (Bardini, 1996) et d'« ingénieur-sociologue » (Callon, 1987). Parce qu'il prévoit à son objet une utilité, le concepteur y *inscrit* sa vision du monde. Le résultat est ce que l'ANT nomme un « script » et définit comme « la mise en forme technique, par le concepteur, de son point de vue sur les relations nécessaires entre son objet et les acteurs qui doivent s'en saisir. [Cette mise en forme] se veut une prédétermination des mises en scène que les utilisateurs sont appelés à imaginer à partir du dispositif technique et des prescriptions (notices, contrats, conseils...) qui l'accompagnent » (Akrich, 2006, p. 161). Nous retrouvons ici l'idée exprimée en SIC selon laquelle les technologies de communication sont des « objets intentionnels : fabriqués par des individus et adressés à d'autres » (Esquenazi, 2006, p. 22). Cependant, selon l'ANT, le script n'est pas l'objet mais plutôt une supposition conçue à partir de ce que le concepteur sait, veut, apprécie, déteste, croit, etc. Ce sera ensuite à l'utilisateur de se saisir ou non de l'objet, à sa manière, comme il l'entend et le peut, selon ses expériences, ses objectifs, ses intérêts et représentations : en somme, son script à lui. L'usage pourra dès lors être différent de ce qui avait été prévu, comme il pourra ne jamais être. Ainsi, le script n'est pas l'action mais son cadre ou *programme d'action* : « série d'objectifs, d'états et d'intentions qu'un agent peut parcourir » (Latour 1999a, p. 178). C'est un ensemble de possibles qui n'est pas actualisé tant que l'objet n'a pas été saisi et que le script n'a pas rencontré d'autres scripts, d'autres possibles, d'autres intérêts.

Pour faire état de la chaîne de traductions qui nous intéresse, il nous faudra étudier ce que Google prévoit à propos de ce que sera sa relation avec les éditeurs. Puis il nous faudra considérer la manière dont les éditeurs actualisent le projet de la firme en fonction de qu'ils savent, pensent et veulent eux-mêmes. Finalement, les clefs données par l'ANT nous permettront de décrire la façon dont le processus d'infomédiation prévu par Google se recompose autour des intérêts de ceux qui y prennent part, ainsi que d'analyser les effets d'une telle recomposition en matière de traitement de l'actualité.

1.3.4 Enrôlement et gouvernementalité

La traduction est un processus que l'on peut déployer, pour une analyse plus fine, en quatre étapes, détaillées par Michel Callon dans un article célèbre concernant la pêche des coquilles Saint-Jacques (Callon, 1986). Il y décrit de quelle manière trois chercheurs ont tenté d'intéresser les marins-pêcheurs à leur projet d'étude, lequel visait à savoir s'il était possible de maîtriser la domestication d'une certaine variété de Saint-Jacques. Il fallait pour cela que les chercheurs persuadent les pêcheurs de collaborer à leur projet, en leur assurant que celui-ci permettrait d'assurer leurs profits à moyen et long termes, et qu'ils réussissent également à « convaincre » les coquilles qu'ils seraient capables de les aider à se reproduire en les protégeant de leurs prédateurs. Ils formulèrent une seule et même question à laquelle tous les acteurs, eux y compris, avaient intérêt qu'une réponse fût apportée : *La variété de coquilles Saint-Jacques Pecten Maximus se fixe-t-elle ?* Michel Callon nomme cette première étape du processus de traduction « la problématisation » et la question formulée « Point de Passage Obligé » (PPO), c'est-à-dire le point par lequel tous les acteurs doivent passer s'ils veulent parvenir à leurs fins. Les chercheurs doivent ensuite montrer à l'ensemble des actants concernés qu'ils ont intérêt à ce que cette question soit résolue. Dans ce cas, ils accepteront la définition que les chercheurs ont donnée de leurs objectifs. Cette définition est une interprétation d'une situation donnée, parmi d'autres interprétations possibles (éventuellement concurrentes), à partir de laquelle un système d'alliances est projeté. Michel Callon nomme cette seconde étape « l'intéressement ». « Intéresser, explique-t-il, c'est se placer entre » (Callon, 1986, p. 185).

La troisième étape de la traduction désigne « le mécanisme par lequel un rôle est défini et attribué à un acteur qui l'accepte » (*ibid.*, p. 189) et se nomme « l'enrôlement ». Les chercheurs deviennent les représentants des intérêts des marins-pêcheurs et des coquilles Saint-Jacques, qui acceptent de collaborer à leur projet, quitte à faire pour cela certaines concessions, certains détours, et même à modifier leurs conditions de vie : les marins-pêcheurs ne pêcheront plus dans la zone de l'étude tandis que les Saint-Jacques vivront dans des collecteurs artificiels. A la suite de négociations multilatérales, les chercheurs deviennent les « porte-parole » au sein

d'un collectif mêlant indifféremment des individus, des bateaux, des coquilles, des collecteurs, des articles scientifiques, des élus locaux, etc. Les différentes entités du réseau acceptent de jouer le rôle que les chercheurs leur assignent. N'étant pas forcés pour autant de jouer ce rôle, ils peuvent à tout moment cesser de collaborer. Mais puisqu'ils ont été intéressés à la résolution du problème formulé par les chercheurs, ils se sont enrôlés librement dans le but de poursuivre leurs propres intérêts. Ainsi l'enrôlement, explique Callon, est un « intéressement réussi ». Les chercheurs, dès lors qu'ils auront réussi à « enrôler d'autres volontés en traduisant ce qu'elles veulent et en réifiant cette traduction de manière à ce qu'aucune d'elles ne puissent vouloir autre chose » (Callon et Latour, 2006, p. 28), pourront « faire faire » certains détours aux marins comme aux coquilles. Toutefois, et nous tenons à insister sur ce point, il ne faut surtout pas considérer l'enrôlement comme une mesure totalisante, mais comme la définition d'un terrain d'entente auquel les acteurs adhèrent librement, quitte à effectuer certaines concessions dans la mesure où ils considèrent que leurs intérêts pourront être davantage satisfaits que s'ils n'y avaient pas adhéré.

Nous retrouvons dans le concept d'enrôlement une idée proche de ce que Michel Foucault a nommé la « gouvernementalité » (Foucault, 1994), néologisme « barbare mais incontournable » (Barthes, 1970) désignant la capacité à conduire les conduites d'autres acteurs. Selon Foucault, les acteurs, quels qu'ils soient, jouissent toujours d'une marge de manœuvres, même infime, susceptible de remettre en cause l'autorité que l'un d'entre eux prétend exercer. Le pouvoir quant à lui « s'exerce plutôt qu'il ne se possède » et constitue l'effet d'un ensemble de « positions stratégiques – effet que manifeste et parfois reconduit la position de ceux qui sont dominés » (Foucault, 1975, p. 35). Ainsi, dès qu'il y a un rapport entre des acteurs, il y a une zone de frottement et un pouvoir qui « ne s'applique pas purement et simplement, comme une obligation ou une interdiction, à ceux qui "ne l'ont pas" ; il les investit, passe par eux et à travers eux ; il prend appui sur eux, tout comme eux-mêmes, dans leur lutte contre lui, prennent appui à leur tour sur les prises qu'il exerce sur eux » (*ibid.*). Une fois qu'ils ont réussi à enrôler les marins-pêcheurs et les coquilles, les chercheurs exercent une forme de pouvoir à leur endroit leur permettant de demander aux marins de ne pas pêcher dans la zone de l'étude et aux coquilles de

s'y fixer. En ce qui concerne notre objet, c'est cette forme de pouvoir et ce qu'il induit qu'il s'agira d'identifier dans le cas de Google et des éditeurs.

La dernière étape du processus de traduction est la « mobilisation ». Cette étape indique que la traduction a réussi : les chercheurs parleront au nom des pêcheurs, des coquilles Saint-Jacques et de la communauté scientifique, qui agiront conformément à ce qui leur a été demandé. Dans notre cas, Google sera en charge de publiciser et de hiérarchiser l'information produite par les éditeurs. Dans l'exemple extrême d'une traduction parfaite, les internautes ne s'informeront jamais plus par d'autre biais que par celui de Google et les éditeurs construiront systématiquement leurs contenus de manière à aller dans le sens de ce qui est préconisé par la firme, quitte à effectuer pour cela de nombreux déplacements. Cependant, une telle traduction pourrait être contestée à tout moment dans la mesure où un acteur peut « trahir » le réseau en n'agissant plus que dans ses seuls intérêts. La « force »¹³ du porte-parole s'étioule alors immédiatement, car il ne peut obliger les dissidents à rentrer dans le rang. Parce que ces derniers ne sont plus intéressés, ou parce qu'ils ne le sont plus assez, l'enrôlement prend fin, et avec lui la mobilisation. La gouvernementalité n'opère plus. Dans le cas d'étude de Michel Callon, alors que tout avait fonctionné durant la première année, l'expérience est renouvelée l'année suivante et les chercheurs constatent à leur grand désarroi que les coquilles ne se fixent plus. Quant à celles qui s'étaient fixées, elles sont pêchées la veille de Noël par des marins intéressés par le profit à court terme davantage que par l'expérience des chercheurs. Dans notre cas, un éditeur pourra désindexer ses contenus du moteur de recherche dès qu'il le souhaitera ; et dans le cas extrême où tous les éditeurs procéderaient à une telle opération, le moteur n'aurait plus de raison d'être, puisqu'il ne pourrait plus aiguiller les internautes. Ainsi, les éditeurs ont du pouvoir eux-aussi, et peuvent menacer l'action de Google s'ils estiment que ce n'est plus leur intérêt d'y collaborer¹⁴. Google doit donc faire en sorte de continuer à les intéresser.

¹³ La force, selon l'ANT, relève d'une capacité à associer des éléments hétérogènes et à intervenir sur, et par, le réseau d'actions ainsi formé. En effet, selon Michel Callon et Bruno Latour, un acteur « est d'autant plus solide qu'il peut associer fortement le plus grand nombre d'éléments – et, bien sûr, dissocier d'autant plus rapidement les éléments enrôlés par d'autres acteurs. La force, c'est donc le pouvoir d'interrompre ou d'interrelier. La force, c'est plus généralement l'intervention, l'inter-ruption, l'inter-prétation, l'intérêt comme l'a magistralement démontré Serres (Serres, 1980). Un acteur est d'autant plus fort qu'il peut intervenir davantage » (Callon et Latour, 2006, p. 25).

¹⁴ Nous verrons comment ce fut le cas, notamment dans le chapitre 6.

1.3.5 Performativité

L'ANT sera d'autant plus heuristique dans notre cas qu'elle nous permettra, grâce à la notion de *performativité*, de nous saisir de ce que nous avons présenté comme une « partie de ping-pong » dont l'énoncé et son objet sont à la fois les stimuli, les causes et le résultat (cf. 1.2.1). La performativité est un concept forgé au départ par le philosophe du langage John Austin (Austin, 1970) pour désigner la manière dont un énoncé peut agir sur ses objets plutôt que de simplement en rendre compte. John Austin a eu beau abandonner cette notion, celle-ci a largement été reprise, révisée, opérationnalisée, discutée et complétée (Denis, 2006). En France, Pierre Bourdieu a consacré à la linguistique une critique célèbre où il préconisait de prendre en compte dans l'analyse de la performativité la position des énonciateurs dans l'espace social et leurs rapports de pouvoir (Bourdieu, 1982). C'est à partir de ce constat que l'ANT a repris la notion à son compte et proposé un « programme de la performativité » (Muniesa et Callon, 2009, p. 294-295). Michel Callon (1998), Donald McKenzie, Yuval Millo (2003) et Fabian Muniesa (Muniesa et Callon, 2009) ont mis en avant la performativité des énoncés scientifiques, ainsi que la nature collective de cette performativité et le caractère hybride, indifféremment humain et non-humain, sociotechnique, des *agencements* qui la rendent possible¹⁵. C'est en circulant que des inscriptions, des énoncés, des faits, des théories, des formules, des cartes, des graphiques, des typologies, gagnent en réalité et se stabilisent. C'est ainsi que les énoncés « arrivent à durer, c'est-à-dire à s'inscrire dans le monde (par l'intermédiaire d'objets, de textes, de dispositifs techniques complexes) et que leur performativité s'accomplit » (Denis, 2006, p. 6). Cependant, pour circuler le long des chaînes de traductions, il leur faut nécessairement passer par certains détours, être modifiés ou complétés, de manière à tisser un réseau d'associations suffisamment solide autour du projet qu'ils portent. C'est *en circulant* et *pour circuler* que l'énoncé fera preuve de performativité. L'ANT s'éloigne ainsi « d'une épistémologie classique qui oppose le monde des énoncés et le monde (plus ou moins réel) des choses auxquelles les énoncés réfèrent et qui constituent en quelque sorte le contexte de ces

¹⁵ Le travail le plus connu a porté sur l'étude de la performativité des sciences économiques (*economics*) sur l'économie réelle (*economy*) (Muniesa et Callon, 2009)

énoncés. Les références ne sont pas extérieures à l'univers des énoncés : elles circulent avec eux et avec les inscriptions dont ils sont issus » (Callon, 2006, p. 269).

Considérer la performativité dans notre cas reviendra à considérer que les énoncés et les méta-énoncés produits par Google peuvent agir sur les informations produites par les éditeurs et sur les conditions et les modalités de leur production, de leur mise en ligne et de leur valorisation, les informations produites par les éditeurs pouvant elles-mêmes performer les conditions d'énonciation des méta-énoncés. Il nous faudra donc nous intéresser aux détours par lesquels la firme Google doit elle-même passer pour que la performativité de ses énoncés et la performativité des énoncés des éditeurs ne nuisent pas à ses intérêts. Ainsi, l'énoncé produit par les concepteurs (algorithmes et recommandations), le méta-énoncé généré par le moteur (liste de liens) et l'énoncé produit par les éditeurs (textes, architextes, hypertextes) sont susceptibles de se performer les uns les autres.

1.3.6 Médiation sociotechnique

Pour fonctionner, le moteur Google fait appel à des techniques dites de crawling, d'indexation, d'appariement algorithmique et de design¹⁶. Les éditeurs, en plus de certaines techniques d'écriture, peuvent mettre en œuvre des techniques de « tagging », implémenter certains « Sitemaps XML », méta-tags, balises, microdonnées, liens hypertextes¹⁷. Ces techniques sont au cœur du processus d'infomédiation, tandis que les outils qui les rendent possibles sont des « médiateurs » au sens de l'ANT, c'est-à-dire qu'ils peuvent influencer de manière décisive le processus et son résultat (Latour, 2007). Les pratiques sous-tendues induisent « une manière de penser investie dans une manière d'agir, un art de combiner indissociable d'un art d'utiliser » (De Certeau, 1980, p. XLI). Dès lors, il s'agit de voir ce que ces techniques permettent de faire, et ce qu'elles font effectivement, mais également ce que la technique *dit* : quelles valeurs elle transporte, quelle vision du monde, quels énoncés. Car en effet, dès lors que nous considérons que les « techniques ne sont rien que des discours » (Latour, 1996,

¹⁶ Nous décrirons ces techniques dans le chapitre 2 (Partie 1) et le chapitre 4 (Partie 2).

¹⁷ Nous décrirons ces techniques dans les chapitres 4 (Partie 2), 8 et 9 (Partie 3).

p. 43), il s'agit de décrypter ces discours, l'intention de leurs émetteurs et leurs effets, leur performativité. La notion de *médiation technique* sera à cette fin particulièrement utile dans la mesure où elle « renvoie à cette épaisseur, cette profondeur de la technique » (Akrich, 1993, p. 88) et permet au chercheur de décrire la manière dont l'action est composée à partir d'éléments hétérogènes. Nous en présenterons ici les quatre dimensions, à partir de la description effectuée par Bruno Latour (1993, 1999a) et de ce que Bernhard Rieder y a ajouté dans sa thèse (2006).

Première dimension : *l'interférence*, c'est-à-dire la rencontre des scripts. Prenons l'exemple d'un individu qui voudrait conduire une voiture pour aller quelque part. Il ne suffit pas à l'homme de vouloir conduire une voiture, encore faut-il qu'il puisse le faire : il doit avoir pour cela un permis (*savoir* et *pouvoir* conduire) et un véhicule, lesquels une fois réunis lui *permettent* de conduire. Ainsi, lorsque l'homme et la voiture se rencontrent, chacun des scripts rencontre un autre script avec lequel il doit composer. Le potentiel d'action du conducteur ne se réduit alors ni à son envie de conduire ni à sa capacité de conduire ni à la voiture. « Les capacités, objectifs et démarches émergent de la rencontre concrète et spécifique entre l'homme et [la voiture]. Voilà pourquoi on ne peut parler d'un déterminisme, ni social, ni technique. Chaque acteur apporte son propre script et le résultat n'est jamais réductible à l'un d'entre eux » (Rieder, 2006, p.103).

Deuxième dimension : le *principe de composition*. En agissant, les actants donnent naissance à un acteur-réseau, dont la capacité n'est pas limitée à sa composante humaine ni à un seul de ses composants. « L'action composée est distribuée sur la chaîne formée par les actants et elle doit son origine à l'ensemble du réseau d'*attachements* qui se met en place. La possibilité qu'un élément soit à un moment plus dominant qu'un autre n'invalide pas le fait que l'action est toujours constituée par une pluralité de forces » (*ibid*, p. 104).

Troisième dimension : la *mise en boîte noire*. Une fois les acteurs et actants attachés les uns aux autres et l'ensemble stabilisé, il devient impossible de les séparer l'un de l'autre. Dans notre exemple, quand une voiture roule à grande vitesse, on ne voit pas le conducteur ni ce qu'il fait dans la voiture. Seul un accident peut arrêter la voiture et dissocier à nouveau les éléments qui composent l'action. C'est pourquoi

l'ANT préconise d'étudier les innovations ou les controverses plutôt que des ensembles déjà stabilisés.

Quatrième dimension : la *délégation*. « Lorsque nous parlons de délégation, nous entendons la substitution d'un acteur à un autre dans le contexte d'une tâche à effectuer ou d'une responsabilité à prendre en charge. Dans notre cas, cet « autre » est un objet technique. Pour comprendre ce qui est délégué à un « personnage » technique, Latour nous dit que nous n'avons « qu'à imaginer ce que d'autres humains ou d'autres non-humains auraient à faire si ce personnage n'était pas en place » (Latour 1993, p. 58) » (Rieder, 2006, p.105). Si, dans la voiture de notre conducteur, il y a une sonnerie stridente qui retentit et un voyant qui s'allume à chaque fois qu'il n'a pas mis sa ceinture de sécurité, on ne peut considérer la sonnerie et le voyant ni comme des purs signes ni comme des pures choses : ils parlent au conducteur qui, même s'il ne veut pas mettre sa ceinture, voudra peut-être se défaire de cette sonnerie et de ce clignotant horripilants. Ainsi, la sonnerie et le voyant sont des artefacts signalétiques qui *performent* l'action du conducteur (Denis et Pontille, 2010). On dira, selon l'ANT, que le concepteur du véhicule a *délégué* au voyant et à la sonnerie une fonction faite de matière et de sens, visant à obliger le conducteur à mettre sa ceinture. Cela peut être lié à la volonté de l'ingénieur de limiter les risques d'accidents, ou bien à une loi qui lui impose de prévoir un tel dispositif.

S'il est vrai que l'ingénieur délègue, l'utilisateur délègue lui aussi : le conducteur dit à la voiture « emmène-moi ici » et la voiture fonctionne, sans quoi il devrait se déplacer à pied. Le concepteur ne peut pas savoir à l'avance de quelle délégation sa machine fera l'objet. « S'il est vrai, explique Madeleine Akrich, que la plupart des choix opérés lors de la conception peuvent se décrire comme des décisions portant sur ce qui va être délégué à qui ou à quoi, cela signifie que l'objet technique contient et produit une certaine géographie des responsabilités ou plus généralement des causes » (Akrich, 2006, p. 162). Nous nous intéresserons précisément à cette géographie dans le cas de Google et des éditeurs (considérés alors comme des usagers de techniques mises à leur disposition par Google) : qui est responsable de quoi, quand, à quel moment, vis-à-vis de qui et à quel point, étant donné quels intérêts, quelles influences, quelles pressions, pour quelle cause et dans quel écosystème

informationnel ? C'est pour aborder d'autant mieux ces questions, indissociables de la question du *pouvoir*, que nous articulerons les préceptes théoriques de l'ANT avec le concept de dispositif.

1.4 Le concept de dispositif appliqué à l'infomédiation

Dans un texte de présentation des sciences de l'information-documentation, Jean-Paul Metzger explique que « le dispositif est la composante instrumentale du système d'information. C'est l'instrument que sollicite et utilise le public pour bénéficier du service. Cet instrument peut être examiné de plusieurs points de vue. Il s'agit d'abord d'un ensemble de procédés qui, mis en œuvre, concourent à faciliter l'accès à la documentation et l'information. Il s'agit corrélativement d'un ensemble d'outils construits pour que ces procédés puissent être mis en œuvre. Il s'agit aussi d'une organisation humaine et sociale qui élabore ces outils et qui agit selon ces procédés. Il s'agit enfin d'un acteur socio-économique qui se positionne et intervient dans un contexte social, économique et culturel » (Metzger, 2006, p. 54). Cette définition a cela d'intéressant qu'elle englobe les procédés, les outils, les organisations et les acteurs qui, ensemble, donnent lieu à une situation de communication. Ainsi, dans le cas de Google et des éditeurs de presse, nous désignerons par « dispositif » aussi bien les caractéristiques techniques du moteur et des sites de presse (procédés qui concourent à faciliter l'accès à l'information) que la firme Google et les groupes de presse (organisations humaines et sociales qui élaborent ces outils et agissent avec leur concours), les ingénieurs et les journalistes (qui se positionnent et interviennent dans un contexte social, économique et culturel). Cependant, nous ne pouvons pas nous suffire de la définition donnée par Jean-Paul Metzger, car c'est précisément à cause d'une définition qui ne consiste qu'à faire état de l'hétérogénéité que le concept de dispositif, très à la mode en SIC, a souvent été galvaudé, trahi, sclérosé (Gavillet, 2010). C'est pourquoi nous reviendrons ici sur la genèse du concept créé par Foucault et repris par Deleuze, avant de présenter l'opérationnalisation qui selon nous est la plus pertinente et de considérer ce que celle-ci implique et peut révéler dans le cadre de notre étude.

1.4.1 Foucault, Deleuze, l'ANT et les SIC

L'histoire du concept commence avec une citation de Michel Foucault désormais canonique. Le dispositif y est présenté par le philosophe comme « un ensemble résolument hétérogène, comportant des discours, des institutions, des aménagements architecturaux, des décisions réglementaires, des lois, des mesures administratives, des énoncés scientifiques, des propositions philosophiques, morales, philanthropiques, bref : du dit, aussi bien que du non-dit [...] [dont] chaque effet positif ou négatif, voulu ou non voulu, vient entrer en résonance, ou en contradiction, avec les autres, et appelle à une reprise, à un réajustement, des éléments hétérogènes » (Foucault, 1994, p. 299). Ainsi, le concept est employé pour désigner des ensembles pluridimensionnels, donnant au chercheur une certaine liberté quant aux terrains investigués et aux outils employés. Il semble qu'une telle plasticité ait fortement contribué à son succès (Peeters et Charlier, 1999, p. 15). Mais souvent, le terme a été repris à Foucault dans une définition partielle, attachée à la seule hétérogénéité dont les vertus heuristiques étaient provoquées plutôt que véritablement justifiées (Gavillet, 2010). Car le dispositif, selon Foucault, ne désignent pas seulement des éléments hétérogènes mais également, et autant, le réseau qui s'établit entre ces éléments hétérogènes. Aussi ne suffit-il pas de faire état de l'hétérogénéité des éléments mais d'étudier ce qui lie ces éléments et pas d'autres, comment, pourquoi, quand, où, pour combien de temps et à quel point. « [E]ntre ces éléments, discursifs ou non, il y a comme un jeu, des changements de position, des modifications de fonctions, qui peuvent, eux aussi, être très différents » (Foucault, 1994). D'autre part, le dispositif foucauldien prétend répondre à une situation d'urgence, et possède dès lors une « dimension stratégique dominante » dans la mesure où certains acteurs tenteront de positionner et d'agréger les éléments autour de leurs intérêts propres. Enfin, le concept est indissociable des notions de pouvoir et de rapports de force. La nature du pouvoir exercé dépendra en effet de l'agencement du dispositif, c'est-à-dire du champ d'action (la *force*) de ceux qui l'exercent et de ceux par qui il s'exerce.

Logiquement¹⁸, l'ANT a largement repris la notion de dispositif en y juxtaposant l'épithète « sociotechnique » — le dispositif sociotechnique étant défini par elle comme « un assemblage d'actants humains et non humains où les compétences et les performances sont distribuées » (Akrich et Latour, 1992, p. 259). C'est de la distribution de ces compétences et de ces performances (ces *forces*) que dépendra la géographie du pouvoir au sein du réseau, qui est toujours et nécessairement selon l'ANT un réseau de pouvoir (Hughes, 1979, 1983 ; Law, 1986, 1991). Les dispositifs « font des choses. Ils articulent des actions ; ils agissent ou en font agir d'autres » (Muniesa *et al.*, 2007). On y mesure la solidité des associations, on y identifie ce qui peut les affaiblir ou les rompre, en cherchant à déterminer la capacité de certains actants à infléchir ou modifier le cours d'action d'autres actants.

Gilles Deleuze a à la fois précisé et assoupli le concept de dispositif de Michel Foucault. Il a notamment proposé de considérer les dispositifs comme des machines à faire voir et à faire dire : « Les deux premières dimensions d'un dispositif, ou celles que Foucault dégage d'abord, ce sont des courbes de visibilité et des courbes d'énonciation. [Les dispositifs] sont des machines à faire voir et à faire parler [...] » (Deleuze, 1989, p. 185). Ainsi, selon Deleuze, le dispositif produit des énoncés et agit en tant que régime de véridiction. « Un dispositif, expliquent Julia Bonaccorsi et Virginie Julliard en reprenant et en opérationnalisant l'apport de Deleuze, se définit par les régimes d'énoncés qu'il fait naître » (Bonaccorsi et Julliard, 2010, p. 5). Nous retrouvons ici un concept proche de celui de la performativité, selon lequel un énoncé peut devenir plus vrai de par son effet performatif (Muniesa et Callon, 2009, p. 296). En outre, Deleuze n'omet pas la dimension fondamentale des travaux de Foucault : le pouvoir. Le dispositif deleuzien est en effet un lieu de monstration des rapports de forces. « Gilles Deleuze insiste principalement sur l'idée que le pouvoir relève d'une stratégie davantage que d'une propriété : il s'exerce plus qu'il ne se possède. Il relève

¹⁸ « C'est sans doute le terme [dispositif] qui convient le mieux pour désigner tous ces assemblages sociotechniques d'humains et de non-humains auxquels s'intéressent ces sociologues, qu'il s'agisse de décrire les « programmes d'action » (Latour, 1996) ou les « scripts » (Akrich, 2006) inscrits dans des objets, ou encore d'accorder à ces derniers le statut de « médiateurs », capables d'introduire de la différence, d'ajouter ou de retirer quelque chose aux actions, et d'en modifier le cours (Hennion et Latour, 1993). Et que serait la traduction sans dispositifs pour la rendre matériellement possible, sans ces assemblages d'éléments hétérogènes d'énoncés, d'agencements techniques, de compétences incorporées qui font les « chaînes de traduction » (Callon, 1995, pp. 50-51) » (Beuscart et Peerbaye, 2006)

d'une relation. Il n'a pas d'essence, et prend justement corps au sein du dispositif, lorsqu'il réalise la fonction pour laquelle il a été créé » (Badouard, 2012, p. 54).

Cependant, et son principal apport est sans doute celui-là, Gilles Deleuze déplace la notion de rapports de forces en rendant possible, par le concept de « ligne de fuite », le dépassement du dispositif (Monnoyer-Smith, 2013). C'est dans cette dimension que s'exprime la volonté « que le dispositif ne soit pas complètement cristallisé par les relations de pouvoir et laisse la possibilité à l'individu de s'exprimer en tant que sujet. Les acteurs peuvent mettre en place des stratégies pour contester les lignes de partage du pouvoir » (Mabi, 2011, p. 11). A condition de faire preuve de créativité, des acteurs dégagent certains espaces à l'intérieur desquels ils sont affranchis des pressions que pourraient exercer d'autres acteurs à leur endroit, ou bien ils font en sorte de « siphonner » le pouvoir de ces acteurs pour l'exercer désormais, en partie ou totalement, à leur place. Quand le chercheur en sciences sociales fait référence à ce concept, c'est la prise en compte de cette créativité, et le fait de penser le dispositif *en mouvement*, qui lui permettront de ne pas enfermer *a priori* les acteurs « dans des lignes infranchissables que sont celles de l'organisation stratégisée du pouvoir » (Monnoyer-Smith, 2013). Nous avons ici une prise en considération subtile du pouvoir, qui nous semble être en mesure de décrire comment celui-ci va être opéré et mis en scène au sein d'un dispositif où aucun schème d'action n'est purement contraint et où les acteurs, en se saisissant de leurs marges de manœuvres, vont être à même de rééquilibrer, ou de contrebalancer, leurs rapports mutuels. En somme, les acteurs, où qu'ils se situent, ont la capacité de subjectiver le dispositif existant, c'est-à-dire de se l'approprier pour le dépasser, le modifier ou le *fuir*. « [O]n a beau faire, hiérarchiser, énoncer, produire des lois et des institutions, des normes et des comportements, le dispositif *fuit* » (Monnoyer-Smith, 2013).

En SIC, le concept a été largement mobilisé, et cela d'autant plus en France depuis qu'un numéro entier de la revue *Hermès* y a été consacré (1999, n°25). Il est courant en effet de trouver des travaux concernant « les dispositifs d'information et de communication » (Appel *et al.*, 2010). Cependant, les chercheurs qui utilisent ce terme ne le font pas de la même façon, dans le même but ou dans les mêmes conditions. Pour notre part, nous nous positionnons dans la lignée des travaux

effectués en particulier par Julia Bonaccorsi et Virginie Julliard (2010), Laurence Monnoyer-Smith (2013), ainsi que les travaux de thèse de Romain Badouard (2012) et de Jean-Christophe Plantin (2012), qui se sont référés au dispositif foucaldien dans sa relecture deleuzienne, en veillant à donner toute son épaisseur à la dimension technique des phénomènes décryptés et en effectuant systématiquement le lien avec l'ANT. Le concept a grâce à eux montré son utilité dans l'étude des situations de communication aussi bien pour décrire la manière dont certains réseaux hétérogènes font parler et font voir l'information, que pour l'analyse des rapports de forces et des lignes de fuite dont un espace de médiation constitue l'épicentre. Par ailleurs, nous tenons à souligner que dès lors que ces chercheurs ont opérationnalisé un concept provenant de la philosophie pour en tirer des clefs de lecture empiriques, d'autres interprétations sont possibles aussi bien des travaux de Foucault que de ceux de Deleuze. Notre référence au dispositif doit donc être justifiée dans le cas de notre objet en particulier, afin qu'aucune ambiguïté ne puisse gêner la compréhension des développements qui suivront.

1.4.2 Dispositif d'infomédiation

Google Search et Google Actualités ont pour vocation de générer des liens hypertextes vers des documents, qu'ensuite ils hiérarchisent et publient via leurs interfaces respectives. Ainsi, le dispositif fait voir les contenus produits par les éditeurs. Il montre un traitement de l'actualité qui n'est plus seulement celui des journalistes, mais également le sien avec les leurs, les leurs à travers le sien. Il s'agit de répondre à l'urgence créée par la quantité de documents mis en ligne et la difficulté de s'y repérer. En guise de réponse à cette situation, le moteur propose « un environnement producteur de *feed-back* immédiat par rapport à l'action des usagers, capable de traiter l'imprévu et la nouveauté » (Peeters et Charlier, 1999, p. 17).

Le résultat de l'infomédiation est conditionné en partie par ce que les ingénieurs croient être une bonne manière de s'informer à propos de l'actualité, et affectée par ce qui, à leurs yeux, peut faire qu'une source, un document ou un auteur sera plus pertinent qu'une autre source, un autre document, un autre auteur. Le dispositif porte en cela un projet technique et éditorial (Bonaccorsi et Julliard, 2010). La « conception » de l'actualité (l'idée

que les concepteurs s'en font) intègre la « conception » du moteur (ce qu'ils font concrètement) et sous-tend l'énonciation de l'actualité une fois que des internautes se seront saisis du dispositif (ce qu'ils auront fait finalement). Ainsi, le dispositif ordonne des espaces de véridiction qui existent indépendamment de lui (l'internaute peut accéder à ces contenus sans passer par Google), mais qui, dès lors qu'un internaute utilise Google Search et/ou Google Actualités, sont traversés par lui et le traversent. Un nouvel espace de véridiction se crée, supplantant ceux qu'il ordonne. Une « machine de vision » (Virilio, 1988) est produite et constitue le lieu même où le procès de communication s'opère. C'est ce qui fait dire à Olivier Bomsel que « schématiquement, la prolifération des contenus indexés sur Internet engendre en quelque sorte un nouveau codage, une nouvelle chaîne de sens, dont le moteur s'institue décodeur universel » (Bomsel, 2007, p. 60).

En plus d'être une machine à faire voir, le moteur est aussi une machine à faire dire, car il produit un énoncé, ou méta-énoncé, à propos de l'actualité. En effet, Google fait parler l'espace-web, comme le topographe fait parler le territoire *via* la carte qu'il en dresse. Les concepteurs créent et paramètrent des indicateurs qui s'expriment conjointement pour élaborer un même énoncé. Ils font parler ces indicateurs, qui sont le fait des actions cumulées des éditeurs et des internautes, et non des ingénieurs directement. En revanche, c'est le concepteur qui décide de « donner la parole » à ces indicateurs (critères algorithmiques) ainsi que de calibrer la portée qu'aura la « voix » de chacun d'eux (pondération des critères).

1.4.3 Analyser où se situe le pouvoir

Le pouvoir, selon Foucault, est « un faisceau plus ou moins organisé, plus ou moins pyramidalisé, plus ou moins coordonné, de relations. » (Foucault, 1994). Les rapports de forces, ajoute Deleuze, sont « microphysiques, stratégiques, multiponctuels, diffus » (Deleuze, 1986, p. 44). Il nous faudra par conséquent rendre méticuleusement compte des véhicules du pouvoir (les forces) et des zones d'exercice (ce qui combine ou oppose ces forces). Ainsi, en étudiant par exemple certains critères de l'algorithme, nous chercherons de quelle manière Google peut faire « fonctionner des relations de pouvoir dans une fonction, et une fonction par ces relations de pouvoir » (Foucault, 1975, p. 208). Nous chercherons également à tracer

les relations de pouvoir dans les salles de rédaction. Et nous verrons comment Google, en assortissant ses services d'un ensemble de prescriptions, de contraintes et d'incitations, met en place une stratégie visant à inviter les éditeurs à optimiser leurs contenus de manière à ce que le moteur fonctionne au plus près de ce qui a été prévu par ses concepteurs. Là encore, nous retrouverons un élément central du dispositif, qui « a une visée d'efficacité, d'optimisation des conditions de réalisation, il est soudé au concept de stratégie. [...] À ce titre, le dispositif peut être défini comme la concrétisation d'une intention au travers de la mise en place d'environnements aménagés » (Peeters et Charlier, 1999, p. 18).

Pour déployer analytiquement le dispositif, il nous fallait être en mesure d'établir « la rationalité à l'œuvre dans la coordination des relations et des médiations qui s'établissent entre les différents éléments (institutions, acteurs, discours, lois, énoncés) et la manière dont elle contraint les acteurs selon leur situation » (Monnoyer-Smith, 2013). Or, pour cela, nous nous sommes aperçus qu'il nous faudrait différencier certains types de pouvoir, qui ne s'exercent pas de la même façon selon qu'un acteur cherche à « inciter, induire, détourner, rendre facile ou difficile, élargir ou limiter, rendre plus ou moins probable » (Deleuze 1986, p. 77). C'est pour effectuer une telle différenciation que nous avons mis au point une grille de lecture en nous appuyant sur les travaux de celui qui est sans doute le principal théoricien anglo-saxon de l'ANT : John Law. Selon ce dernier, le pouvoir doit être traité par le chercheur comme un « potentiel ou un ensemble de conditions tant qu'on n'oublie pas qu'il est aussi le produit d'un ensemble de relations structurées et plus ou moins précaires » (Law, 1991, p. 170). A partir de ce constat, et des travaux de Foucault, John Law définit deux formes de pouvoir : le « pouvoir de » (*power to*) et le « pouvoir sur » (*power over*). Le « pouvoir de » s'apparente à l'influence qu'un acteur peut avoir sur un autre acteur ou une catégorie d'acteurs. Il conduit leur conduite, sans que cette conduite soit à proprement parler contraignante, dès lors qu'il n'y a ni menace ni sanction possible. A l'inverse, le « pouvoir sur » sanctionne et menace, de façon à ce que d'autres acteurs se plient à la volonté de celui qui l'exerce.

Le pouvoir, quelle que soit sa forme, peut selon John Law être stocké et déployé, dès lors qu'il permet à un acteur de maintenir sa position d'une part, au moins pour quelques temps, et, d'autre part, de choisir entre différentes lignes d'actions possibles. Cette notion de choix est un des composants fondamentaux à la fois du « pouvoir sur » et du « pouvoir de » : en fait, elle désigne un « pouvoir de ne pas ».

Pour les besoins de notre enquête, nous avons conservé le concept de « pouvoir sur » et nous avons subdivisé la notion de « pouvoir de », car elle nous paraissait englober trop de cas différents et, par conséquent, ne pas permettre une analyse suffisamment fine. Nos subdivisions sont les suivantes :

- Le « pouvoir faire » désigne ce que Google et les éditeurs peuvent faire étant donné les caractéristiques du web. Ce terme englobe l'ensemble des possibles en termes techniques et éditoriaux pour celui qui crée des contenus ou des services et les publie sur le web, indépendamment des recommandations et des intérêts de Google ou des éditeurs en particulier.
- Le « pouvoir de faire » désigne le « pouvoir faire » qu'un éditeur doit à Google en particulier. La firme met en effet en œuvre certaines techniques, certains dispositifs, certains modes d'emploi et certaines recommandations dont les éditeurs peuvent se saisir, qui font alors office de « prescriptions » adressées à l'éditeur « en vue de rendre son environnement conforme à celui qui est prévu pour le dispositif » (Akrich, 2006). Il s'agit bien de pouvoir dans la mesure où les éditeurs ont la capacité de se saisir ou non des possibilités et des préconisations que Google met ainsi à leur disposition.
- Le « pouvoir de faire faire » désigne la capacité de Google à conduire la conduite des éditeurs pour qu'ils aillent dans le sens de ses intérêts, et cela librement, en défendant ce qu'ils estiment être les leurs.

Enfin, nous reprenons à John Law la notion de « pouvoir sur » :

- Le « pouvoir sur » désigne la capacité de Google à contraindre, éventuellement en les menaçant, les éditeurs à obéir à ce que ses porte-parole préconisent, sous peine de sanction. Si les éditeurs obéissent, ils reçoivent éventuellement une « grâce » au sein du dispositif. Le « pouvoir sur » fonctionne ainsi sur deux leviers : crainte de la sanction, désir de la grâce.

Nous délimiterons donc pour les éditeurs ce qui est de l'ordre de la possibilité d'action (pouvoir faire) et d'une possibilité d'action accordée (pouvoir de faire) et, d'autre part, pour Google, ce qui est de l'ordre de la gouvernementalité (pouvoir de faire faire) et de la contrainte (pouvoir sur). Bien sûr, une telle grille comporte les défauts de toute clef typologique : elle peut être remise en cause, notamment dans la mesure où certains champs d'action relèvent de plusieurs catégories. C'est pourquoi nous l'utiliserons comme un outil visant à clarifier notre exposé, et ne comptons pas en faire une proposition théorique. Nous l'avons d'ailleurs forgé au moment d'écrire notre thèse, et non comme un outil conçu *a priori* pour éclairer notre terrain.

1.4.4 Lignes de fuite

La notion de ligne de fuite nous permettra de décrire certaines situations où le référenceur¹⁹ ou les journalistes feront appel à leur créativité pour s'affranchir des pressions liées au référencement sur les services de Google, ou bien pour optimiser leur référencement d'une manière qui n'aura pas été prévue ou n'aura pas été souhaitée par les concepteurs du moteur. Ainsi, pour contourner le « pouvoir sur » et le « pouvoir de faire faire » de Google, certains éditeurs élaborent un savant cocktail entre leur « pouvoir faire » et leur « pouvoir de faire ». Ils développent des compétences et créent un « art opératoire » (Flichy, 2003) pour affronter ou contourner le programme d'action du concepteur. La créativité et la réactivité sont ainsi sources de subjectivation, c'est-à-dire que les éditeurs opposent une résistance à la machine à faire dire et à faire voir comme elle a été pensée. C'est en ce sens qu'il

¹⁹ Nous désignons par le terme « référenceur » la personne embauchée par une entreprise de presse dans le but de maximiser le trafic apporté par Google Search et Google Actualités.

est possible de dire que « tout réagencement est susceptible de produire de nouvelles relations de pouvoir, de nouvelles pratiques, et de nouvelles productions symboliques, bref : de nouvelles médiations » (Bonaccorsi et Julliard, 2010, p. 8). Une fois que la machine à faire dire et à faire voir aura été contestée ou contournée, sa matérialité sera éventuellement modifiée, déplacée, transformée : *traduite*. C'est pourquoi il nous faudra analyser comment les éditeurs peuvent « regimber » en créant certaines zones de contestation et certaines zones d'affranchissement, puis en s'en saisissant comme des *lignes de fuite*.

Cependant, dès lors que leur pouvoir est contesté ou contourné, les concepteurs du dispositif peuvent tout à fait réviser leur programme d'action à leur tour dans le but de maintenir intacte leur influence. Les ingénieurs de Google vont par exemple redéfinir les « pouvoirs de faire » accordés aux éditeurs, pour reprendre la main sur l'échelle de visibilité. Ils pourront également modifier certaines composantes techniques et sémiotiques du processus (changement d'algorithme, d'interface...), auquel cas le dispositif deviendra un « autre dispositif ». La ligne de fuite, résultant de la créativité des éditeurs, pourra alors éventuellement être « rattrapée » par le concepteur, qui aura fait preuve de créativité à son tour dans le but de maintenir son pouvoir quant à la nature et au résultat du processus. Il nous faudra donc analyser comment Google se déplace, s'adapte, fait front et contourne les cours d'action des dirigeants des entreprises de presse, des référenceurs et des journalistes. Nous voyons ici très bien comment et pourquoi le concept de dispositif, articulé avec l'ANT et opérationnalisé en SIC, permet de décrypter ce que nous avons présenté comme un « match de ping-pong » performatif entre l'énoncé, son objet, ceux qui ont pour but de concevoir cet énoncé et ceux qui évoluent au contact de son objet.

1.5 Problématique et hypothèses

Problématique. *La mise en énoncé effectuée par le dispositif d'infomédiation Google performe-t-elle le processus de production et de valorisation du contenu mis en ligne par les entreprises de presse ?*

« Performer, expliquent Fabian Muniesa et Michel Callon, c'est provoquer, instaurer, constituer, faire que quelque chose a lieu » (Muniesa et Callon, 2009, p. 301). La performance, ajoutent-ils, est « un ensemble d'activités et d'événements qui instaurent ou modifient un agencement » (*ibid.*, p. 302). Ainsi, lorsque nous utilisons ce verbe dans notre problématique, nous sous-entendons qu'il nous faut à la fois étudier ce qui provoque, ce qui est provoqué, la situation que cela instaure, ce qui est modifié, ce qui s'agence, ce qui a lieu et le lieu même de cet « avoir lieu ». En outre, le concept de performativité nous permet de ne pas considérer la seule influence exercée directement par la firme Google sur l'action des éditeurs, mais l'ensemble des effets que peut avoir le fait de produire un méta-énoncé sur l'activité des éditeurs, leurs contenus, les circuits de valorisation et les représentations des webjournalistes. Cela nous conduira bien sûr à étudier comment la firme Google exerce son influence directement ou indirectement, mais également comment d'autres acteurs peuvent eux aussi influencer l'action des éditeurs et son résultat par la *médiation* du dispositif d'infomédiation.

Hypothèse générale. *Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent radicalement la manière dont le fait d'actualité est mis en récit, mis en circulation et valorisé par les entreprises de presse sur le média web.*

En parlant dans notre hypothèse générale à la fois de « conditions » et de « modalités », nous faisons référence à ce qui peut contraindre les acteurs en présence en *conditionnant* leur comportement, autant qu'à la marge de manœuvres grâce à laquelle ils sont à même de *moduler* celui-ci.

Nous avons ensuite décliné notre hypothèse générale en cinq hypothèses qui nous permettraient de répondre à notre problématique. Nous avons veillé dans leur formulation à ce que les principales dimensions du dispositif puissent être prises en compte. En les testant, nous pourrions finalement évaluer ce qu'il en est de notre hypothèse générale et répondre au mieux à notre problématique.

Hypothèses

1/ Le dispositif d'infomédiation Google contribue à créer de nouvelles formes de valorisation de la production journalistique.

2/ Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google entraînent une reconfiguration des organisations de presse.

3/ Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent les choix quant à la structure du site, des pages et du code HTML.

4/ Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent les webjournalistes dans leurs pratiques de production et de publication.

5/ Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent les lignes éditoriales des sites de presse en ligne.

1.6 Terrains

Nous pensons, à l'instar de Cyril Lemieux, qu'« une bonne description, en sciences sociales, est littérale, non pas interprétative ou explicative. Autrement dit, [...], elle doit être la plus « mince » possible. » (Lemieux, 2009, p. 97) Pour effectuer cette description, il nous a fallu d'abord choisir ce que nous allions observer et décrire : quels acteurs, quels actants, quelles dynamiques. L'objet de cette sous-section est d'expliquer cette démarche et de présenter les terrains et les méthodes de notre investigation.

1.6.1 Méthodologie de collecte

L'une des premières inquiétudes du chercheur qui a choisi l'ANT est l'impossibilité apparente de dénombrer l'ensemble des actants participant du réseau qu'il souhaite étudier. En effet, nous devrions potentiellement nous intéresser aussi bien à l'ingénieur de Google qu'à ses dirigeants, ses actionnaires, ses sous-traitants, celui qui installe dans les foyers les lignes permettant de se connecter à Internet, les lignes elles-mêmes, toutes les infrastructures, logiciels, interfaces, algorithmes, code, puis les journalistes, les rédacteurs en chef, les personnels chargés de la publicité et de l'acquisition d'audience, les formateurs, les concurrents amateurs et professionnels, les usagers, les députés, les sénateurs, les juges, les économistes, les associations, les claviers, les souris... La liste pourrait être ainsi continuée indéfiniment. Mais l'ANT dit aussi qu'il n'y a pas de différence insurmontable entre les niveaux micro et macro, et qu'il existe certains centres de traduction qui, à un moment donné, sont à même de mobiliser l'ensemble du réseau afin d'engager certaines actions (Callon, 2006, p. 270). En d'autres termes, il existe des porte-parole par lesquels plusieurs (ou tous) les actants doivent passer pour agir et qui, dès lors, renseignent le chercheur au sujet de l'ensemble du réseau auquel ils prennent part. Nous étudierons ces points nodaux que sont, chez Google, les concepteurs et les dirigeants et, en particulier parmi eux, ceux qui s'expriment en public au nom de la firme. Afin d'étudier le script de Google, en plus des multiples interventions orales et écrites des dirigeants et des ingénieurs, il existe les « centres d'aide ». Il s'agit de sites où sont publiées un ensemble de préconisations visant à optimiser la

construction d'un site afin que ses documents soient le mieux référencés possible. Nous nous intéresserons en particulier au Centre d'Aide pour les Webmasters (CAW), et au Centre d'Aide pour les éditeurs souhaitant être référencés sur Google Actualités (CAA)²⁰. Nous pourrions ainsi savoir dans le détail ce que Google prescrit et prévoit quant au comportement des éditeurs. Ce sera un moyen de voir quel serait l'impact idéal que la firme Google souhaiterait avoir sur les caractéristiques des contenus produits et mis en ligne. De la même manière qu'avec les actants humains, il faut, pour les acteurs non-humains innombrables liés au cours de l'action, saisir les points nodaux. Parmi eux, on trouvera dans le cas de Google les logiciels de *crawling*, les critères de l'algorithme d'appariement et l'interface. Il nous faudra également considérer certains actants qui ne font pas directement partie du moteur mais y sont liés et peuvent également infléchir l'action des participants. Ce sera le cas par exemple de l'outil « Tendances de recherche ». Ce dernier a pour vocation d'agrèger les termes des requêtes effectuées sur Google Search et participe, dès lors, lui aussi à la « mise en récit » effectuée par le dispositif.

Dans les entreprises de presse, nous nous intéresserons aux dirigeants, journalistes et au référenceur qui, eux-aussi, peuvent mobiliser tout ou partie du réseau pour engager certaines actions stratégiques. Pour ce qui est des actants non-humains, nous considérerons le système de management de contenus (CMS) et les outils censés influencer le référencement des documents, notamment les balises HTML, les microdonnées, les « *Sitemaps* » et les liens hypertextes. Nous verrons comment les journalistes sont formés à l'optimisation des contenus pour les moteurs de recherche, quelles préconisations leur sont faites et quels objectifs leur sont fixés. Nous pourrions, là encore, avoir une idée du script initié par les dirigeants et par le référenceur dans l'espoir de maximiser l'audience. Et nous considérerons le script des journalistes eux-mêmes, dans le but de comprendre quel est leur savoir en termes de pratiques SEO et dans quelle mesure, et pourquoi, ils mobilisent ce savoir. En d'autres termes, nous essaierons de cerner la vision qu'ont les journalistes du rôle et de l'importance du moteur vis-à-vis de leur propre rôle et de la qualité de ce qu'ils produisent, et de comprendre de quelle manière ils agissent étant donné cette vision.

²⁰ Toutes les citations que nous ferons du CAW et du CAA ont été récoltées sur ces sites entre novembre 2011 et décembre 2012.

En plus de décrire les actants et leurs scripts, nous observerons un certain nombre de résultats de leur rencontre. Le fait que nous ayons accès aux services de Google autant que nous le souhaitons, ainsi qu'aux sites de presse en ligne, nous permettra de comparer diverses situations. Nous décrirons les interfaces de Google Search et Google Actualités, la disposition des contenus sur les sites de presse en ligne, les adresses URL, les titres et les chapeaux, les articles eux-mêmes. Nous observerons également le code HTML, cela pour voir quel usage y est fait des balises.

1.6.2 Approche qualitative

Nous avons choisi d'effectuer une enquête essentiellement qualitative. Ce type d'enquête « se prête à l'étude des phénomènes complexes et mouvants » (Deslauriers, 1991, p. 14). A partir de nos hypothèses, nous adoptons une logique à la fois descriptive et compréhensive nous permettant de décrire le phénomène d'une part et, d'autre part, de l'appréhender tel qu'il est vécu par les sujets. Par ailleurs, notons bien ici que l'approche qualitative ne recherche pas la représentativité mais la diversité des mécanismes : « La démarche [de la méthode qualitative] [...] n'analyse pas les corrélations statistiques, mais les mécanismes sous-jacents aux comportements et l'interprétation que les acteurs font de leurs propres comportements ; elle ne recherche pas la représentativité mais la diversité des mécanismes » (Alami *et al.*, 2009, p. 4).

Une grande partie de notre collecte s'est déroulée sur le terrain : nous sommes allés à la rencontre des acteurs que nous avons observé et avec qui nous nous sommes entretenus. Le but que nous poursuivions était, chaque fois que nous rencontrions un acteur, « de savoir ce que la personne pensait et d'apprendre des choses qu'on ne peut observer directement comme les sentiments, les idées, les intentions » (Deslauriers, 1991, p. 34). Ainsi, nous avons identifié « les mécanismes stratégiques des relations de pouvoir ou de coopération entre acteurs » (Alami *et al.*, *op. cit.*, p. 15).

1.6.3 De la difficulté d'étudier Google

Il n'est pas aisé d'étudier Google (Langville et Meyer, 2006, p. 346 ; Jarvis, 2009, p. 97 ; Vaidhyathan, 2011, p. 9, p. 52)²¹. Nous avons envoyé des dizaines de courriels à des dizaines d'employés de la firme Google et, le plus souvent, la réponse fut la même : dès le premier ou le second courriel, les personnes contactées acceptaient de nous rencontrer sans nous donner ni lieu ni date de rendez-vous. Nous leur écrivions alors pour les remercier et leur signifier que nous nous tenions à leur disposition, quand ils le voudraient, où ils le voudraient, mais ne recevions aucune réponse. Nous avons beau lancer et relancer : aucune réponse, jamais. Nous ignorons s'il s'agit d'une consigne donnée aux employés, mais une chose est certaine : la méthode qui consiste à dire oui sans jamais préciser où et quand fait perdre son temps au chercheur, qui s'obstine, ce qu'il n'aurait pas fait dans le cas d'un refus explicite. Par ailleurs, nous ne pourrions jamais reprocher aux employés de Google de ne pas s'être prêtés au jeu de la recherche (dont leur entreprise est issue !) puisqu'ils nous ont effectivement dit « oui » avant de ne plus rien nous dire du tout.

Malgré tout, et après certains stratagèmes, nous avons réussi à interroger une personne par téléphone, à rencontrer physiquement la responsable en France des relations institutionnelles, et à interroger par courriel le responsable en Europe des relations entre Google et les éditeurs de presse.

Tableau 1. Employés de Google interrogés

Google	Michel Benard	University Relations Manager	22/12/2011 30 minutes (téléphone)
	Elizabeth Bargès	Policy Manager	24/01/2012 1h (brasserie – 75009)
	Madhav Chinnappa	Head of Strategic Partnerships, Google News EMEA	21/02/2012 (Courriel)

²¹ Nous n'en citons que trois ici, mais de nombreux chercheurs se sont plaints de la difficulté d'étudier la firme. de Mountain View. A titre d'exemple, voici une des citations dont nous nous servons ici : « But I must acknowledge the irony of advocating transparency in a book about Google, which in many ways is as opaque and secretive as Dick Cheney. You can't get into a Google office without signing a nondis-closure agreement. Google won't reveal details of its revenue split with sites that run its ads. It refuses to list its Google News sources. It won't tell us how many servers it has. It chooses not to use open- source software for some functions, like managing its cloud of computers, so it can retain a proprietary advantage » (Jarvis, 2009, p.97).

Ces entretiens ne constituent pas une grande richesse, hélas, même si quelques points sont tout de même intéressants. La plupart des questions centrales ont été éludées par un « je ne sais pas », « ce n'est pas mon domaine de compétence », ou « adressez-vous à X » à qui nous nous sommes adressés et qui nous a dit « oui, rencontrons-nous » mais qui, comme les autres, n'a pas donné suite à son accord. Il nous a donc fallu aborder la question autrement et procéder à notre enquête là où nous pouvions obtenir des informations, c'est-à-dire :

- 1) Les livres biographiques, nombreux, parfois hagiographiques, à propos de Google où nous avons trouvé un certain nombre d'informations à propos de l'entreprise, de son organisation, mais aussi de nombreux verbatim des employés que leurs auteurs ont réussi à interroger (réussissant où nous avons échoué). Parmi ces livres, l'un d'eux est écrit par un ancien employé lui-même (Edwards, 2011).
- 2) Les sites de Google, centres d'aide, blogs, où sont données les prescriptions aux usagers et aux éditeurs, et où les employés s'expriment à propos de leur entreprise, de ses objectifs, du fonctionnement du moteur, du cadre légal, et répondent à certains des reproches qui leur sont adressés. Nous avons également utilisé les communiqués de presse publiés par Google.
- 3) Les documents financiers, publiés chaque trimestre puisque Google est une entreprise cotée en bourse (Nasdaq) et doit par conséquent répondre à un certain nombre d'obligations de présentation et d'explication des chiffres, tout en présentant aux investisseurs quelles sont ses prévisions et ses stratégies pour l'avenir.
- 4) Les sites spécialisés, et notamment le site *Search Engine Land*, où un travail de veille est fait quotidiennement par des experts de haut niveau. Ceux-là ont souvent accès aux employés de Google pour des interviews. En France, il existe également des sites comme *WebRankInfo*, *Café référencement* et *Abondance*, qui nous ont été très utiles.
- 5) Les compte-rendu des affaires judiciaires pendant lesquelles certains représentants de Google sont forcés de s'exprimer et de justifier leurs

décisions. Nous avons épluché ces compte-rendu, les analyses des experts et les décisions des tribunaux.

- 6) Les publications scientifiques à propos de Google, dont la plus célèbre est celle des fondateurs eux-mêmes (Brin et Page, 1998). Dans d'autres publications, très rares, les chercheurs ont réussi à interroger les employés à propos des sujets qui nous intéressent. En France notamment, Nikos Smyrnaioi a réussi à interroger l'ancien Pdg de Google France (Smyrnaioi, 2005). Même si l'entretien date un peu, il nous a également aidé.
- 7) Les concurrents : Nous avons pu nous entretenir avec le Directeur des Affaires Publiques et Juridiques de Microsoft, Marc Mossé, à propos de la concurrence entre le moteur Bing et le moteur Google.
- 8) L'observation en ligne qui nous a permis de voir quel était le résultat apparent des choix et des stratégies mis en œuvre par Google.
- 9) Les brevets : Google a déposé certains brevets nous permettant d'en savoir d'avantage sur les techniques mises au point.
- 10) Les interventions par voie de presse : nous avons relevé de nombreux verbatim extraits d'articles et d'interviews publiés dans la presse au sujet de Google, et pour lesquels les journalistes ont eu accès à des porte-parole de la firme.

Nous n'avons pas eu accès à ce qu'il se passait à l'intérieur de la firme Google : quelles sont les négociations entre les ingénieurs ? entre les ingénieurs et les dirigeants ? Aussi avons-nous considéré la firme en aval de ces négociations. Quand nous disons « Google » ou « la firme », c'est à une entité stabilisée que nous faisons référence, tandis que lorsque nous disons « Google Search », « Google Actualités », « le moteur », nous désignons le dispositif d'infomédiation au sens large et donc un ensemble d'acteurs et d'actions dépassant largement le seul cadre de la firme.

1.6.4 Editeurs et référenceurs

Chez les éditeurs, notre terrain s'articule sur trois niveaux. Le premier est d'ordre macro : nous considérons « l'éditeur » comme un ensemble stabilisé, désignant tous les éditeurs de presse en ligne à la fois, dès lors qu'ils embauchent des

journalistes, traitent d'information d'actualité politique et générale et sont essentiellement financés sur un modèle de gratuité pour l'internaute et de facturation d'encarts publicitaires à des annonceurs, de manière à ce que les enjeux liés à la maximisation du trafic soient forts. Cela nous permettra d'aborder les éditeurs d'un point de vue théorique, afin de voir comment leur modèle économique est lié à celui de Google et quelle est la marge de manœuvres dont ils disposent pour optimiser le référencement de leurs documents.

Le deuxième niveau de notre terrain est mezzo. Ainsi, une fois que nous aurons identifié grâce au niveau macro tout ce qu'il est possible de faire pour un éditeur de presse en ligne, quel qu'il soit, au sein d'un dispositif d'infomédiation dont nous aurons déployé analytiquement les tenants actionnels et intentionnels, nous pourrons aller à leur rencontre et suivre leurs déplacements. En rencontrant leurs employés et leurs dirigeants, nous avons pu entrer dans les salles de rédaction, relever certains points communs et certaines différences, ainsi que certaines divergences, certaines frictions, certaines négociations, certaines inquiétudes et certaines issues mises en place ou envisagées. Sur ce terrain mezzo, nous avons choisi de concentrer notre terrain sur les seuls Sites Issus de la Presse Imprimée (SIPI ci-après). Lors de nos entretiens exploratoires, nous avons interrogé certains porte-parole de sites nés en ligne, comme *Rue 89*, où nous avons rencontré les fondateurs Pascal Riché et Laurent Mauriac, et *Mediapart*, où nous avons interrogé le journaliste Vincent Truffly. Cependant, il ne nous a pas paru justifié d'intégrer ces sites au champ de notre enquête, cela car *Mediapart* fonctionne sur un modèle payant tandis que la part d'audience de *Rue 89* provenant de Google est de 15% (selon Pascal Riché), ce qui est très peu relativement aux autres sites de notre échantillon et a constitué à nos yeux un motif d'exclusion. S'il n'y a pas non plus de sites issus de la radio ou de la télévision, c'est parce qu'après avoir interrogé une journaliste de *France24*, nous avons réalisé que la grande différence dans le cas de la télévision et de la radio est que la stratégie de référencement n'a pas, ou pas encore, d'influence sur la manière de filmer (cadrer, zoomer, déplacer, étalonner) ou d'enregistrer du son, tandis qu'elle peut influencer directement la manière d'écrire (choix des mots, disposition des phrases). Nous ne prétendons pas qu'il ne serait pas intéressant d'étudier les

stratégies de référencement des sites issus de la télévision et de la radio, mais que cela constituerait un autre champ de manœuvres HTML, un autre ensemble de stratégies possibles, et donc un autre champ d'investigation. Par ailleurs, il serait également intéressant de considérer d'autres sites nés en ligne traitant de l'information politique et générale en continu, comme le *Huffington Post*, mais nous n'avons hélas pas eu le temps dans le cadre de cette enquête, préférant nous concentrer sur des sites comparables.

L'avantage de notre positionnement du côté des SIPI est qu'il nous a permis d'étudier les stratégies d'entreprises préexistantes au web et à Google où, dès lors, il pouvait exister une certaine mise en tension entre plusieurs visions du rôle social joué par le journaliste, de la ligne éditoriale du titre et des traitements adéquats de l'actualité. Cette mise en tension et les frictions qu'elle occasionne n'auraient sûrement pas, ou en tout cas beaucoup moins, pu être observées dans des rédactions créés dans un environnement dont Google Search était déjà un des points nodaux. Le fait d'avoir centré notre étude sur les sites issus de la presse imprimée nous a donc permis : 1) de comparer ce qui était comparable, 2) de rendre compte de mises en tensions liées à la traduction de pratiques journalistiques et organisationnelles préexistantes à Google (i.e. de nous intéresser au *changement*, aux *mutations* et à la *reconfiguration* de l'existant) 3) d'étudier des entreprises qui ont des marques fortes et où, par conséquent, l'enjeu lié au référencement est problématique dans la mesure où une augmentation d'audience ne signifie pas forcément une augmentation durable des recettes publicitaires, dès lors que la marque peut être dévaluée si la qualité perçue par les lecteurs et les publicitaires diminue.

Tableau 2. Employés des SIPI interrogés

PQN	lefigaro.fr	Luc de Barochez	Directeur de la rédaction	19/10/2011 1h (Le Figaro)
		Laurent Suply	Journaliste – Chef de Projet éditorial	27/10/2011 1h (Le Figaro)
		Tristan Vey	Journaliste	26/10/2011 1h (Le Figaro)

		Michaël Vuillaume	Chef de projet SEO	19/10/2011 1h (Le Figaro)
	Leparisien.fr	Guillaume Giraudet	Responsable SEO	30/05/2012 1h30 (Bistro 75015)
Presse magazine	lepoint.fr	Marc Vignaud	Journaliste	05/10/2011 1h (Café – 75019)
		Laure Aubry	Responsable SEO	02/05/2012 2h (Café – 75014)
	Lexpress.fr	Voir ci-après		
	Lexpansion.com	Voir ci-après		
	Lentreprise.fr	Voir ci-après		
PQG	20minutes.fr	Pierre-Jean Bozo	Directeur de Publication	20/02/2012 1h (20 Minutes)
PQR	Ladepeche.fr	Michaël Bourguignon	Directeur Général	21/09/2011 (Email – document Word)
		Jean-Jacques Rouch	Journaliste	21/07/2011 1h (Chez lui)
		Philippe Rioux	Journaliste – Chargé des relations entre la rédaction papier et le web	20/04/2011 1h (La Dépêche)
	Ouest-france.fr	Antoine de Tarlé	Directeur Général adjoint –en charge du multimédia	10/03/2011 1h (Ouest France)

Nous avons également rencontré certains spécialistes de la question du référencement des sites de presse sur Google, notamment des formateurs et des consultants. Ces acteurs sont mandatés par les dirigeants pour former les journalistes, auditer les sites, formuler des recommandations, mettre en œuvre des changements, effectuer la veille. Ils participent donc entièrement « au » et « du » réseau étudié.

Tableau 3. Consultants interrogés

Resoneo	Julien Crenn	Consultant SEO	28/03/2012 2h (Café- 75009)
Resoneo	Jean-Christophe Baudrier	Consultant SEO	10/04/2012 2h (Café- Levallois-Perret)
Awe Search Performance	Olivier Kerdanet	Consultant	27/10/2011 3h (Café- 75003)
Wan-Ifra	Cédric Motte	Formateur	21/06/2011 2h (Café- 75009)

Comme pour Google, nous avons également complété ces entretiens avec des verbatim issus de livres ou d'interviews trouvées dans la presse, de documents financiers, de comptes-rendus judiciaires et de publications scientifiques. Comme pour Google également, nous avons pu observer les sites, leur hiérarchisation, leur organisation, la disposition des contenus, et les contenus eux-mêmes : choix des mots, liens hypertextes, sujets traités, angles, nombre de mots, tags, adresses URL, métadonnées. En particulier, nous avons sélectionné six sites d'information politique et générale adossés à des titres de presse parmi les plus visités en 2012 selon l'OJD (Lemonde.fr, Lefigaro.fr, Lepoint.fr, Lexpress.fr, Leparisien.fr, Nouvelobs.com). Pour chacun de ces sites, nous avons prélevé les trois articles qui se trouvaient en tête de la page d'accueil au moment où nous les visitons, durant quatorze jours en octobre 2012, pour un total de 252 articles. Nous avons ensuite codé à la main un certain nombre d'informations afin de comparer l'utilisation de certains signes de ponctuation, les liens entrants et sortants, l'usage des balises HTML et les modules publicitaires. En outre, pour chacun des six sites, nous avons observé leurs ressource « robots.txt », leurs Sitemaps, leur rubriquage et leur disposition. Même si le nombre d'article est relativement faible, cette observation a eu le mérite de mettre très clairement à jour certains points communs et certaines particularités chez les six sites observés.

1.6.5 Groupe Express Roularta

Le seul niveau mezzo nous a paru insuffisant. Nous avons considéré qu'il était nécessaire d'aller vers le particulier, jusqu'au niveau micro. Ainsi, pour une entreprise de presse donnée, nous avons rencontré des acteurs dont les métiers et les profils étaient très différents. Cette approche monographique nous a permis de préciser et d'exemplifier de nombreux points qui, sans elle, seraient restés au stade de la supposition ou de l'intuition. Si nous avons choisi le Groupe Express Roularta (GER), c'est en premier lieu parce qu'une réflexion approfondie y a été menée à propos des stratégies de référencement, en second lieu parce que *L'express.fr* est un site d'information d'actualité politique et générale parmi les 7 plus importants en France en termes d'audience (OJD), en troisième lieu parce que sa marque est forte aussi bien pour la version papier que pour la version en ligne, et, enfin, parce que les dirigeants et les employés du GER ont très aimablement accepté de collaborer à notre projet.

Tableau 4. Employés de RMF interrogés

L'express.fr	Eric Mettout	Rédacteur en Chef	29/09/2011 1h (<i>L'Express</i>)
L'express.fr	Thomas Bronnec	Directeur adjoint délégué en charge de la home page.	23/11/2011 1h (<i>L'Express</i>)
L'Express (Papier)	Marc Epstein	Rédacteur en Chef du service Monde	25/01/2012 1h (<i>L'Express</i>)
L'Express (Papier)	Julien Bordier	Rédacteur en chef adjoint au service Arts & Spectacles	15/05/2012 1h (Café- 75009)
L'express.fr	Camille Caldini	Responsable d'édition web— FrontPage Editor	02/11/2011 1h (<i>L'Express</i>)
L'express.fr	Catherine Goueset	Journaliste - Monde	26/12/2011 1h (Café- 75009)
L'express.fr	Clément Daniez	Journaliste – Sport	26/01/2012 1h (<i>L'Express</i>)
L'express.fr	Julie Saulnier	Journaliste - Société	22/11/2011 1h30 (Restaurant- 75009)
L'express.fr	Matthieu Deprieck	Journaliste - Politique	02/02/2012 1h (<i>L'Express</i>)

Lexpress.fr	Flavien Hamon	Journaliste animateur de communauté	16/12/2011 45min (<i>L'Express</i>)
Express Roularta Services	Céline Deguy	Responsable marketing	28/10/2011 45min (<i>L'Express</i>)
Express Roularta Services	Laurent François	Responsable marketing	28/10/2011 45min (<i>L'Express</i>)
Lexpress.fr	Emilie Mounier	Chef de projet Marketing - web	27/10/2011 1h30 (Café- 75009)
Groupe Express Roularta	Marc Bouguié	Directeur de l'Acquisition d'audience	07/12/2011 2h (<i>L'Express</i>)
Groupe Express Roularta	Renaud Perrin	Responsable du référencement	28/09/2011 2h (Café – 75009)
Groupe Express Roularta	Sophie Gohier	Directeur adjoint Développement Nouveaux Médias	22/11/2011 1h (<i>L'Express</i>)
lexpansion.com ²²	Yves Adaken	Rédacteur en chef	02/11/2011 1h (<i>L'Expansion</i>)
lexpansion.com	Julie de la Brosse	Journaliste	27/09/2011 1h (Restaurant-75009)
lexpansion.com	Emilie Lévêque	Journaliste	20/10/2011 1h (Café- 75009)
Lentreprise.fr	Emmanuel Colombié	Journaliste	29/09/2011 1h30 (Café-75009)
	Sébastien Pommier	Journaliste	20/10/2011 1h30 (Café-75009)

*

Comme nous l'avons brièvement annoncé en introduction, nous allons maintenant historiciser la situation de communication dont nous prétendons rendre compte (chapitre 2 et 3). Après quoi nous déploierons le dispositif, les marges de manœuvres, les incitations, les conditions et les modalités qui prévalent à la production de l'énoncé (Partie 2). Enfin, nous ferons état des stratégies de captation des éditeurs, et de leurs effets sur le processus de production journalistique et sur la mise en récit de l'actualité (Partie 3).

²² L'Expansion.com est aujourd'hui apparenté à la rubrique « économie » de l'express.fr. Les journalistes sont assis à côté des journalistes de l'Express.fr, et sont partiellement intégrés aux équipes, et ce même s'ils ont leur propre titre papier : le mensuel l'Expansion, qui lui est clairement distinct de l'hebdomadaire l'Express. La technique des deux sites l'express.fr et l'expansion.com est la même. D'ailleurs, depuis 2011, l'expansion.com a changé d'adresse URL et s'appelle « l'expansion.lexpress.fr ». Nous verrons que ce changement a eu lieu en partie pour des histoires de référencement.

CHAPITRE 2. GENEALOGIE DE LA RECHERCHE D'INFORMATION

« Le nombre advient, celui de la démocratie, de la grande ville, de la cybernétique. C'est une foule souple et continue, tissée serrée comme une étoffe sans déchirure ni reprise, une multitude de héros quantifiés qui perdent nom et visage en devenant le langage mobile de calculs et de rationalités n'appartenant à personne. Fleuves chiffrés de la rue. »

(De Certeau, 1980, p. 12)

2.1 Musée imaginaire

2.1.1 Produire pour être stocké, stocker pour être trouvé

Chez les Hébreux, les textes où le nom de Dieu figurait étaient systématiquement stockés dans les *guenizot*²³ sans qu'on en juge la pertinence et sans le moindre effort de classification. Il était impossible de consulter ces textes, et quand bien même : il aurait été inconcevable de trouver un texte en particulier étant donné l'absence de système d'indexation et de classification. Ainsi, comme le rappelle Paul Duguid (Duguid, 2008), pour qu'une recherche ait une chance d'aboutir, les messages doivent avoir été stockés afin de pouvoir être trouvés. Mais il faut également que ce stockage soit possible, ce qui dépend en partie de leur producteur : celui-ci, au moment où il émet un message, peut faire en sorte que l'information puisse être conservée et stockée le mieux possible. Le producteur considère cela dès l'instant où il décide de produire son message : quel support, quelle encre, quelle couleur, quel alphabet, etc. De sa volonté ainsi objectivée dépend en partie la possibilité de transporter et de stocker le support. Si le producteur n'a pas désiré que son message fût mémorisé, ou bien s'il n'a pas eu les moyens de s'en assurer, ce sera au destinataire de s'en débrouiller. Il pourra par exemple recopier le message sur un support plus fiable, plus facile à transporter et à stocker, en y additionnant des méta-informations visant à pallier la négligence, volontaire ou non, du producteur.

²³ « Guenizot » est le pluriel de « guenizah », en hébreu « גניזה » : endroit où l'on stocke les textes dans une synagogue.

2.1.2 Répondre à un besoin d'information

L'Histoire commence lorsque la société des hommes s'organise autour de lois écrites, de contrats, d'écritures monétaires, de démonstrations scientifiques et de textes à caractère religieux et pédagogiques (Serres, 2007). Le *contrat social* n'est ainsi envisageable qu'à condition qu'une technique objective durablement les informations dont il est concomitant : ne pas les altérer ou les modifier, garder intact leur support, pouvoir les transporter, les stocker et être en mesure de les mobiliser dès que cela sera nécessaire. C'est principalement pour répondre à ces besoins que les hommes se sont évertués à innover en matière de supports, de transmission, de stockage et de mobilisation de l'information.

On pouvait transporter les tablettes en cire plus facilement que les tablettes en pierre mais elles s'altéraient avec la chaleur et le message était alors partiellement perdu ou modifié. Les tablettes en pierre, quant à elles, étaient trop lourdes et compliquées à manier. Les rouleaux de papyrus finirent par s'imposer, la plupart des écrivains de l'antiquité considérant leur format comme étant le plus adéquat (Langville et Meyer, 2006). Les rouleaux, ou *volumen*, permettaient de produire une information plus aisée à diffuser. Facile à transporter, elle pouvait être émise pour être reçue, y compris par des destinataires qui n'étaient proches ni dans l'espace ni dans le temps, et être stockée afin d'être trouvée. Dès lors, les éditeurs qui maîtrisaient les procédés de confection des *volumen* augmentaient leurs chances de voir leurs contenus intégrer durablement les bibliothèques et leur message circuler dans l'opinion publique.

2.1.3 A chaque support ses méta-informations

Les Grecs et les Romains inventèrent un système d'étiquetage : attachées aux rouleaux, des étiquettes contenaient un résumé permettant au lecteur potentiel d'éviter de dérouler un long document pour finalement s'apercevoir que les informations qu'il contenait ne l'intéressaient pas. Ainsi, le support était assorti de méta-informations, celles-ci ne faisant pas partie du contenu mais du contenant, qu'elles permettaient d'archiver de manière à ce que le message puisse être mobilisé. Elles ajoutaient au rouleau une fonction : celle de permettre au chercheur de savoir ce

qui s’y trouvait sans avoir à le dérouler, ce qui était contraignant et risquait d’abîmer le rouleau. L’éditeur pouvait ainsi chercher à simplifier la vie de ceux à qui il s’adressait en assortissant son support d’indications méta-informatives. Probablement, cela augmenterait ses chances d’être entendu. Les méta-informations pouvaient également avoir été produites par un intermédiaire, par exemple un bibliothécaire, qui aurait lui-même été en possession du rouleau et aurait voulu en simplifier l’accès aux visiteurs de sa bibliothèque.

Au V^{ème} siècle avant notre ère, Amy Langville et Carl Meyer (2006) racontent que les Grecs précédaient les pièces de théâtre par un résumé chanté par le chœur afin que ceux qui n’étaient pas intéressés puissent passer leur chemin avant le début de la pièce. Dans cet exemple, celui qui manquait le résumé ne manquait en rien la pièce. Les strophes chantées s’adressaient à l’éventuel spectateur afin de lui faire gagner du temps, quitte à commencer par lui en faire perdre. Ce dernier était alors en mesure de répondre à la question : « d’après ce que le contenant me dit du contenu, est-ce que oui ou non le contenu correspond à ce dont j’ai besoin ? » Ce résumé chanté, composé par l’auteur de la pièce ou par quelque autre personne, visait également à appâter le spectateur en faisant la publicité de la pièce, précurseur de ce que sont aujourd’hui les bande-annonce au cinéma — et de ce qu’est le lien hypertexte généré par un moteur de recherche.

2.1.4 A nouveau contenant, nouveaux procédés

Confrontés à une pénurie à la suite du non ravitaillement par les Egyptiens en papyrus, les Pergaméniens eurent l’idée d’inscrire leurs messages sur un nouveau type de support. Fait de peau de bête, le parchemin²⁴ leur permit de ne plus être dépendants du bon vouloir des Egyptiens en matière d’exportation de papyrus, tout en créant une technique que chaque nation pratiquant l’élevage pourrait utiliser dès lors qu’elle la maîtriserait. Par ailleurs, le parchemin ne pouvant être roulé sans risquer d’altérer le support et son contenu, les Pergaméniens décidèrent de le diviser en plusieurs rectangles qu’on empilerait dans un *codex* : l’ancêtre du livre. On construisit de nouvelles bibliothèques, dont les rayonnages furent adaptés à ces objets

²⁴ L’étymologie du mot « parchemin » est dérivée du mot « pergaménien ».

rectangulaires. On s'aperçut ainsi qu'il était plus aisé d'archiver des codex que des volumens et que le feuilletage permettait d'accéder plus rapidement à l'information que le déroulage (Kilgour, 1998). Très vite, on assortit le codex d'une tranche permettant de maintenir les feuilles ensemble, sur laquelle on inscrivit certaines méta-informations : en général, le titre de l'ouvrage, sa date de publication et le nom de son auteur. Une contrainte matérielle et économique, liée à un différend politique entre les Egyptiens et les Pergaméniens, avait ainsi abouti à l'invention d'une nouvelle technique de production, laquelle avait conduit à repenser l'archivage, ce dont on profita pour faciliter la recherche de contenus en créant une forme méta-informative mieux soudée au contenant. Nous voyons ici comment le processus d'innovation ne peut être réduit à une seule dimension technique, sociale, juridique, politique ou économique, mais dépend de toutes ces composantes à la fois. Dès lors qu'il possédait une peau de bête et connaissait les techniques lui permettant de la transformer en parchemin, un individu sachant écrire pouvait constituer un codex et le transmettre à ceux qui détenaient les clés des bibliothèques, espérant qu'ils accepteraient de consulter et d'archiver son message. En maîtrisant les techniques de production et d'écriture, celui-ci optimisait les chances pour son message d'atteindre certains destinataires dans l'espace et le temps, par l'intermédiaire de ceux qui en auraient décidé et assuré l'indexation, lesquels exerçaient alors une forme de *pouvoir* vis-à-vis aussi bien de l'émetteur du message que de ses destinataires.

2.1.5 A nouveau support, nouvelle présentation

En plus de faciliter la production des supports à grande échelle, l'invention de l'imprimerie entraîna à partir du XVI^e siècle une normalisation de ce qui se trouvait écrit dans les livres²⁵. La matérialité du support pouvait avoir une influence

²⁵ « Tout en conservant l'apparence générale du manuscrit relié, le processus de fabrication entraîne des changements dont la portée n'est pas perçue au départ. Ainsi la pagination des ouvrages s'est-elle imposée d'abord pour éviter au relieur d'intervertir l'ordre des feuillets, mais elle est rapidement devenue l'un des outils indispensables à la lecture elle-même. La page de titre, inconnue des manuscrits médiévaux, a été conçue pour donner d'un seul coup d'œil au lecteur potentiel une idée synthétique de l'ouvrage : c'est la première ouverture faite à la « publicité », entendue aussi bien dans son sens originel que dans son acception commerciale. L'imprimerie a eu un impact puissant sur la langue elle-même en obligeant à fixer la graphie des mots, auparavant anarchique, ouvrant ainsi l'âge des Académies : celles qui se créent en Italie puis en France se donnent comme objectif d'imposer le beau langage. Alors que le texte des manuscrits se présente comme un flux ininterrompu et longtemps peu ponctué, ce qui le rend presque indéchiffrable aux non-initiés, le livre imprimé, destiné à un cercle beaucoup plus vaste, met en ordre la pensée en adoptant des conventions visuelles stables » (Racine, 2010, p. 17-18).

sur le contenu, *et vice et versa*. Là encore, l'innovation prétendait répondre au besoin d'émettre, de conserver et de mobiliser les informations dans l'espace et le temps, et l'on ne pourrait l'étudier sans s'intéresser de près au contexte politique, social et économique de cette époque, bouleversée par la facilité avec laquelle on pouvait désormais transmettre, reproduire et trouver des informations. Notamment, l'imprimerie fut l'un des principaux vecteurs de la réforme protestante entreprise par le moine augustin Martin Luther²⁶.

Il est possible d'imaginer qu'un éditeur n'ayant pas respecté les règles de présentation du contenu, de pagination, de structure, de titre, parce qu'il préférerait continuer à concevoir des manuscrits moyenâgeux, risquait de ne pas délivrer son message, ou *moins*, aussi intéressant fût-il. Ainsi le travail de l'écrivain, celui de l'éditeur, du bibliothécaire et celui du chercheur avaient désormais, au moins partiellement, été reconfigurés. Les écrivains et les éditeurs pouvaient rechigner à l'idée d'user de la ponctuation, des conventions visuelles, d'imprimer une page de sommaire, d'obéir aux Académies, mais il ne leur en faudrait pas moins maîtriser ces techniques architextuelles s'ils souhaitaient optimiser leurs chances de *publiciser* leurs textes dans l'espace et le temps.

2.1.6 L'utopie de Paul Otlet

Aucun entrepôt n'a jamais pu contenir « toutes les informations du monde » à la manière de la Bibliothèque de Babel qu'imagina Jorge Luis Borges dans son recueil *Fictions* (1944). Par ailleurs, les bibliothèques n'étaient que dans certaines villes et l'on ne pouvait y accéder sans certaines autorisations parfois difficiles à obtenir. Les livres demeuraient des objets que le feu et l'eau détruisaient, en témoignent l'incendie de la Bibliothèque d'Alexandrie en 47 av. J.C. et, deux millénaires plus tard, l'inondation de la Bibliothèque de Florence en 1966. Dès lors, rien d'étonnant dans le fait que certains penseurs aient appelé de leurs vœux – voire *prédit* – la numérisation et la mise en place de dispositifs permettant de s'y retrouver dans le gigantesque fatras que pourrait constituer l'ensemble des connaissances

²⁶ On prête à Martin Luther la parenté de cette célèbre phrase : « Tout homme est pape une bible à la main », qu'il aurait prononcé en brandissant une bible imprimée.

réunies en un même endroit. L'exemple le plus probant en la matière est sans doute celui de Paul Otlet et Henri La Fontaine. Ces deux avocats pacifistes fondèrent en 1895 l'*Institut international de bibliographie* à Bruxelles, où ils conçurent le projet de constituer le « livre universel du savoir » en « comptabilisant au jour le jour le travail intellectuel des deux mondes » (Mattelart, 2009, p. 24). Dans son testament philosophique, *Traité de documentation* (1934), Paul Otlet imaginait ainsi l'avenir :

« La table de travail ne serait plus chargée d'aucun livre. À leur place se dresse un écran et à portée un téléphone. Là-bas au loin, dans un édifice immense, sont tous les livres et tous les renseignements. De là, on fait apparaître sur l'écran la page à lire pour connaître la réponse aux questions posées par téléphone, avec ou sans fil. Un écran serait double, quadruple ou décuple s'il s'agissait de multiplier les textes et les documents à confronter simultanément ; il y aurait un haut parleur si la vue devait être aidée par une donnée ouïe, si la vision devait être complétée par une audition. Utopie aujourd'hui, parce qu'elle n'existe encore nulle part, mais elle pourrait bien devenir la réalité pourvu que se perfectionnent encore nos méthodes et notre instrumentation. Et ce perfectionnement pourrait aller jusqu'à rendre automatique l'appel des documents sur l'écran, automatique aussi la projection consécutive. » (*cit. in* Levie, 2006)

Le scénario d'Otlet montre à quel point l'ordinateur, le réseau Internet et le média web constituent moins une révolution qu'un aboutissement de la science documentaire, *configuration sociotechnique* qui probablement en entraînera à son tour de nouvelles (Rebillard, 2007), toujours dans le but d'émettre, recevoir, stocker et mobiliser l'information sans l'altérer. En outre, il est intéressant de constater que dès qu'il pense au support – l'écran, le téléphone, le haut parleur – Otlet suggère également de « rendre automatique l'appel des documents sur l'écran, automatique aussi la projection consécutive », c'est-à-dire d'automatiser le processus opératoire de mobilisation du support.

Dans ce système utopiste, le documentaliste maîtrise des méthodes et une instrumentation compliquées, que très peu de gens à part lui sont capables de mettre en œuvre. Cela explique pourquoi, dans de nombreuses traditions, comme l'a très bien expliqué David Kaser (1962), le bibliothécaire est considéré comme le tout puissant gardien d'un lieu éternel. Il jouit d'un fort *pouvoir*. Les éditeurs souhaitant s'assurer que leurs messages demeurent devront s'en remettre à lui, à son savoir, son

savoir-faire, ses techniques. Et il est possible d'imaginer qu'un éditeur « graisse la patte » du bibliothécaire afin que ses documents figurent en bonne place. Il est même possible de supposer qu'un éditeur puisse écrire ce qui plaira au bibliothécaire, dans le seul but de voir ses ouvrages mieux classés et mis en valeur que ceux de ses concurrents, et ce même s'il s'avère que ces ouvrages ne plaisent pas aux lecteurs.

2.2 La science de la recherche d'information

2.2.1 Genèse

Malgré les avancées qu'il représentait, le livre était encore imparfait : volumineux, altérable, cher à produire, limité par le nombre de pages, difficile à transmettre, il ne pouvait recueillir les informations que sous forme de texte ou d'images et n'était pas capable d'assurer une transmission en direct. Ces imperfections expliquent en partie qu'on ait continué à innover dans le domaine des moyens de production, de transmission et de stockage. Au XIX^{ème} et durant la première moitié du XX^{ème}, des progrès considérables furent réalisés (Flichy, 1991). Mais au début des années 1970, les procédés de transmission analogiques avaient encore deux défauts majeurs : le stockage était peu aisé et le signal se dégradait en raison du bruit et des interférences, des distorsions. On passa alors progressivement de l'analogique au numérique, c'est-à-dire d'un procédé basé sur la transmission de signaux continus variant proportionnellement à la grandeur physique, à l'utilisation du code binaire (Balle et Cohen-Tanugi, 2001). Il s'agissait de l'application concrète de la théorie de l'information du mathématicien Claude Elwood Shannon. Les « bits » devinrent à l'information ce que les atomes sont à la matière : particules élémentaires (Negroponte, 1995). La numérisation ne constitua pas pour autant un changement radical : elle s'inscrivait dans une longue lignée de techniques dont le but avait été de permettre « de mieux contrôler des collections complexes et des flux d'informations » (Halavais, 2009, p. 11).

Puisqu'on pouvait transformer les informations en chiffres et qu'on possédait des ordinateurs de plus en plus puissants, il fut question d'avoir recours à l'outil mathématique pour se charger d'indexer puis de trouver les informations. Aussi

commença-t-on, dès le milieu du XX^{ème} siècle, à chercher à confier à la machine le travail autrefois fait à la main par le bibliothécaire. L'appellation « science de la recherche d'information » (*information retrieval*, SRI ci-après) fut employée pour la première fois en 1948 par l'informaticien Calvin Mooers tandis qu'il achevait son travail de thèse. Dix ans plus tard, la première conférence dédiée à ce thème se tint à Washington D.C.. Il s'agissait pour les pionniers – Cyril Cleverdon, Brian Campbell Vickery, Peter Luhn – de définir le meilleur moyen possible d'indexer un document numérique à un temps t afin d'être à même de le mobiliser à un temps $t+x$. Pour cela, la machine devrait réaliser trois étapes : connaître les documents, les indexer puis les mobiliser au moment de la requête. Il fallait concevoir un outil en mesure de consulter les informations contenues dans un document et de générer à partir d'elles certaines méta-informations, ou *descripteurs*, – mots clefs, notions, date – qui permettrait grâce à un procédé appelé « indexation matière » de savoir de quoi il était question dans le document, quand il avait été rédigé, par qui et dans quel but, afin de pouvoir finalement le retrouver.

En 1958, Luhn publia une thèse selon laquelle la fréquence d'utilisation d'un mot dans un document ainsi que sa position permettaient de mesurer le degré de pertinence du document en question. Ces deux paramètres constituèrent les prémices de la recherche par mots-clés. En 1968, Gerard Salton publia l'ouvrage canonique *Automatic Information Organization and Retrieval*, après quoi, durant les années 60, 70 et 80, l'analyse statistique des textes connut des années prolifiques, dont les projets les plus marquants furent sans doute *Cranfield* (1957/67), *Medlars* (1968), *Smart* (1971), ainsi que les *Stairs* (1985) et *Trec* (1992).

2.2.2 Algorithme d'appariement

Le mot « algorithme » provient de la traduction latine de la dernière partie du nom d'un mathématicien perse du neuvième siècle avant notre ère : Abu Abdullah Muhammad ibn Musa al-Khwarizmi. Selon l'*Encyclopaedia Britannica* (15ème édition), il désigne « les procédures mathématiques systématiques qui produisent – en un nombre fini d'étapes – la réponse à une question ou la solution à un problème ». Dans leur ouvrage *Notions sur les grammaires formelles*, Maurice Gross et André

Lentin définissent l'algorithme comme « une procédure mécanique qui, appliquée à une certaine classe de symboles (symboles d'entrée), fournit, éventuellement, un symbole de *sortie* » (Gross et Lentin, 1967, p. 43-44). Ces deux définitions nous permettent de comprendre à quoi sert un algorithme en SRI : dès lors que l'on peut établir le profil d'un document, on est en mesure d'indexer certains symboles d'entrée qu'on rapproche ensuite computationnellement des symboles procurés par la requête (l'algorithme les *apparie*), obtenant ainsi plusieurs symboles de sortie appelés, dans leur ensemble, « résultat ».

Pour que l'algorithme d'appariement fonctionne correctement, il est impératif de définir quels symboles seront considérés parmi la masse de ceux que contiennent les documents et la requête. Le concepteur de l'algorithme doit par conséquent faire des choix visant à ce que l'outil puisse appairer le mieux possible un document et une requête. Ainsi, si c'est le rôle de la machine de transformer l'information en symboles puis d'appairer ceux-là correctement, c'est le rôle de son concepteur de la paramétrer afin qu'elle puisse trouver lesquels sont pertinents et donner à l'utilisateur ce qu'il considèrera comme une réponse satisfaisante à la question posée.

Le résultat généré par un algorithme ne peut toutefois être considéré comme le seul fait de son concepteur : ceux qui émettent les signaux d'entrée, producteurs et chercheurs d'information, y ont leur part de responsabilité, conscients ou non du rôle qu'ils auront joué. Le chercheur décide des mots qu'il emploie lorsqu'il effectue sa requête. Quant au producteur, les choix qu'il fait au moment de concevoir un document influenceront, et cela qu'il le veuille ou non, l'interprétation qu'en produira le dispositif de SRI. Dès lors, il peut décider de concevoir son message en fonction de ce qu'il sait des choix formulés par celui qui a conçu l'algorithme, ainsi que de ce qu'il sait des intentions des chercheurs, pour tenter de fluidifier le processus opératoire et d'augmenter les probabilités pour son document d'être consulté : chercher à être trouvé.

2.2.3 Arrivée du web

Lorsque Tim Berners-Lee venait tout juste d'inventer le web, il comprit qu'il ne suffirait pas de proposer aux internautes de pouvoir naviguer grâce à des liens hypertextes. Il fallait également leur indiquer où se trouvaient les informations pertinentes. Nous l'avons vu avec l'exemple du *codex* : dès lors que l'on met au point une nouvelle manière de consulter les informations, il semble indispensable d'inventer une façon de s'y retrouver. Aussi Tim Berners-Lee décida-t-il de mettre en place *The Virtual Library*, catalogue de liens constitué à la main et servant à indiquer où l'information se trouvait à propos de tel ou tel sujet. Non seulement l'invention de Tim Berners-Lee permettrait de produire et de diffuser des informations mais, en plus, elle permettrait de mobiliser parmi ces informations celles que lui et ses équipes considéreraient comme pertinentes.

De leur côté, les chercheurs en SRI se sont très vite intéressés au défi que représentait pour eux ce qui semblait appelé à devenir un chaos informationnel. Puisqu'ils étaient déjà en mesure d'indexer des documents et d'apparier une requête et un résultat, il leur fallut concevoir des outils qui leur permettraient de récolter sur le web les symboles d'entrée à indexer. Des logiciels robots furent conçus, ou *crawleurs*²⁷, chargés de scanner les documents. Ceux-là étaient capables de générer des méta-informations pour indexer les données de manière à ce qu'elles puissent être mobilisées. L'ensemble du processus opératoire serait ainsi confié à la machine, la principale action humaine consistant à paramétrer le dispositif en amont et, en aval, à formuler une requête.

Mais les avancées en SRI avaient été faites à une époque où les informations étaient déjà classifiées par bases thématiques tandis qu'avec le web, les documents pouvaient être mis en ligne par n'importe qui, à propos de n'importe quoi, sans que rien n'oblige le producteur à respecter une quelconque convention. La question se posait donc du rôle joué par les éditeurs. Ainsi, le problème des premiers crawlers résidait dans le fait que les concepteurs des pages web devaient « prévenir » le robot

²⁷ Nous appellerons tout au long de notre thèse « crawler » ce qu'en anglais on appelle « web crawler ». Cette appellation du robot d'indexation nous permettra de distinguer sans ambiguïté le logiciel (crawler) du verbe désignant l'action de ce logiciel (crawler), tout en évitant le terme trop vague de « robot ».

s'ils voulaient figurer dans l'index²⁸. L'action *a priori* de l'éditeur conditionnait ainsi le fonctionnement du moteur. En 1993, les crawleurs *JumpStation*, *World Wide Web Worm* et *Repository-Based Software Engineering* furent mis au point. Ils archivaient automatiquement les titres, les accroches et les URL, mais les méta-informations n'étaient pas indexées de manière à pouvoir être retrouvées si le chercheur ne connaissait pas exactement les termes employés dans ces titres, accroches et URL. Là encore, l'éditeur devait y penser au moment de définir l'adresse des documents et d'en rédiger le titre et l'accroche.

Au fur et à mesure que le nombre de documents sur le web augmentait, des progrès furent faits en matière de paramétrage des crawleurs, des systèmes d'indexation et des procédures algorithmiques. De nombreux logiciels virent le jour dans les années 90²⁹. Le premier d'entre eux à crawler des documents entiers fut *Webcrawler*, qui permit de raffiner en nombre et en qualité les symboles d'entrée. *Altavista* fut le premier à permettre d'effectuer des recherches en langage naturel (« Quelle est la couleur du cheval d'Henri IV ? »). Et l'on vit apparaître les premiers méta-moteurs capables de combiner les résultats de plusieurs moteurs, avec notamment *Hotbot* (1996) et *Dogpile* (1996).

Toutefois, malgré le nombre de solutions proposées, les moteurs ne fonctionnaient pas encore suffisamment bien en 1998 pour répondre à toutes les requêtes, comme le montrèrent par exemple John Brown et Paul Duguid avec le moteur *Sherlock* (Brown et Duguid, 2000, p. 41-44) . Le mieux semblait donc encore de consulter manuellement les documents disponibles, puis de les thésauriser sous forme d'annuaire en effectuant un travail éditorial dont n'étaient pas capables les machines livrées à elles-mêmes.

²⁸ Ce fut le cas par exemple pour le crawler conçu par Matthew Gray développé dans le but d'observer la croissance du nombre de documents en ligne, ou encore celui que Martijn Koster développa dans le cadre de son projet *Archie-Like Indexing of the Web* (ALIWEB).

²⁹ *Excite* (1993), *Infoseek* (1994), *Lycos* (1994), *Webcrawler* (1994), *Altavista* (1995), *Excite* (1995), *Echo* (1996), *Ask Jeeves* (1997), *Google* (1998), *AllTheWeb* (1999)

2.2.4 Analyse des liens

A partir de 1996, certains chercheurs comme Robin Li³⁰, Jon Kleinberg, Larry Page et Sergey Brin planchèrent sur une méthode différente de celle qui consistait à analyser statistiquement le contenu des documents. Leur idée fut d'exploiter les informations générées par le média lui-même, le web et les liens hypertextes, c'est-à-dire de générer des méta-informations à partir de ce que l'on pouvait apprendre au sujet de l'emplacement topologique des informations. En ajoutant les symboles d'entrée liés à cet emplacement à ceux donnés par l'analyse statistique et la requête, ils cherchaient à obtenir des symboles de sortie plus pertinents. C'est ce qu'Olivier Ertzscheid désigne comme « la recherche d'information augmentée » (Ertzscheid *et al.*, 2009).

Jon Kleinberg détermina l'importance que l'on devait attribuer à la position des pages en se référant au mode d'organisation du *Science Citation Index* (SCI) et en particulier à la méthode bibliométrique conçue par Pinski et Narin à partir des travaux de Garfield, laquelle consistait à attribuer une valeur à une publication scientifique proportionnelle au nombre de publications qui la citaient. Il développa pour le laboratoire Almaden d'IBM l'algorithme HITS dont le fonctionnement était basé sur l'identification de certaines pages comme des « *authorities* », étant donné le nombre de pages pointant vers elles, et d'autres comme des « *hubs* », caractérisées par la présence de liens pointant vers les *authorities* (Kleinberg, 1999). Chaque document recevait ainsi deux valeurs : l'une utilisée pour déterminer la pertinence d'un document, l'autorité, et l'autre, *hub*, indiquant la pertinence des liens pointant de ce document vers d'autres documents. Ces mesures permettaient potentiellement de donner à la fois une idée de la façon dont le web était organisé et d'identifier au moment d'une requête certaines pages qui, probablement, contiendraient une information pertinente. La même année, l'équipe d'IBM proposa de représenter les liens hypertextes par la méthode mathématique des graphes orientés (Kleinberg *et al.*, 1999). Toujours au sein de ce laboratoire, le projet CLEVER (*Clientside Eigenvector*

³⁰ Robin Li est à l'origine de la mise au point de l'algorithme « RankDex », en 1996, avec lequel il contribua à créer le moteur chinois aujourd'hui leader dans ce pays : *Baidu*. Pour être concis, nous nous concentrons ici sur Kleinberg, Brin et Page, ce qui ne doit surtout rien enlever aux mérites de Li !

Enhanced Retrieval) visa à améliorer le fonctionnement de l'algorithme HITS en raffinant la définition de ce qu'étaient des *hubs* et des *authorities*. D'autres ajustement furent ensuite proposés, à l'issue du projet SALSA (*Stochastic Approach for Link-Structure Analysis*), permettant d'appliquer un méta-algorithme aux *hubs* et aux *authorities* afin de ne retenir que les plus pertinents. HITS fut ensuite utilisé par plusieurs moteurs destinés au grand public, notamment *Teoma* lancé en avril 2000. Le moteur *Wisenut* et la *Webfontain* d'IBM, tous deux conçus au début des années 2000, l'utilisaient également.

En 1997, Larry Page conçut un moteur nommé « *BackRub* », dont les crawlers étaient capables non pas seulement de suivre les liens allant d'une page vers une autre mais également de trouver, pour chaque page, les liens pointant vers elle, ou *back-links*. « En un sens, expliquait-il alors, sur le web n'importe qui peut annoter n'importe quoi grâce au lien hypertexte. Mais le problème avec les premières versions de la navigation hypertexte était de ne pas pouvoir suivre les liens en sens inverse » (Battelle, 2005, p. 72). Une fois *BackRub* en état de marche, Page eut l'idée de traiter les informations collectées afin de pouvoir les indexer et, le moment venu, les mobiliser³¹. C'est ainsi qu'il conçut avec Sergey Brin, doctorant comme lui à Stanford, l'algorithme *PageRank* : reprenant les idées formulées par Kleinberg, Larry Page et Sergey Brin considérèrent que chaque lien pointant vers une page était un vote pour cette page. Plus une page recevait de votes, plus elle était considérée comme pertinente, et plus son vote, lorsqu'elle pointait elle-même vers d'autres pages, était significatif (Brin et Page, 1998 ; Page *et al.*, 1999). Brin et Page couplèrent ensuite leur analyse topologique avec une méthode d'analyse de contenu (Bianchini *et al.*, 2005). Le texte de chaque page était ainsi analysé en entier, tandis qu'une attention particulière était consacrée au texte des liens eux-mêmes, appelé « ancrés ». Le moteur fut nommé *Google Search*. Sa grande nouveauté, en plus de donner des résultats jugés plus satisfaisants que ceux d'*Excite* et d'*AltaVista* (Battelle, 2005), était de considérer le web comme un *système social* à part entière : « Chaque document du corpus est considéré comme un membre d'un réseau ou d'une société stratifié(e), et ce avant même que la moindre requête n'ait été formulée. Le

³¹ Au départ, Page n'avait pas prévu de concevoir un moteur de recherche, mais très vite, cela lui parut envisageable étant donné les informations récoltées par *Backrub* (Battelle, 2005, p. 73-81).

concept central en SRI de *pertinence* – lié à un « besoin d’information » spécifique – est complété par le concept sociométrique de *statut* et d’*autorité* » (Rieder, 2012).

2.2.5 Fonctionnement opaque

Après que PageRank et HITS furent mis au point, et étant donné la croissance du nombre de documents sur le Web, les procédés de recherche automatique l’emportèrent sur l’indexation manuelle, moins coûteux, plus rapides et de plus en plus efficaces. Au fur et à mesure que le nombre d’utilisateurs de Google augmentait, les méthodes utilisées pour crawler, indexer et apparier les documents et les requêtes devinrent également de plus en plus opaques. Le moteur, peu à peu, devenait une *boîte noire*. Cela souleva dès le début des années 2000 certaines questions centrales à propos du *pouvoir* de ce nouveau type d’intermédiaire et de la médiation que, réellement, il opérait— questions qui n’ont pas cessées depuis lors (Brown et Duguid, 2000, p. 45 ; Mathias, 2009, p. 39-41)

2.3 ADN de Google

2.3.1 Université et entrepreneuriat

Avant que la bulle internet n’éclate, les étudiants de Stanford qui se lançaient dans des projets avaient à l’esprit que, peut-être, leurs inventions vaudraient plusieurs millions de dollars. S’intéresser à Google nécessite dès lors de comprendre qu’étant donné certaines caractéristiques propres à l’Université Stanford, son ADN est double : à la fois académique et entrepreneurial. D’une part, son patrimoine socioculturel est intimement lié à la recherche universitaire, son fonctionnement, ses exigences. D’autre part, il est lié à la figure de l’entrepreneur nord-américain, parti de rien et devenu millionnaire en quelques années³².

³² « Le campus de Stanford cumule la plupart des traits de caractère consubstantiels au dieu Google. Le premier d’entre ces traits est l’excellence académique et la vénération des rituels de transmission de connaissance qui l’accompagnent. La connaissance est en ce lieu le bien suprême, comme il l’est pour Google. Le deuxième trait tient à l’inscription profonde de Stanford dans le nouveau monde de la Silicon Valley et sa fièvre des start-up de nouvelles technologies de l’information et de la communication. [...] Point tout aussi essentiel, mais vécu sans doute de façon plus diverse par les étudiants : à ce culte de la start-up se mêlent en un paradoxe très américain un élitisme féroce, un esprit et une esthétique hippies ou du moins post-hippie et une générosité naïve très « néo-Kennedy » et « jeunes loups du Parti démocrate ». Ce mariage de ce que bien des Européens considèrent comme des éléments antinomiques éclaire la singularité de l’esprit Google : une entreprise

2.3.2 Ingénierie

Sergey Brin et Larry Page sont des ingénieurs, passionnés par l'innovation technique et peu regardant pour ce qui n'est pas du domaine de l'ingénierie. C'est en tout cas la thèse étayée par un de leurs anciens collaborateurs (Edwards, 2011), lequel, parce qu'il n'était pas ingénieur, avait tendance à se retrouver au ban des processus de décision.

Quand ils embauchèrent leurs premiers employés, Brin et Page choisirent des ingénieurs de Stanford ou d'universités comparables, adhérents aux mêmes valeurs qu'eux. Al Gore, qui fut consultant pour la firme en 2001, raconte à ce sujet : « Ceux qui attribuent le succès de Google à son algorithme ou à la loi des rendements d'échelle oublient de considérer à quel point le génie de Google réside également dans son recrutement d'employés extrêmement compétents et dans l'expérience que ces employés ont du travail chez Google [...] Ce n'est d'ailleurs pas seulement du recrutement et de la rétention de ces employés dont il s'agit, mais également de leur appartenance à une même communauté de valeur »³³ (Auletta, 2010, p. 22).

Pour Brin et Page, il était important dès le départ de travailler à des innovations commercialisables. En cela, ils étaient plus proches d'une démarche d'ingénieur que de la démarche académique, laquelle n'a pas toujours un but utilitariste et lucratif. « Mon but, explique Larry Page, était de travailler sur quelque chose de réel et d'intéressant sur le plan académique, mais il n'y avait pas de raison, parce que je me livrais à un travail universitaire, que je travaille sur des choses qui ne seraient pas pratiques. Il y a plein de problèmes intéressants qui sont aussi des problèmes pratiques. [...] J'ai donc décidé finalement de concevoir quelque chose qui serait

décomplexée, qui ne perçoit aucune contradiction entre sa soif de puissance économique et son utopie d'un monde gouverné par la connaissance » (Kyrrou, 2010, p.43).

³³ Those who attribute Google's success to its algorithms or the law of increasing returns, fail to fully appreciate the extent of which Google's superior talent recruitment stems from its unusual empowerment of employees and the attention they pay to the quality of the employee experience at Google [...] It's not only in the recruiting and retention of the higher quality employees. It also has to do with their alignment with community values, with trying to make the world a better place » (Auletta, 2010, p. 22, notre traduction)

potentiellement intéressant pour beaucoup de gens... et qu'ensuite je commercialiserais »³⁴ (Battelle, 2005, p. 74).

Au début des années 2000, les fondateurs de Google refusèrent les offres de rachat faites par les géants du monde des médias et des télécommunications (Van Couvering, 2008). « Ces compagnies, raconte Larry Page, ne prétendaient pas se concentrer sur le *search* - elles devenaient des portails. Elles ne comprenaient pas le *search*, et il n'y avait pas d'ingénieurs là-bas »³⁵ (Levy, 2011, p. 30). Ainsi, les deux fondateurs de Google semblent s'être toujours méfiés de ceux qui n'étaient pas des ingénieurs. L'univers médiatique des éditeurs leur était étranger, et ils ne souhaitent pas, ou ne souhaitaient pas au départ, y être liés.

Au moment de l'introduction en bourse de la firme en 2004, le PDG Eric Schmidt expliquait aux investisseurs : « Google est gérée par trois ingénieurs informaticiens. Nous allons donc faire toutes les erreurs que les ingénieurs informaticiens font quand ils gèrent une entreprise. Mais une des erreurs que nous n'allons pas faire est celle que font ceux qui ne sont pas ingénieurs. Nous allons faire des erreurs basées sur des faits, des données et des analyses »³⁶ (Auletta, 2010, p. 23).

La culture d'ingénieur informaticien est donc très présente dans les propos des dirigeants de Google³⁷. La performance technique en est la clef de voûte, et l'analyse statistique considérée comme étant l'appui le plus solide dès lors qu'il s'agit de

³⁴ My goals were to work on something that would be academically real and interesting, but there is no reason if you are doing academic work to work on things that are impractical. There are plenty of interesting problems that are also practical. I wanted both, and I didn't think there was much of a trade-off to be made. I figured if I ended up building something that was going to potentially benefit a lot of people... then I would commercializing it » (Battelle, 2005, p. 74, notre traduction).

³⁵ « These companies weren't going to focus on search – they were becoming portals. They didn't understand search, and they weren't technology people » (Levy, 2011, p. 30, notre traduction).

³⁶ « Google is run by three computer scientists. We're going to make all the mistakes computer scientists running a company would make. But one of the mistakes we're not going to make is the mistake that nonscientists make. We're going to make mistakes based on facts and data and analysis » (Auletta, 2010, p. 23, notre traduction).

³⁷ Michel Freyssenet a donné dès le début des années 90 une vision très éclairante de ce qu'une telle culture pouvait avoir de particulier : « [La culture « orientée ingénierie »] se manifeste tout d'abord par la conviction que la solution technique est toujours plus efficace et plus définitive que tout autre solution (organisationnelle, sociale, gestionnaire) pour élever rapidement la productivité ou pour résoudre un problème organisationnel ou social. La technique est vue ensuite comme l'application à un domaine utilitaire des lois scientifiques. Il ne peut donc y avoir qu'une seule bonne solution technique, celle qui applique correctement ces lois et dès lors s'impose avec la force de la vérité. L'idée de plusieurs techniques possibles et valables en fonction des objectifs poursuivis [...] est dans cette culture difficilement acceptable. / La vision d'une science établissant des lois de la nature qui la régiraient mécaniquement donne naissance à un idéal de la perfection technique évacuant l'aléa et l'incertitude. Plus un système est « bouclé », plus il réduit l'intervention humaine, plus il est censé être performant et parfait. L'ingénieur se doit donc de tout prévoir, [...] il ne peut laisser à l'appréciation suspecte et inconsciente de l'utilisateur le bon fonctionnement des machines qu'il conçoit. » (Freyssenet, 1990, p.11-12 *cit. in* Flichy, 2003, p. 66)

prendre une décision³⁸. C'est une culture « orientée ingénierie » (Auletta, 2010, p. 19).

Pour ce qui concerne le traitement de l'information actualité, nous verrons que les porte-parole de Google mettent en avant l'absence d'intervention humaine comme un gage de performance et de pertinence. Et c'est cette culture orientée ingénierie qui, à bien des reprises, entre en conflit avec la culture des entreprises issues du monde des médias.

2.3.3 Esprit Montessori

« We're Montessori kids. We've been trained
and programmed to question authority. »
(Levy, 2011, p. 122)

Sergey Brin et Larry Page ont la particularité d'avoir tous deux été élevés dans des écoles Montessori. Or la pédagogie Montessori, contrairement aux méthodes d'enseignement traditionnelles, veut que l'individu devienne ce qu'il est non pas grâce à ce que ses maîtres lui ont dit mais grâce à ses propres accomplissements. Cette pédagogie, fortement inspirée par Jean-Jacques Rousseau, consiste également à contester les autorités et les légitimités instituées. L'enseignant doit être un exemple d'humilité pour l'enfant, il le respecte et ne l'aide qu'à condition que cela soit nécessaire (Nourot, 1993, p. 15 ; Röhrs, 1994). Les enfants des écoles Montessori sont libres de choisir les ateliers auxquels ils participent, ce qui donne de l'eau au moulin des détracteurs de cette pédagogie selon lesquels l'éducation risque d'être incomplète et puérocentrique.

Au sujet de Brin et Page, Marissa Mayer, ex Vice-Présidente de Google, explique : « [L'esprit Montessori] a vraiment intégré leur personnalité. Poser leurs propres questions, faire leurs propres trucs. Ne pas respecter l'autorité. Faire quelque chose parce que ça a du sens, et non parce qu'une figure autoritaire vous a dit de le faire. Dans les écoles Montessori, vous peignez parce que vous avez quelque chose à

³⁸ Douglas Bowman, *designer* chez Google depuis 2006, souligne à ce sujet : « Quand une compagnie est gérée par des ingénieurs, chaque problème devient un problème d'ingénierie [...] Les données sont finalement au centre de chaque décision, paralysant la firme et l'empêchant de prendre certains virages audacieux en matière de design »³⁸ (Auletta, 2010, p. 20).

exprimer ou parce que vous voulez justement peindre cet après-midi là, et non pas parce qu'un professeur vous a dit de le faire. Ceci est vraiment intégré à la manière dont Larry et Sergey abordent les problèmes »³⁹ (Levy, 2011, p. 122). Quant à Sergey Brin, il dit lui-même : « Je pense que j'ai bénéficié grâce aux écoles Montessori d'une éducation qui donne aux étudiants la liberté de faire les choses à leur manière et de découvrir⁴⁰. [...] Je pense qu'une part de ma volonté de faire des choses qui vont dans mon propre intérêt, je la dois à l'éducation de Montessori »⁴¹.

Nous retrouvons une volonté d'émancipation très proche de « l'esprit Montessori » dans de nombreuses prises de positions des dirigeants de Google⁴². Ainsi, par exemple, Sergey Brin disait le 15 avril 2012 au *Guardian* : « Si nous pouvions d'un coup de baguette magique ne plus être sujets à la loi des Etats-Unis, ce serait bien. »⁴³ Dans le cas de la relation entre Google et les éditeurs, nous verrons comment, et dans quelle mesure, les fondateurs de Google et leurs employés ont pu contester aux médias traditionnels leur autorité.

2.3.4 La publicité et l'argent

Dans l'annexe A de la publication qui présentait le fonctionnement du moteur Google, Brin et Page se prononcèrent très clairement contre toute idée de partenariat avec des annonceurs publicitaires : « Nous pensons que des moteurs de recherche financés par la publicité seront forcément incités à travailler pour leurs annonceurs et non pas pour les besoins des consommateurs » (Brin et Page, 1998), opposant ainsi l'intérêt de l'annonceur et celui du consommateur.

³⁹ « It's really integrated in their personalities. To ask their own questions, do their own things. To disrespect authority. Do something because it makes sense, not because some authority figure told you. In Montessori school you go paint because you have something to express or you just want to do it that afternoon, nor because the teacher said so. This is really baked into how Larry and Sergey approach problems » (Levy, 2011, p. 122, notre traduction).

⁴⁰ On retrouve chez Brin la volonté d'expérimenter à Stanford ce dont il a envie, sans être pressé d'achever sa thèse, une liberté d'action qui fait penser aux écoles Montessori : « Sergey found Stanford an intellectual feast of opportunity for the curious. "I tried so many different things in grad school," he said. "The more you stumble around, the more likely you are to stumble across something valuable. » (Vise, 2005)

⁴¹ « I do think i benefited from the Montessori education which in some ways gives the students a lot more freedoms to do thing at their own pace, to discover [...] I do think that some of the credit for the willingness to go on your own interests – you can tie that back to the Montessori education », notre traduction, vu sur une interview vidéo : <http://www.youtube.com/watch?v=3OCAdXjlLBA> (dernière visite le 20 août 2013)

⁴² Il est probable que la remise en cause systématique des pouvoirs institués par les dirigeants-fondateurs de Google soit liée au fait que, comme le souligne Yves Cusset dans un article discutant de la méthode Montessori, l'école soit « le prodrome de l'espace public, et l'éducation qu'elle dispense appelle à libérer le potentiel d'autoréflexion dont chacun peut disposer pour se ménager un accès critique à ses propres intuitions éthiques et politiques » (Cusset, 2006).

⁴³ <http://www.guardian.co.uk/technology/2012/apr/15/web-freedom-threat-google-brin> (dernière visite le 20 août 2013)

Douze ans plus tard, la firme Google était devenu la plus importante régie publicitaire en ligne du monde (Idate, 2010). Cependant, il n'en demeure pas moins que l'entreprise semble avoir conservé un sentiment ambivalent à l'égard de la publicité. Par exemple, sur le site présentant la philosophie de l'entreprise⁴⁴, on trouve parmi « les dix points clefs de Google » un point nommé : *Il est possible de gagner de l'argent sans vendre son âme au diable*. Cette phrase laisse sous-entendre qu'en général, quand on gagne de l'argent, c'est qu'on a vendu son âme au diable, mais que la firme Google, elle, a trouvé comment rester « pure ». Toujours sur cette page présentant la philosophie d'entreprise, il est expliqué : « L'un des fondements de Google est [...] de bien distinguer les publicités des résultats sur les pages de résultats. Nous ne manipulons jamais les classements pour favoriser nos partenaires dans les résultats que nous fournissons. Personne ne peut payer pour obtenir un meilleur classement PageRank. Nos utilisateurs font confiance à l'objectivité de Google et rien ne pourrait justifier la remise en cause de cette confiance. » Ainsi, l'activité publicitaire est clairement séparée dans le discours de Google du fonctionnement du moteur, comme s'il s'agissait de deux métiers tout à fait différents, l'un à visée lucrative, l'autre à visée philanthropique. Nous verrons cependant (chapitre 5) pourquoi et comment il est malgré tout possible de nuancer une telle affirmation.

2.4 Comment Google est devenue leader sur le marché des moteurs

Elizabeth Van Couvering (2008) divise l'histoire du marché des moteurs de recherche en trois phases distinctes : *entrepreneurs technologiques* (1994-1997), *portails et intégration verticale* (1997-2001), *syndication et consolidation* (2001-...). Ces périodes sont intéressantes à résumer pour notre démarche d'historicisation. Nous y ajouterons deux périodes, afin de mieux cerner ce qui s'est produit pendant les années 2004-2012.

⁴⁴ <http://www.google.fr/about/company/philosophy/> (dernière visite le 20 août 2013)

2.4.1 Portails et intégration verticale (1997-2001)

La première des périodes identifiées par Van Couvering – *entrepreneurs technologiques* (1994-1997) – fut marquée par une prolifération d'acteurs au sein d'un environnement hautement compétitif. Les barrières à l'entrée étant peu nombreuses, de nouveaux acteurs firent leur apparition. Mais la plupart de ces acteurs étaient issus de projets universitaires effectués en SRI et n'avaient pas de modèle économique. La seconde période – *portails et intégration verticale* (1997-2001) – fut quant à elle marquée par l'arrivée sur le marché d'acteurs historiques des médias et des télécommunications, notamment les géants de l'édition et les Fournisseurs d'Accès à Internet (FAI). Détenteurs de portails, ceux-ci achetèrent et intégrèrent des moteurs à leurs plates-formes afin de développer des synergies en attirant une audience la plus large possible. Il s'agissait de « portails pleinement intégrés » (Meisel et Sullivan, 2000, p. 484). Cette période vit un certain nombre des détenteurs d'annuaires ou de moteurs être rachetés alors que de nouveaux acteurs indépendants comme *Google* et *AskJeeves* entraient sur le marché. Mais les portails étaient sujets aux suspicions étant donné les conflits d'intérêts possibles : ils prétendaient fournir du contenu tandis qu'ils proposaient de guider les internautes sur l'ensemble du web et donc, potentiellement, de les mener vers des sites concurrents. C'est ce qui explique en grande partie selon Elizabeth Van Couvering les échecs d'acteurs comme Disney après le rachat d'Infoseek ou @Home après le rachat d'Excite. Le conflit d'intérêts était mal perçu par les internautes, qui voyaient, derrière la boîte noire, la possible manipulation des résultats. La *délégation*, pour laquelle la confiance est fondamentale⁴⁵, était entravée. Ainsi, le rôle d'éditeur et celui de moteur semblaient dès le début des années 2000 difficiles à concilier.

2.4.2 Syndication et consolidation (2001-2004)

A la suite de l'éclatement de la bulle Internet, quand débuta la troisième période, *syndication et consolidation*, les acteurs historiques revendirent, souvent à perte, leurs moteurs et annuaires au plus offrant – dans la plupart des cas *Yahoo*, dont

⁴⁵ « La délégation n'est donc pas un acte singulier mais une relation continue, basée sur un aller-retour perpétuel et fondée sur un acte de confiance. » (Rieder, 2006, p. 92)

les deux acquisitions les plus importantes furent *Iktomi* en 2002 et *Overture* en 2003. Cette période fut marquée par la concentration en faveur de deux acteurs âgés de moins de dix ans et tous d'eux issus du campus de Stanford : Yahoo et Google.

Google et Yahoo avaient, et ont toujours, des stratégies sensiblement différentes. Google paria dès le début sur la recherche automatisée tandis que Yahoo fit d'abord le pari de l'annuaire tout en payant à Google le droit d'intégrer son moteur à ses pages. Etant donné le nombre de documents en ligne, et la difficulté croissante de les indexer manuellement, Yahoo finit par s'émanciper en acquérant des entreprises spécialistes du traitement automatisé et en mettant en ligne son propre moteur en 2004. Yahoo proposait sur son portail une interface complète, éditorialisée, indiquant à l'internaute dès la première page la plupart des services et des contenus auxquels il aurait accès sur Yahoo, tandis que Google proposait une interface la plus simple possible. Google opta ainsi pour une stratégie que Robert Boure et Nikos Smyrnaiois qualifient de « positionnement moteur », quand Yahoo opta pour un « positionnement média » (Smyrnaiois et Boure, 2006).

2.4.3 Diversification et différenciation (2004-2009)

De 2004 à 2009, période que nous proposons d'ajouter à celles identifiées par Elizabeth Van Couvering et de nommer *diversification et différenciation*, Google développa de nouveaux services et réalisa de nombreuses acquisitions. Son but était à la fois d'améliorer son moteur et de se lancer sur de nouveaux marchés. Les nouveaux services et les rachats soulignent son « positionnement moteur » dans la mesure où ces services fonctionnent en général de manière automatisée, y compris le versant publicitaire de ses activités sur lequel les annonceurs peuvent s'inscrire, fixer leurs prix et mener leurs campagnes sans qu'il n'y ait d'intervention manuelle de la part des employés de Google⁴⁶. Le métier de la recherche en ligne est au cœur de l'identité de Google (Girard, 2008), et la plupart des services que l'entreprise propose ont un lien avec la recherche d'information : ils permettent par exemple d'effectuer des recherches plus fines grâce à des moteurs verticaux, dont le champ d'indexation est restreint à certains sites ou à certains types de documents. Les services qui ne

⁴⁶ Nous y reviendrons dans le chapitre 5.

paraissent pas directement liés à la fonction recherche, comme la messagerie Gmail, permettent à Google de récolter des données sur le comportement et les centres d'intérêt d'un internaute et ainsi d'optimiser les réponses à ses requêtes lorsqu'il se rendra sur le moteur généraliste ou sur l'un des moteurs spécialisés. D'autres acquisitions et développements concernent quant à eux le versant publicitaire de son activité, dont le point d'orgue fut le rachat de la régie DoubleClick en 2008⁴⁷.

La firme *Yahoo*, quant à elle, continua durant ces années de se concentrer sur son positionnement média. En mars 2005, l'entreprise acheta le site de partage de photos *Flickr*, puis le site de « social bookmarking » *Del.icio.us* en décembre de la même année. L'entreprise acquit ensuite en 2006 le service d'édition vidéo en ligne *Jumpcut* et le service de création de concours *Bix*, puis le réseau social *MyBlogLog* en 2007. Yahoo multiplia les canaux d'informations spécialisés : *Yahoo! Finance*, *Yahoo! Sport*, *Yahoo! News* et *Yahoo! Pour Elles*. Le plus souvent ces sites avaient leurs propres rédactions, dont les contenus s'ajoutaient aux contenus agrégés par Yahoo à la suite de partenariats noués avec les éditeurs (Attias, 2007). Leur poids dans le paysage médiatique est indéniable : *Yahoo! Finance* était selon *comScore* le site d'actualités financières le plus visité aux Etats-Unis en 2008 devant les sites de *CNN*, *Dow Jones* ou *Reuters*⁴⁸, avec des revenus de 110 millions de dollars⁴⁹. *Yahoo! Finance* a noué des partenariats avec plus de deux cent éditeurs et Nathan Richardson, qui dirigeait la chaîne en 2005, a été élu cette même année la personnalité la plus influente du web financier⁵⁰.

Alors que Google rechigne à négocier avec les éditeurs (Smyrnaio et Rebillard, 2009), préférant que l'ensemble de la procédure d'indexation demeure algorithmique, Yahoo noue des partenariats comparables aux contrats établis par un éditeur de presse qui externaliserait une partie de sa production (Attias, 2007). Se voulant plus qu'un relais d'information, Yahoo choisit d'en être également la source. Tout en continuant à œuvrer dans le domaine de la recherche algorithmique, son but était ainsi de

⁴⁷ Nous décrivons le modèle économique de Google dans le détail lors du chapitre 5.

⁴⁸ http://www.comscore.com/Press_Events/Press_Releases/2008/07/Yahoo!_Finance_Top_Financial_News_and_Research_Site_in_US (dernière visite le 20 août 2013)

⁴⁹ Ce chiffre provient d'une interview donnée au magazine *Forbes* par l'ancien responsable de *Yahoo! Finance*, Nathan Richardson : http://www.forbes.com/2009/02/03/nathan-richardson-Yahoo!-business-media-0203_richardson.html (dernière visite le 20 août 2013)

⁵⁰ *Ibid.*

conserver les internautes sur son site, ce qui n'est pas sans faire penser aux stratégies menées par les acteurs historiques au début des années 2000 et aux conflits d'intérêts qui leur ont alors été reprochés. Mais sur le marché des moteurs, malgré les énormes investissements effectués par Yahoo à partir de 2003, Google semblait l'avoir largement emporté avec, en 2007, 81 % des parts au Royaume-Uni, 60 % des parts aux Etats-Unis et 84 % des parts en Australie, contre 4%, 19% et 3% pour Yahoo dans ces mêmes pays (Pollock, 2010).

2.4.4 Monopolisation (2009-...)

Après avoir acquis *Powerset* en 2008, Microsoft lança le 1^{er} juin 2009 le moteur de recherche automatique Bing. Microsoft avait alors son propre portail appelé MSN mais, contrairement à Yahoo, préféra attribuer à son moteur un site dédié, différent de son portail, portant un nom où ne figurait pas le nom de Microsoft et doté d'une interface simple. Il s'agissait de ce qu'on pourrait identifier comme une stratégie d'imitation de Google, visant à dissimuler les conflits d'intérêts possibles dès lors que Microsoft était également éditeur de contenus. Si l'on s'en réfère à la typologie stratégique de Nikos Smyrniotis et Robert Boure (*op. cit.*), Bing est ainsi l'objet d'un « positionnement moteur » plutôt que d'un « positionnement média ».

Un mois après la mise en ligne de Bing, et à la suite du rachat avorté de Yahoo par Microsoft, le positionnement moteur l'emporta sur le marché lorsque Yahoo annonça que le moteur hébergé sur ses pages serait désormais Bing, avec lequel un partenariat fut constitué, l'entreprise préférant se concentrer sur d'autres activités, notamment la production de contenus. Carol Bartz, alors dirigeante de Yahoo, alla jusqu'à nier que Yahoo ait jamais été un moteur de recherche⁵¹. Bruxelles et Washington approuvèrent ce rapprochement malgré la note de mars 2010 du Conseil de l'Europe pointant du doigt les risques potentiels de concentration sur le marché.

Le marché était concentré, en effet, et l'est toujours. Ainsi, selon le cabinet d'études *AT Internet*, Google possédait plus de 90% des parts du marché européen en juin 2013, les parts de Bing et Yahoo cumulés ne dépassant pas les 6% (AT Internet,

⁵¹ http://bits.blogs.nytimes.com/2009/08/07/yahoo-ceo-we-have-never-been-a-search-company/?_r=0 (dernière visite le 20 août 2013)

2013). Il n'y a qu'aux Etats-Unis que les parts cumulées de Bing et Yahoo peuvent dépasser les 30% tandis que celle de Google est à plus de 65%. En Chine, en revanche, Google est loin derrière *Baidu*. Mais il n'empêche : si l'on agrège le nombre de requêtes dans le monde, Google en reçoit environ 85% (ZenithOptimedia, 2011). La rareté de la concurrence s'explique en premier lieu parce que le nombre de documents en ligne a explosé depuis le milieu des années 90 et que, par conséquent, les infrastructures à mettre en place pour crawler et indexer ces informations coûteraient extrêmement cher. Ces coûts fixes agissent comme une « barrière à l'entrée » – que seuls des géants comme Microsoft peuvent franchir⁵².

Il semblerait toutefois que les coûts fixes n'expliquent pas à eux seuls la position de quasi-monopole de Google. En effet, il est possible de considérer que ce quasi-monopole est structurel étant donné les « effets de réseaux »⁵³ dont la firme profite grâce à sa position sur le marché publicitaire. Déjà dans les années 60, l'économiste George Stigler expliquait que « dès lors que le coût de collecte de l'information est (approximativement) indépendant de son usage, [...] il existe une forte tendance au monopole en matière d'approvisionnement en informations : en général, il n'y aura plus qu'une source standard »⁵⁴ (Stigler, 1961, p. 220). C'est ce qui semble s'être passé dans le cas de Google, dès lors que « les usagers valorisent la qualité mais ne sont pas facturés, tandis que les annonceurs valorisent les usagers et sont facturés, [ce qui] explique la nature très concentrée du marché des moteurs et rend probable une évolution du marché vers le monopole »⁵⁵ (Pollock, 2010, p. 28)

⁵² « Le seuil d'entrée pour un nouvel acteur sur [le marché des moteurs de recherche] est extrêmement élevé parce que pour proposer un produit intéressant, il faut un énorme investissement matériel et intellectuel, ce qui exclut tout acteur économique qui ne serait pas déjà extrêmement puissant ; même Microsoft, entreprise qui détient une énorme capacité financière et une base installée sans pareille, peine à mettre en cause la domination de Google. » (Rieder, 2006, p. 299)

⁵³ Nous définirons ce terme en détail dans le chapitre 5.

⁵⁴ « [s]ince the cost of collection of information is (approximately) independent of its use [...], there is a strong tendency toward monopoly in the provision of information: in general, there will be a "standard" source » (Stigler, 1961, p. 220, notre traduction)

⁵⁵ « users care about quality but are not charged, while advertisers care about users and are charged, explains the highly concentrated nature of the search engine market and makes it probable that the market will continue to evolve down this path towards monopoly » (Pollock, 2010, p. 28, notre traduction).

En plus des coûts fixes et des effets de réseaux, il existe trois autres barrières à l'entrée protégeant la position dominante de Google :

1. Il existe une « habitude » des internautes vis-à-vis de Google. Certes, ils peuvent changer de moteur en un seul clic, mais ils supporteront alors un coût d'opportunité. Jansen *et al.* (2007) attribuent cela à un « effet de marque » en montrant que, même lorsque les résultats donnés par différents moteurs sont les mêmes, les internautes font davantage confiance à Google et jugent préférables les résultats qui leur sont donnés par ce moteur. Par ailleurs, Rufus Pollock souligne que pour les usagers utilisant les paramètres de requête avancés, souvent spécifiques pour un moteur donné, il y a un « coût de réapprentissage » dans le cas d'un changement de moteur, qui contribue à fortifier la position du moteur dominant (Pollock, 2010, p. 17). Par ailleurs, il avait été souligné dès la fin des années 90 que les utilisateurs n'avaient pas toujours le savoir nécessaire pour changer de moteur (Telang *et al.*, 1999).
2. Google a noué de nombreux partenariats avec des sites auquel il fournit sa technique. Sur ces sites apparaît en général une barre à l'intérieur de laquelle le visiteur peut formuler une requête. Ces sites, s'ils voulaient changer de technique, auraient également un coût d'opportunité à supporter : mettre en place un nouveau partenariat, l'implémenter et s'assurer que leurs visiteurs sont au moins aussi satisfaits qu'avec Google. Cela contribue à fortifier la position de Google Search sur le marché. Il nous faut ajouter à cela les partenariats avec les navigateurs, comme Mozilla⁵⁶ et Opera⁵⁷, qui font de Google leur moteur par défaut. Là encore, il y a un coût d'opportunité pour les navigateurs qui changeraient de fournisseur. Il leur faudrait négocier un nouveau partenariat, et risquer alors de décevoir leurs usagers, soit parce que la technique sera jugée moins performante, soit à cause de l'effet de marque.

⁵⁶ <http://blog.mozilla.org/blog/2011/12/20/mozilla-and-google-sign-new-agreement-for-default-search-in-firefox/> (dernière visite le 20 août 2013)

⁵⁷ <http://www.generation-nt.com/opera-google-accord-actualite-1616942.html> (dernière visite le 20 août 2013)

3. Les données récoltées par Google à propos du comportement des internautes, qui ne sont pas rendues publiques, lui permettent de jouir d'un avantage concurrentiel décisif par rapport à la concurrence réelle et potentielle, lui assurant d'être le meilleur moteur, non pas parce que sa technique est la plus performante, ou pas seulement, mais aussi parce que ses données sont les plus nombreuses (Argenton et Prüfer, 2012).

Comme l'ont remarqué John Brown et Paul Duguid dès le début des années 2000, la *désintermédiation* ne signifie donc pas nécessairement qu'il n'y a plus d'intermédiaire, mais peut signifier que l'intermédiation est le fait de moins d'acteurs qu'auparavant. Dès lors, ceux qui continuent d'assurer cette fonction exercent un *pouvoir* plus grand que leurs prédécesseurs, car leur contrôle est centralisé (Brown et Duguid, 2000, p.28⁵⁸). Il y a eu, finalement, un double mouvement de *réintermédiation*, au sens d'Enrique Bustamante (Bustamante, 2004), et de *concentration*.

*

Maintenant que nous avons remis en perspective la position de Google sur le web, nous procéderons de même pour ce qui est de la diffusion de la production journalistique afin d'obtenir un point de vue panoramique sur la généalogie des deux univers qui se rencontrent au sein du dispositif d'infomédiation dont nous avons fait notre objet d'étude.

⁵⁸ Cela rejoint également l'observation de Fukuyama et Shulsky selon lesquels les NTIC et la désintermédiation n'ont pas produit la décentralisation qu'on attendait d'elles (Fukuyama et Shulsky, 1997).

CHAPITRE 3. HISTOIRE DE LA DISTRIBUTION-DIFFUSION DES NOUVELLES

« [La distribution de la presse] appelle un régime spécifique qui assure, tout à la fois, la liberté mais aussi une certaine égalité des chances et des conditions. Ce sont là des préoccupations qui, dans d'autres cadres ou dans d'autres domaines, ont retrouvé aujourd'hui une certaine actualité ! »

(Derieux, 2002, p.382)

Dans ce chapitre, nous procédons à une phylogénèse de la diffusion de la presse imprimée, afin de montrer comment, depuis que l'information d'actualité existe, sa production est en partie conditionnée par les procédés de publication et de diffusion, et leur cadre indifféremment économique et légal, politique, historique, social, technique. A la fin du chapitre, nous présentons en particulier l'histoire de *L'Express*, afin de mettre en perspective la frange monographique de notre terrain.

3.1 Parler à des statues parlantes

Nous prendrons ici divers exemples tirés de l'histoire de la presse imprimée dans le but de situer les relations entre les éditeurs de presse en ligne et le dispositif de distribution-diffusion⁵⁹. Nous prenons volontairement certaines situations dans lesquelles le pouvoir de censure exercé par les intermédiaires est fort afin de montrer comment cela peut influencer les éditeurs. Nous expliquons également que lorsque la censure est moindre, les rapports de forces demeurent malgré tout.

3.1.1 Place Pasquino

Dès le XIV^{ème} siècle en Europe⁶⁰, les nouvelles devinrent des marchandises. En Italie, les *menanti* organisaient pour les princes et les marchands des services de correspondance manuscrite (Albert, 2010). Ils distribuaient des *avvisi* aux abonnés,

⁵⁹ Si nous avons choisi de nous concentrer sur le cas de la presse imprimée, c'est parce que notre terrain est exclusivement composé de sites issus de la presse imprimée. Cependant, nul doute qu'une telle analyse socio-historique aurait également été très intéressante dans les cas de la radio et de la télévision.

⁶⁰ En Europe, car en Chine, où l'on avait inventé l'imprimerie mille ans plus tôt, la dynastie des T'ang, régnant de 618 à 907, lança une gazette officielle qui s'appelait le *Ti Pao* (Jeanneney, 2011, p.29).

assurant à la fois la rédaction et la distribution de ces nouvelles mises en circulation une ou deux fois par semaine (Dooley, 1999, p. 1317). En plus d'avoir en commun leur caractère politique, les *avvisi* avaient la même forme, qui demeura inchangée pendant plus de deux siècles.

A Rome, au XVI^{ème} siècle, l'habitude du peuple d'utiliser l'écrit garantissait aux *avvisi* un public dans chaque classe sociale et corps de métier. Partout dans la ville, des *avvisi* étaient placardés, parfois illégaux, souvent anonymes. Les auteurs choisissaient en général certaines places et rues où ils étaient sûrs que de nombreux Romains passeraient. Ils augmentaient ainsi leurs chances d'être lus et commentés (Dooley, *op.cit.*). La tradition voulait notamment qu'on recouvre les statues de *Pasquino* et de *Marforio* de divers pamphlets à propos des gouvernants, de leurs femmes, de la religion et des prélats. Le passant qui s'arrêtait devant ces « statues parlantes » savait qu'il y trouverait de l'information, vraie ou fausse, renouvelée, variée, à propos de la cité. Ainsi, les auteurs-afficheurs, grâce à cette forme de publication autoritative⁶¹, optimisaient leurs chances d'être lus.

Brendan Dooley explique que le nombre de lecteurs potentiels des *avvisi* était considérable, car après avoir quitté la boutique de leur inventeur, les *avvisi* « acquéraient une vie bien à eux et passaient entre les mains de copistes qui les reproduisaient maintes fois » (Dooley, 1999, p. 1321). Cela montre comment, déjà au XVII^{ème} siècle, l'information d'actualité était recopiée, agrégée, compilée, échangée et commentée. Des légions de livreurs disséminaient les *avvisi* tandis que des colporteurs les lisaient à haute voix sur la place Navone et la place Pasquino. Le travail de ces livreurs et de ces colporteurs complétait ainsi celui de l'auteur, de l'éditeur et de l'imprimeur. Les *avvisi* nourrissaient grâce à eux un contre-pouvoir. L'affaire Cenci⁶² en témoigne, à l'issue de laquelle le pape Clément VIII donna l'ordre de fouiller toutes les boutiques et les échoppes des producteurs d'*avvisi*, copistes, libraires, imprimeurs, bibliothécaires, colporteurs, afin de supprimer définitivement les documents qui le mettaient en cause. Les autres villes d'Europe

⁶¹ La pratique autoritative désigne « la pratique selon laquelle l'auteur (ou les auteurs) s'auto-publie sans passer par l'assentiment d'un médiateur extérieur » (Rebillard, 2006, p.57).

⁶² Le peuple de Rome demanda au Pape Clément VIII de gracier les assassins de Francesco Cenci (1598), étant donné les crimes dont l'aristocratie s'était rendu coupable à leur endroit. Cependant, le Pape refusa la grâce et les fit exécuter. L'une d'entre eux, Béatrice Cenci, fille de Francesco Cenci, est devenue ce jour-là un symbole italien de la résistance aux abus de pouvoir perpétrés par l'aristocratie.

avaient elles aussi leurs *avvisi* : les *occasionnels* en France et les *zeitungen* en Allemagne. Elles avaient leurs places Pasquino, leurs libraires, leurs bistrots : ces lieux où l'on prenait connaissance de l'actualité.

Ainsi, il existe depuis plusieurs siècles des lieux privilégiés pour consulter les nouvelles, les partager et les comparer. Cela, souvent, sans payer, ou en tout cas en payant moins que pour posséder le support où les nouvelles avaient été fixées. Par exemple, le chevalier de Meslé créa en 1751 des bureaux d'adresse où il était possible de lire *La Gazette* sans acheter d'exemplaire. Certains libraires louaient quant à eux la lecture à leur clientèle, ou leur faisait la lecture, dès 1650. De surcroît, diverses sociétés, notamment dans la deuxième moitié du XVII^{ème} siècle, s'abonnaient à frais communs à plusieurs titres et se retrouvaient dans des « chambres de lectures » pour lire et commenter l'information d'actualité (Cooper-Richet, 2002, p. 132).

Si ces sociétés et ces cabinets accueillaient le plus souvent des nobles et des bourgeois, le peuple, quant à lui, organisait des lectures orales et des conversations « chez les limonadiers, dans les tabagies ou les tripots, l'équivalent de nos cafés. On les trouve aussi dans la rue, sur les places des villes, lors des grandes tensions de l'opinion » (Feyel, 2007, p. 33). A partir du XIX^{ème} siècle, les cafés, restaurants, auberges, s'abonnent aux journaux dans le but d'offrir à leur clientèle un service supplémentaire (Albert, 2002, p. 155). Ainsi, la « place Pasquino », où l'on prend connaissance des nouvelles, où on les compare, les partage et les commente, sans payer ou en payant moins que si l'on achetait directement le journal, existe depuis que les nouvelles existent : elles sont *concomitantes*.

3.1.2 Monopole des maîtres-courrier

En 1630, soit un an avant la création de la *Gazette* par Théophraste Renaudot, Richelieu créa une surintendance générale des postes, dont le surintendant nomma un maître-courrier dans chaque chef-lieu du territoire. Les gazettes, une fois créées, purent ainsi être acheminées rapidement par une poste en situation de monopole. Le surintendant et ses maîtres-courrier, forts d'une organisation politique, économique et technique extrêmement efficace pour l'époque, exerçaient au nom du Roi un pouvoir de censure sur l'acheminement des messages. Nous apprenons ici deux choses

fondamentales : 1) la création de la presse périodique est concomitante à celle d'un moyen d'acheminer et de mobiliser rapidement les supports 2) ce moyen (ou *média*) est l'objet d'un monopole avec lequel les éditeurs devront composer s'ils veulent être sûrs de satisfaire leurs abonnés. Ces abonnés, qui étaient au début quelques privilégiés, devinrent de plus en plus nombreux dans la première moitié du XVIII^{ème}, à Paris aussi bien qu'en province (Feyel, 2007).

Dès le départ, Renaudot avait compris de quel pouvoir étaient détenteurs ceux qui maîtrisaient les réseaux d'acheminement, cependant il était persuadé que si ces réseaux empêchaient les gazettes d'arriver à bon port, celles-ci emprunteraient d'autres voies. Aussi demanda-t-il aux princes et aux Etats étrangers de « ne perdre point inutilement le temps à vouloir fermer le passage [aux] nouvelles dont le commerce ne s'est jamais pu défendre et qui tient en cela de la nature des torrents qu'il se grossit par la résistance » (Albert, 2010, p. 11). L'histoire de la presse lui a donné raison, mais il n'empêche : ne pas s'entendre avec les intermédiaires contrevient à la diffusion optimale d'un titre, laquelle permettrait aux nouvelles d'accéder à un public le plus large possible.

En 1740, un imprimeur avignonnais signa avec la Poste royale le premier contrat d'abonnement dans le but que ce ne soient plus les abonnés qui payent la Poste lors de la réception, mais l'imprimeur lui-même, pouvant ainsi proposer des abonnements « franco de port ». Les négociations avec l'intermédiaire se dérouleraient désormais du côté du producteur. Se posa alors le problème de la taxe postale, car certains éditeurs bénéficiaient d'un traitement de faveur (Feyel, 2007, p. 49). On finit par instaurer un traitement similaire pour tous les éditeurs autorisés à la suite de la délibération du 13 novembre 1763. Selon l'article 14 de cette délibération, les ouvrages périodiques devaient être envoyés « sous bande », chaque exemplaire se trouvant dans un « paquet séparé ». L'Intendant général des postes et ses administrateurs devaient quant à eux recevoir gratuitement et systématiquement un exemplaire de chaque ouvrage. Il fut indiqué que l'abonnement n'aurait lieu « qu'autant de temps qu'il plaira[it] à MM. Les administrateurs de vouloir bien continuer de l'accorder » (Feyel, 2002, p. 17). Nous trouvons ici l'idée d'un pouvoir de censure exercé par l'intermédiaire en situation de monopole, qui reçoit

gratuitement un exemplaire de chaque ouvrage et accorde aux éditeurs, ou non, la grâce de l'abonnement.

3.1.3 Contournements

Les rois du XVII^{ème} et du XVIII^{ème} siècles demandaient à leurs douaniers de contrôler la presse qui franchissait la frontière, et notamment la presse politique, car de nombreux éditeurs désireux d'éviter la censure et des prix jugés trop élevés imprimaient leurs feuilles à l'étranger et cherchaient ensuite à les rapatrier en France. Pierre Bayle publia ainsi en Hollande les *Nouvelles de la République des lettres* entre 1684 et 1687. En plus d'imprimer à l'étranger, il choisit de donner à ses feuilles un caractère littéraire qui serait plus facilement toléré par les douaniers, lesquels censuraient surtout la presse politique (Jeanneney, 2011). Malgré leur caractère littéraire, les critiques d'ouvrages étaient l'occasion d'une remise en cause virulente des politiques menées par le Roi. Ainsi, certaines contraintes liées à l'acheminement des nouvelles pouvaient avoir une influence, parfois subtile, sur la ligne éditoriale d'un titre autant que sur les méthodes d'acheminement.

Par ailleurs, la pratique de réimpression de certaines gazettes, qui permettait d'imprimer sur place à partir d'un modèle plutôt que d'importer en quantité, diminuait les coûts liés à la Poste et fut largement pratiquée en France⁶³. Elle finit par être interdite pour les journaux Parisiens, qui eurent alors l'obligation de passer par voie postale pour être diffusés en province, ce qui permit aux gouvernements du XVIII^{ème} et du XIX^{ème} siècles d'utiliser la taxe postale pour contraindre la presse (Feyel, 2007. p. 68).

Enfin, il est intéressant de noter qu'au temps du monopole postal, ceux qui transportaient des colis en essayant de contourner le monopole étaient tenus pour responsables de l'infraction au même titre que ceux qui imprimaient et empaquetaient les colis, et ce même s'ils ignoraient leurs contenus (Bataillé, 2002, p. 58-59). Nous reparlerons dans le cas de Google du problème lié à la responsabilité d'un intermédiaire vis-à-vis du contenu qu'il met en lumière et/ou achemine (chapitre 6).

⁶³ L'absence de copyright jusqu'en 1852 entre la France et l'Angleterre ont rendu certaines de ces réimpressions très lucratives, notamment pour Giovanni Antonio Galignani qui rééditait à Paris les publications britanniques à moindre frais et en faisait bénéficier une clientèle luxueuse, parmi laquelle Stendhal figurait (Cooper-Richet, 2002, p. 124).

3.1.4 Fin du monopole postal

Le monopole postal, pour les journaux politiques, dura en France jusqu'en 1870 alors que celui des autres journaux avait été aboli en 1856. Cette suppression visait à faciliter la circulation des nouvelles, afin de pratiquer « à fond la liberté de la presse », comme le souhaitait Gambetta, et de diffuser au mieux les idées républicaines (Bataillé, 2002). Elle faisait suite également à la difficulté, exprimée par la délégation de Tours, de différencier les journaux politiques de ceux qui ne l'étaient pas, les éditeurs jouant avec les définitions et essayant ainsi de se soustraire au monopole en réorientant leurs lignes éditoriales. Durant la Troisième République, la Poste, dès lors qu'elle n'avait plus le monopole, baissa ses tarifs. Ce fut notamment le cas après la loi de 1878 promulguée dans une logique de libre concurrence. A la même période, le développement des chemins de fer permit d'atteindre plus rapidement encore les abonnés, et plus largement (Muhlmann, 2004, p. 19).

Dès qu'ils l'ont pu, les éditeurs de presse imprimée ont tenté de maîtriser leurs réseaux de distribution, cela dans le but d'éviter, en plus de la censure, nombre de retards, et pour ne plus dépendre des taxes fixées par la Poste. Ainsi, dès sa création en 1863, *Le Petit Journal* mit en place « une campagne de propagande et de vente en s'appuyant sur un réseau ramifié de dépositaires régionaux et locaux » (Taveaux-Grandpierre, 2002, p.165). Il effectuait la vente au numéro, une méthode de diffusion qui permettait d'échapper aux contraintes des Postes. Des colporteurs du *Petit Journal* parcouraient les rues « jouant de la trompette, criant le titre du journal, offrant d'une voix retentissante les primes, annonçant le feuilleton nouveau » (Morierval, 1934, p. 200, cit. in. *ibid.*). Ceux-là mettaient en lumière le contenu journalistique, et appâtaient ainsi les lecteurs potentiels.

3.1.5 La stratégie Hachette

Pour ce qui est de la vente au numéro, dépassant dans la deuxième moitié du XIX^{ème} siècle la vente par abonnement et marquant les débuts de la consommation « individuelle » des nouvelles (Albert, 2002), différentes messageries de presse se créèrent à partir de 1848 et jusqu'à la fin du XIX^{ème} siècle. Elles approvisionnaient kiosquiers, libraires et colporteurs, et leur activité finit par faire l'objet d'un

« quasi-monopole »⁶⁴ lorsque Hachette eut racheté les *Messageries quotidiennes des journaux* et la société *Faivre* (ex-messagerie du Figaro) (Albert, 2002, p. 153).

Dès lors, la plupart des éditeurs furent obligés de se lier par contrat avec Hachette. En contrepartie, l'entreprise s'engageait à ce que leurs titres soient présents le plus rapidement possible chez les dépositaires et mis en vente en même temps que les concurrents. Elle acceptait également le droit d'inspection, les administrations de journaux étant dotées d'un droit de contrôle « sur la prise de papier, sur le transport, les conditions de vente et le retour » (Taveaux-Grandpierre, 2002, p.180-181).

3.1.6 Presse autorisée et presse clandestine

Durant la seconde guerre mondiale, les journalistes qui voulaient être certains de ne pas être censurés devaient accepter – et servir – l'idéologie officielle (Mathien, 2007, p. 112). Un journal comme *Le Cri du Peuple*, organe de doctrine créé par Jacques Doriot, voyait ainsi sa diffusion protégée par la sécurité militaire allemande et encouragée par la *Propaganda-Abteilung* (Feyel, 2007, p. 176). *Le Figaro*, quant à lui, refusant de respecter les consignes dictées par l'Allemagne après l'invasion de la zone Sud, souhaita annoncer sa disparition dans le numéro du 11 novembre 1942, mais celui-ci fut saisi et sa diffusion empêchée (D'Almeida et Delporte, 2003, p. 128). La presse clandestine se mit en place dès 1940, polycopiée en secret, dactylographiée, ronéotypée, multigraphiée, distribuée dans les boîtes aux lettres la nuit souvent par ses auteurs eux-mêmes. Ceux-là s'assuraient également de la complicité de certains imprimeurs et ouvriers du livres, linotypistes, typographes, clicheurs (*op. cit.*, p.137-140). « Au total, plus de 1200 feuilles clandestines paraissent en France, entre 1940 et 1944, parfois limitées à un seul numéro, parfois locales ou conçues pour une catégorie professionnelle » (*ibid.*). Etant donné les coûts et les difficultés d'approvisionnement en papier, ces journalistes de l'ombre devaient faire bref et percutant. C'est pourquoi les poèmes furent une forme d'expression très utilisée par les Résistants. A la Libération, les résistants reprochèrent d'ailleurs au collaborationniste Robert Brasillach d'avoir écrit et publié les *Poèmes de Fresnes* car

⁶⁴ « En reliant ses services d'approvisionnement des bibliothèques et les Messageries de journaux qu'elle vient d'acquérir, la maison Hachette devient incontournable dans la diffusion de la presse, dont elle assure ainsi la vente dans l'espace public, mais aussi privé. Hachette est indispensable... » (Taveaux-Grandpierre, 2002, p.177).

il s'agissait selon eux d'une forme d'expression qui, en quelque sorte, leur *appartenait* (Kaplan, 2001). On comprend ainsi comment des contraintes liées à la distribution peuvent aboutir dans certains cas à *l'appropriation* d'une forme d'expression, d'un style, d'une langue et d'une façon d'aborder la critique du pouvoir en place.

Bien sûr, le cas de l'occupation allemande est extrême dès lors que le pouvoir est concentré dans les mains d'un groupe d'acteurs désireux de pratiquer une censure systématique. Mais il montre comment et pourquoi les journalistes et les éditeurs ne peuvent éviter de composer avec leurs intermédiaires, ou au moins d'agir en fonction de ce qu'ils savent de la manière dont ces intermédiaires agissent.

3.2 De la Libération à nos jours

3.2.1 Loi Bichet

Après la Libération, et à la suite du comportement collaborationniste de la maison Hachette durant l'occupation, la distribution se réorganisa d'abord de manière relativement anarchique. On créa les *Messageries Françaises de Presse* (MFP), qui jouirent d'un monopole de fait sur la distribution de la presse parisienne, tandis que certaines messageries régionales existaient en province (Eck, 2002). Les MFP, dont la gestion était assurée par des proches du Parti Communiste, utilisèrent le pouvoir qu'elles avaient en tant qu'intermédiaire pour saboter la diffusion de certains titres associés à une opinion adverse, notamment ceux de la mouvance du MRP (*ibid.*). Par ailleurs, il se trouve que « la Poste se préoccupe mal de la presse, qui lui semble un client imposé et quelque peu nécessaire. Pendant des décennies, les entreprises de presse demeurent tributaires du bon vouloir des revendications des postiers, qui n'assurent pas un service de qualité » (Eveno, 2002, p.357).

En 1947, la grève des rotativistes mit à mal les MFP, qui accumulèrent des dettes considérables. La loi Bichet fut votée cette même année, en avril, dans la confusion et la crainte qu'Hachette ne retrouve sa situation de monopole. Son premier article rappelle que « la diffusion de la presse imprimée est libre » (alinéa 1) et que « toute entreprise de presse est libre d'assurer elle-même la distribution de ses

propres journaux et publications périodiques par les moyens qu'elle jugera les plus convenables à cet effet » (alinéa 2). Par ailleurs, la loi Bichet a comme objectif d'assurer un traitement égalitaire pour les différentes entreprises de presse : « en France, aucune publication ne peut faire l'objet d'un traitement discriminatoire fondé sur le contenu, aussi bien de la part des coopératives que du réseau de diffusion » (Dupuy-Busson et Sonnac, 2002, p.393). Les MFP ayant fait faillite, les Nouvelles Messageries de la Presse Parisienne (NMPP) furent créées après l'adoption de la loi sous l'influence de Hachette, qui s'empessa de réclamer, via les NMPP, la levée des réquisitions dont elle faisait et pouvait encore faire l'objet. La maison obtint gain de cause en 1949 lorsque les NMPP firent l'offre de racheter une partie des sommes dues par les MFP (*ibid.*). Hachette redevenait ainsi une entreprise dont la prédominance était incontestable : malgré la liberté de diffusion, la plupart des entreprises de presse considéraient plus économique et moins dangereux stratégiquement la solution offerte par les NMPP. Ces dernières jouissaient d'un « monopole de fait ».

Par ailleurs, Hachette avait pris le contrôle de certains titres de presse, notamment *France-Soir* et *Paris-Presse*. A la fois éditeur et distributeur, il semble qu'elle ait connu le même genre de conflit d'intérêts que celui dont nous avons parlé à propos des portails web : Hachette était encouragée à privilégier la distribution de ses propres publications⁶⁵. En outre, des procédures furent entamées dès février 1949 accusant la maison Hachette d'avoir utilisé les fonds des NMPP pour financer sa prise de participation au capital des titres de presse : là encore, il y avait des frictions dès lors que l'entreprise était à la fois un distributeur, partenaire de nombreux éditeurs, et un éditeur concurrent de ces mêmes éditeurs.

Etant donné leur position dominante, les NMPP firent l'objet de nombreuses critiques. On leur a notamment reproché de se livrer à une inégalité de traitement des publications, même si cela n'a jamais été constaté ni sanctionné par les tribunaux (Derieux, 2002, p. 385). On leur a également reproché de s'appuyer sur leur position dominante sur le marché de la distribution pour fausser la concurrence sur le marché du mobilier d'agencement de diffuseurs de presse : en 1987, le Conseil de la concurrence a accusé les NMPP d'imposer aux candidats à l'obtention d'agrément

⁶⁵ Nous ne disons pas ici que ce fut effectivement le cas, mais qu'un conflit d'intérêts existait.

nécessaire à l'exploitation d'un point de vente de s'équiper en mobiliers fournis par elles. Elles pouvaient en effet exercer des pressions sur les diffuseurs en vue de les inciter à choisir son mobilier (*ibid.*). Nous verrons comment cela peut être rapproché du comportement de Google sur le marché de la publicité (chapitre 5).

Toujours en vigueur, la loi Bichet prévoit également que « les éditeurs restent propriétaires de leurs titres tout au long du processus de distribution et décisionnaires dans le volume alloué » (Dupuy-Busson et Sonnac, 2002, p. 408). Cela conduit certains éditeurs à surdoser la quantité de titres qu'ils font parvenir aux marchands de journaux, se livrant à une double stratégie de prolifération et de spécialisation (*ibid.*). Ils le font parce que le coût des invendus n'est pas trop élevé et parce qu'ils considèrent comme important d'être présents dans les rayonnages afin de minimiser la possibilité de voir des lecteurs se reporter sur des titres concurrents. Tous les journaux ne pratiquent pas cependant ce sur-approvisionnement, en témoigne l'exemple de *Libération* et de *La Tribune*, qui demandèrent en 2006 d'être retirés d'un grand nombre de points de vente, car « quand on vend 75 000 exemplaires sur toute la France, voire 10 000, il faut se représenter que, dans des centaines voire des milliers de communes ou de quartiers, on n'en vend qu'un seul. Et parfois aucun. On arrive donc à ce paradoxe : des journaux en quête désespérée d'acheteurs refusent d'être distribués sur tout le territoire » (Schwartzberg, 2007, p. 85). En outre, les éditeurs vendent des espaces dans leurs pages aux annonceurs publicitaires, ce qui les incite à se saisir d'un réseau de diffusion le plus large possible afin de montrer aux annonceurs que leur titre est visible dans de nombreux points de vente. « Le fait que le titre soit chez tous les marchands de journaux augmente sa probabilité d'être vendu et, par voie de conséquence, augmente la probabilité pour les lecteurs de prendre conscience de la publicité des produits pour lesquels les annonceurs font la promotion » (Dupuy-Busson et Sonnac, *op. cit.*, p.403).

La presse étant une composante fondamentale de la démocratie, son système de distribution, nous dit Nadine Toussaint Desmoulins, ne doit pas « souffrir d'exclusion, de sélectivité, ce qui est contraire aux pratiques commerciales courantes qui privilégient le référencement limité et la mise en valeur de certains produits au détriment d'autres » (Toussaint Desmoulins, 2002, p. 409). Cela est prévu par la loi

Bichet, dans le but de contrecarrer la « tendance naturelle des diffuseurs » qui serait de « refuser des titres qui se vendent mal ou qu'on ne connaît pas et de favoriser les titres connus qui se vendent bien et qui génèrent des recettes importantes. [...] Les diffuseurs, qui ne sont pas pénalisés par la mévente, n'ont donc pas le droit de refuser de diffuser un titre au motif qu'ils risquent des pertes » (*ibid.*, p.411).

Les NMPP, devenues *Prestaliss*, assurent aujourd'hui 70% de la distribution du secteur de la presse imprimée, quotidiens et magazines confondus⁶⁶. Leur système de distribution est toujours censé être équitable, basée sur la solidarité et l'entente entre les éditeurs, sous forme de coopérative, même s'il est l'objet de nombreuses critiques. Leur fonctionnement, et leur respect de la loi Bichet, notamment de l'égalité de traitement entre les différents titres, est contrôlé par l'Autorité de régulation de la distribution de la presse (ARDP), autorité administrative indépendante créée à cette fin le 20 juillet 2011 et travaillant de concert avec le Conseil supérieur des messageries de presse (CSMP), qui réunit des représentants des professionnels. L'ARDP veille notamment « au respect de la concurrence et des principes de liberté et d'impartialité de la distribution »⁶⁷.

3.2.2 Distribution et ligne éditoriale

S'intéressant en particulier à la distribution du quotidien *Le Monde*, Patrick Eveno a expliqué comment il était possible « de suivre les aléas de la politique rédactionnelle à travers la périodisation de la diffusion » (Eveno, 2002, p. 362). Dans une certaine mesure, ce sont les lecteurs qui donnent l'impulsion à la distribution : la demande crée l'offre. « Encore faut-il accompagner les lecteurs, afin d'adapter la maquette du journal aux goûts du public, afin de faire évoluer la pagination, l'écriture rédactionnelle, les styles rédactionnels en fonction des tendances de l'époque, de préférence en les précédant » (*op. cit.*, p. 363). La presse imprimée prend donc en considération ce qu'elle sait de son lectorat et de ses envies pour adapter son format, son style, les sujets qu'elle traite. Ainsi, les lecteurs, à partir de ce que l'on sait d'eux, contribuent à déterminer (sans être seuls déterminants) ce qu'est la ligne

⁶⁶ <http://www.lejdd.fr/Medias/Presse-ecrite/Actualite/Distribution-de-la-presse-un-risque-de-crise-systemique-488145> (dernière visite le 20 août 2013)

⁶⁷ Article 17 de la loi Bichet, comme aménagée le 5 juillet en 2011.

éditoriale du titre. Or nous verrons comment, et dans quelle mesure, cette pratique peut être rapprochée de ce que font certains titres de presse en ligne lorsqu'ils utilisent « Google Tendances de Recherche ».

La « Une » du *Monde*, explique Eveno, est considérée comme une « vitrine, qui doit attirer chaque jour de nouveaux acheteurs afin d'inciter des non-lecteurs à prendre en main le journal » (*ibid.*). C'est pourquoi une attention particulière est portée aux études de marché menées par la SOFRES sur la réception de la « Une ». Les mesures statistiques permettent de mesurer l'impact de tel ou tel sujet traité en « Une » sur les ventes. Au milieu des années 90, la « Une » du *Monde* présente un titre principal sur quatre colonnes, six appels de papiers, « dont trois illustrés de photos et un d'un dessin, un ventre quelque peu décalé et un pied de page où commence un éditorial ou point de vue. Après une période de tâtonnement, le titre principal est devenu plus informatif, tout en alternant les sujets informatifs et les sujets magazine » (Eveno, 2002, p.363). L'heure de publication compte également : le fait d'être un journal du soir, arrivé dans l'après-midi dans les kiosques, permet d'augmenter la durée d'exposition et donc, potentiellement, le nombre de ventes.

Il sera intéressant pour notre étude de garder cet exemple en tête, et de considérer l'usage fait par les éditeurs de ce qu'ils savent du comportement des utilisateurs de Google, puis d'analyser comment les caractéristiques du réseau de distribution-diffusion sont susceptibles d'influencer la présentation des nouvelles, les choix des sujets traités, et un tempo de publication censé maximiser l'exposition.

3.2.3 Exposition sur le web

Nous avons expliqué comment les propriétés du dispositif de distribution pouvaient avoir une influence sur les stratégies des éditeurs de presse imprimée. Sur le web, il existe également des statues de Pasquino : lieux de lectures où les informations provenant de divers éditeurs sont lues, comparées et commentées. Il existe également des intermédiaires, Google en tête, dont les profils sont très différents de ceux de la presse écrite. Quant à l'environnement légal, nous verrons qu'il est loin d'être stabilisé. Dès lors, la nature du pouvoir que les intermédiaires

peuvent exercer, et les effets de ce pouvoir, ainsi que la responsabilité de ces intermédiaires vis-à-vis du contenu diffusé, doivent être interrogés.

« Plus un journal est visible, plus il a de chances d'être acheté et plus cela coûte cher » (Schwartzberg, 2007, p. 83). Est-ce également vrai pour un site web ? Que sont les coûts, alors, que doivent supporter les éditeurs pour être visibles ? Comment peut-on maximiser l'exposition d'un titre sur Google ? Y a-t-il un impact sur le contenu lui-même, la ligne éditoriale, le traitement de l'actualité par les journalistes ? Y a-t-il besoin d'un équivalent de la Loi Bichet, qui assurerait, comme pour la presse imprimée, l'égalité et l'impartialité du traitement par Google ? Nous touchons ici au cœur de notre problématique.

Grâce à cette historicisation du dispositif, nous avons pu mettre en perspective et mieux comprendre notre problématique. Il nous reste maintenant à situer *L'Express* en particulier, étant donné l'attention que nous lui portons dans ce travail.

3.3 *L'Express*

3.3.1 Arbre généalogique

D'abord supplément politique accompagnant une fois par semaine le quotidien *Les Echos*, *L'Express* fut fondé en 1953 par Jean-Jacques Servan-Schreiber, ancien éditorialiste du *Monde* et de *Paris-Presse*, avec l'aide de Françoise Giroud, Pierre Viansson-Ponté et Leone Georges-Picot. Historiquement anti-communiste, anticolonialiste et soutien politique de Pierre Mendès France, le titre accueillait des auteurs aux opinions parfois contraires : Jean-Paul Sartre, François Mauriac, Albert Camus. Cependant, *L'Express*, dans ses débuts au moins, prenait des positions claires dans certains débats d'actualité. Et cela put influencer sa stratégie de production et de diffusion. Il devint par exemple un quotidien à la fin de l'année 1955, au moment des élections législatives, dans le but de soutenir activement le Front Républicain et la paix en Algérie, se positionnant comme l'organe de ce que Servan-Schreiber nommait lui-même la « Nouvelle gauche » (Siritzky et Roth, 1979). Après quoi, pour des raisons économiques, Servan-Schreiber décida de revenir à une version hebdomadaire et transforma en 1964 le titre en *newsmagazine* sur le modèle de ce qui existait déjà

en Angleterre, avec le *Time*, dont il imita le style, et en Allemagne, avec le *Der Spiegel*, dont il imita le format. *L'Express* fut ainsi le premier titre français de presse magazine d'information politique et générale.

Quand, en 1966, Claude Imbert devint rédacteur en chef, il afficha une ligne éditoriale qui se voulait neutre. Son arrivée et le changement de format correspondirent à un accroissement du nombre d'articles non signés, dont le but était de rendre compte des faits sans donner d'opinion à leur sujet. Le troisième changement important fut lié à l'orientation du modèle économique. La nouvelle formule était motivée par la convergence entre le lecteur « cadre » de *L'Express* et le lecteur « utile » des annonceurs, et par la volonté de la direction d'augmenter la part des recettes publicitaires par rapport à la part de la vente au numéro. De 40 % en 1962, Servan-Schreiber voulait que les recettes publicitaires passent à 75 % de la recette globale (Jamet, 1983, p. 89). Pour attirer une nouvelle clientèle, expliquait alors Roland Cayrol, *L'Express* avait posé deux règles : « d'une part, passer du prophétisme au professionnalisme [...], c'est-à-dire non plus présenter les réflexions personnelles de journalistes-moralistes destinées à un public averti, mais offrir chaque semaine à un large public la synthèse des informations importantes, quitte à les expliquer et les commenter – peut-être d'ailleurs amènera-t-on plus facilement de cette façon un public nouveau à des positions "progressistes" ; d'autre part renouveler les méthodes de conception et de fabrication du journal pour lui permettre d'assumer ce rôle » (Cayrol, 1966, p. 495). Cette modification de ligne éditoriale et le nouveau format étaient indissociables de la « course au tirage » et d'une « osmose voulue entre le marketing [...] et la rédaction » (Jamet, 1983, p. 90) destinée à augmenter les recettes du titre. Grâce à cette stratégie et à une campagne de publicité qui coûta un millions de francs, la diffusion passa de 152 919 exemplaires (décembre 1963 à novembre 1964) à 261 823 (mars 1965 à février 1966). Les journalistes voyant d'un mauvais œil le changement de format et de ligne éditoriale furent écartés. Servan-Schreiber remplaça ceux qui déclaraient « ne pouvoir s'accommoder de la transformation du journal d'opinion en champion de "l'information objective" et de la substitution de la *cover story* à l'éditorial, par des journalistes plus soucieux de métier que d'idéologie. » (*ibid.*, p. 90)

En 1971, moins d'un an après que le Groupe Express S.A fut devenu une société à directoire, son président Servan-Schreiber devint le représentant du Parti radical valoisien. Cette double-casquette provoqua le mécontentement de nombreux journalistes, et notamment de Claude Imbert qui démissionna pour fonder *Le Point*, ce qui marqua en France la fin du monopole de *L'Express* en matière de presse magazine hebdomadaire d'information politique et générale. Les années 70 furent également marquées par plusieurs aller-retour quant à la prise de position vis-à-vis des gouvernements successifs, et par le départ de nombreux journalistes. Finalement, Servan-Schreiber céda, en 1977, 45 % du capital au patron de la *Générale Occidentale* James Goldsmith. Ce dernier confia la direction à Jean-François Revel et la rédaction en chef à Olivier Todd. La présence dans la rédaction de Raymond Aron, éditorialiste et président du comité éditorial, qui comptait parmi les principaux opposants de Jean-Paul Sartre, donna au titre une orientation jugée à droite, qui fut amplifiée lorsque Goldsmith renvoya Olivier Todd en raison une couverture jugée défavorable au Président de la République Valéry Giscard d'Estaing, et que Jean-François Revel démissionna par solidarité⁶⁸.

En 1987, Goldsmith céda la propriété du groupe à la Compagnie générale d'électricité (CGE) qui, sous la houlette de la nouvelle PDG du groupe *L'Express* Françoise Sampermans racheta 40% des parts du magazine *Le Point* en 1992. Un an après ce rachat, Sampermans, alors qu'elle répondait aux questions du journaliste Jean-Pierre Elkabbach⁶⁹, expliqua que le journal était un produit vendu à des publicitaires et qu'elle évoquait par conséquent très régulièrement avec le directeur de la rédaction « l'évolution et de la stratégie rédactionnelle du journal ».

A partir du début des années 90, et avec l'arrivée de Christine Okrent au poste de rédactrice en chef en 1994, succédant à Yann de L'Écotais (1987-1994) lui même ayant succédé à René de Laportalière (1986-1987), *L'Express* se positionna à nouveau comme un magazine neutre au regard des politiques revendiquées par le Parti Socialiste et les partis de la droite modérée. Les années Okrent furent marquées à la fois par une progression des ventes mais aussi par la perte de la place du

⁶⁸ Dans une interview donnée à Bernard Pivot (Grand Soir 3, 7 février 1997), Jean-François Revel indiqua cependant que la couverture n'avait été qu'un prétexte pour Goldsmith, visant à se séparer de directeurs qui ne lui obéissaient pas.

⁶⁹ Interview disponible sur le site de L'INA : <http://www.ina.fr/video/I00009973> (dernière visite le 20 août 2013°

magazine le plus diffusé en France en 1996, devancé désormais par *Le Nouvel Observateur*⁷⁰. En 1996, la CGE céda le groupe à la *Compagnie européenne de publications* (CEP), filiale de *Havas* (futur *Vivendi Universal Publishing*). Et c'est parce que le PDG de la CEP, Christian Bréguou, voulait élargir la diffusion payée de *L'Express* à 800.000 exemplaires (contre un peu plus de 500.000 en 1996), en prenant en exemple le magazine allemand *Focus* et en changeant la ligne éditoriale, qu'Okrent fut remerciée⁷¹. Nous voyons ici très bien comment les volontés liées à une exposition maximale ont pu entraîner des changements dans l'organisation et le positionnement éditorial.

Denis Jeambar, ancien journaliste du *Point*, succéda à Christine Okrent. Ce fut sous sa direction que le premier site web fut mis en place en 1996. Il s'agissait essentiellement d'un site vitrine, où l'on présentait les produits du groupe sans qu'aucun article exclusif n'y soit mis en ligne. L'internaute pouvait y consulter les archives de *L'Express* depuis 1993 grâce à un système d'abonnement forfaitaire et pouvait s'abonner au magazine papier (Smyrniotis, 2005, p. 385). Il faudrait attendre 1999 pour qu'une version dynamique du site soit mise en place par Corinne Denis.

En 2002, les groupes *L'Express* et *L'Expansion* furent fusionnés sous la houlette de leur propriétaire *Vivendi Universal*, qui vendit le tout à la *Socpresse*, société éditrice du *Figaro*, propriété du groupe *Hersant*. La direction des sites internet des deux groupes fut alors confiée à Corinne Denis. En 2004, le groupe *Dassault* prit le contrôle de 80 % des parts de la *Socpresse* en rachetant les parts des héritiers de Robert Hersant. Puis, en décembre 2005, le groupe belge *Roularta*, numéro Un des médias en Belgique, racheta 35 % de *L'Express-L'Expansion* à la *Socpresse* de Serge Dassault, puis les 65 % restants six mois plus tard, devenant alors le seul propriétaire du groupe. *Roularta* créa l'entité *Roularta Media France* (RMF), rassemblant les magazines *L'Express*, *L'Expansion*, *L'Étudiant*, *L'Entreprise*, *Mieux vivre votre argent* et *Lire*. Après quoi, en 2006, Denis Jeambar quitta le groupe pour prendre la présidence des Editions du Seuil. Chef du service politique de *L'Express* depuis 1996 et directeur adjoint de la rédaction depuis 2001, Christophe Barbier prit sa place en

⁷⁰ <http://www.liberation.fr/medias/0101173933-ockrent-quitte-l-express-la-tete-haute-et-le-coeur-gros-son-depart-et-son-remplacement-a-la-tete-de-la-redaction-par-denis-jeambar-sont-officiels> (dernière visite le 20 août 2013)

⁷¹ *Ibid.*

août 2006. La Société des Journalistes de *L'Express* salua ce changement de direction et l'idée d'un renouvellement, en indiquant rester attentive à son « autonomie éditoriale »⁷².

3.3.2 Roularta Media France

Roularta Media France est une filiale de la multinationale belge Roularta Media Group (RMG) cotée à Bruxelles, ayant réalisé 712 millions d'euros de chiffres d'affaires en 2012, dont 75 % dans la presse écrite. Le résultat net était déficitaire cette même année à moins 3 millions d'euros. Roularta Media a deux mille employés, dont 700 en France qui réalisent 40 % du chiffre d'affaires de la société mère. Le PDG de RMG est Rick de Nolf, fils du fondateur. Deux directeurs généraux opérationnels sont en charge de RMF : la Directrice en charge de la partie business, Corinne Pitavy, et le Directeur en charge de la partie éditoriale, Christophe Barbier. Les titres du groupe sont répartis en trois pôles :

- Pôle Information-Business : *L'Express*, *L'Expansion*, *L'Expansion Tendances*, *L'Entreprise*, *Mieux Vivre Votre Argent*.
- Pôle Femmes : *L'Express Styles*, *Maisons Côté Sud*, *Maisons Côté Ouest*, *Maisons Côte Est*, *Vivre Côté Paris*, *Maison Française*, *Maison Magazine*, *Zeste Cuisinons Simple et Bon*, *Vivre Côté Cuisines et Bains*, *Point de Vue*, *Images du Monde*.
- Pôle Culture-Loisirs : *Studio Ciné Live*, *Lire*, *Classica*, *Pianiste*, *Start up*.

L'organisation de RMF est matricielle. Chacune des marques des trois pôles a sa propre rédaction, dirigée par un duo composé d'un directeur en charge de la partie commerciale des activités, et d'un rédacteur en chef, en charge de la partie éditoriale. Le modèle économique de chaque titre repose principalement sur la diffusion (vente au numéro, abonnements) et sur la publicité. En plus de ces marques, il existe des fonctions transverses, agissant pour chaque titre : la direction de la diffusion, les finances, la régie commerciale, la fabrication, la direction des nouveaux médias.

⁷² <http://tempsreel.nouvelobs.com/medias/20060809.OBS7840/jeambar-quittel-express.html> (dernière visite le 20 août 2013)

Il n'existe pas 21 sites pour les 21 titres, mais 6 sites web : Lexpress.fr, lexpansion.com, lentreprise.com, Votreargent.fr, L'Express Style et Côté maison. Chacun de ces sites a sa propre rédaction web, distincte de celle du papier⁷³. Les articles des autres titres sont ensuite ventilés, selon la thématique, parmi les 6 sites. Ainsi, tout ce qui concerne la décoration (*Maisons Côté Sud, Maisons Côté Ouest, Maisons Côte Est, Vivre Côté Paris, Maison Française, Maison Magazine*) sera intégré au site Cotemaison.fr, tandis que ce qui concerne la culture (*Studio Ciné Live, Lire*) sera intégré à Lexpress.fr, rubrique Culture.

La direction des nouveaux médias, fonction transverse dirigée par Corinne Denis, s'occupe principalement des sites et des applications pour tablettes et smartphones. Elle comprend une centaine d'employés, dont la moitié sont des journalistes répartis sur les six sites, la plupart d'entre eux travaillant pour Lexpress.fr. Le reste des employés de la direction est composé d'une vingtaine de techniciens, ainsi qu'une direction de projets fonctionnels, une division marketing, une direction de l'acquisition d'audience et une direction commerciale en charge des activités de régie publicitaire.

Le site Lexpress.fr est considéré, selon les propos de la directrice adjointe des nouveaux médias (bras droit de Corinne Denis) et Editrice de Lexpress.fr Sophie Gohier, comme le « vaisseau amiral » des nouveaux médias. La gestion du site est dévolue à Sophie Gohier pour la partie business, et à Eric Mettout, rédacteur en chef, rattaché à Christophe Barbier mais « très autonome dans sa gestion », selon ce que nous a dit Sophie Gohier lorsque nous l'avons interrogée. C'est donc essentiellement le duo Mettout / Gohier qui est en charge de la gestion du site, sous le contrôle de Corinne Denis et de Christophe Barbier.

RMF, sur son site officiel, indique toucher avec ses titres papier et ses sites web 29 % des CSP+ en France. Ces derniers constituent ainsi la principale cible publicitaire du groupe, et de *L'Express* en particulier, ce qui est dans la continuité du modèle déjà mis en place par Servan-Schreiber. Mais si la cible publicitaire est restée la même, en revanche, au moment où Barbier prit la tête du titre, l'objectif de Roularta était l'inverse de celui de Servan-Schreiber : augmenter la diffusion et les

⁷³ Nous ne mentionnons pas ici L'Etudiant, qui fait l'objet d'une gestion relativement autonome du reste du groupe. Ce titre est organisé en « news room », c'est-à-dire que les journalistes écrivent indifféremment pour le papier et le web.

ventes au numéro, pour augmenter les recettes totales tout en augmentant la part qu'y représente la vente au numéro, et enfin fidéliser le lectorat féminin. Barbier a également revu la maquette dès son arrivée à la rédaction en chef, et rompu la collaboration avec plusieurs chroniqueurs, dont Claude Allègre, Bernard Guetta, Daniel Rondeau et Jean-Luc Petitrenaud. Au total, une vingtaine de départs de journalistes ont eu lieu en 2006, dont la moitié de pigistes.

Pour ce qui est de la ligne éditoriale, selon Christophe Barbier, *L'Express* n'est aujourd'hui « ni de droite ni de gauche » et a choisi « le camp des lecteurs »⁷⁴. Il rejoint ainsi ce que François Mauriac disait déjà en son temps, à savoir que « ce que *L'Express* a de mieux, ce sont ses lecteurs » (Blandin, 2004, p. 60) Ainsi, le titre conserve un positionnement qui se veut le plus neutre possible et qui, dès lors, est susceptible de s'adresser à un lectorat le plus nombreux possible, ce qui va dans le sens de la stratégie commerciale. Par ailleurs, Christophe Barbier a fait, en mars 2013, l'objet de vives critiques de la part de la rédaction, l'accusant par voie de communiqué de ne pas suffisamment prendre en considération l'avis des journalistes dans la gestion de la ligne éditoriale, ce qui n'est pas sans rappeler les reproches autrefois formulés à l'encontre de Servan-Schreiber.

En termes de diffusion, avec une moyenne de 433 031 exemplaires par numéros en 2012, c'est d'après le classement de l'OJD le troisième magazine d'information après *Le Nouvel Observateur* (502 348) et *Paris Match* (610 156), et devant *Le Figaro Magazine* (431 865), *Le Parisien magazine* (415 302) et *Le Point* (412 278).

3.3.3 Du site compagnon au site d'information

Quand Eric Mettout arrive en 2000, la première version dynamique du site vient d'être lancée. Les journalistes avaient accueilli cette nouvelle version avec une certaine perplexité. Jusque là leurs articles n'étaient publiés que sur un seul support, et lorsqu'un support voulait les republier, par exemple un titre étranger, ils touchaient

⁷⁴ http://www.lexpress.fr/actualite/politique/presidentielle-l-express-a-choisi-le-camp-des-lecteurs_1096535.html (dernière visite le 22 août 2013)

des droits d'auteur supplémentaires. Or, cette fois, ils écrivaient pour deux supports et n'étaient pas mieux rémunérés⁷⁵.

Eric Mettout, toujours rédacteur en chef du site web, a travaillé comme journaliste au quotidien le *Matin de Paris*, au magazine *Actuel*, à la radio chez *Oui FM* et *Nova*, puis, à partir de 1996, chez l'éditeur de bandes-dessinées *Humanoïdes associés*. Sous son impulsion, tout au long des années 2000, le site s'est positionné non pas seulement comme un concurrent des sites adossés aux magazine, mais également des sites adossés à des quotidiens : *Lemonde.fr*, *Lefigrao.fr*, *Liberation.fr*, *Leparisien.fr* ; et des sites nés en ligne : *Rue89*, *Mediapart*.

Le positionnement magazine a été traduit sur le site, étant donné les caractéristiques propres à ce média. Il semblait en effet inapproprié à Eric Mettout de traiter l'actualité de la même manière que sur le magazine papier. Dès lors, le site affiche plusieurs niveaux de traitement de l'actualité : des articles factuels, publiés très peu de temps après les événements, et des articles d'analyse, sur le modèle du magazine, repris à la rédaction papier ou bien rédigés par la rédaction web.

En 2005, avant l'arrivée de Christophe Barbier à la tête de *L'Express*, le site n'avait qu'un rédacteur en chef, chapeautant un journaliste et deux stagiaires. « Leur travail consist[ait] à produire des contenus d'information sur une base quotidienne, essentiellement en réécriture des dépêches d'agences. Cependant, en raison de cet effectif réduit, la production originale [était] très limitée. [...] L'équipe de *lexpress.fr* [était] complétée par un documentaliste chargé d'intégrer les articles du titre au site internet sans traitement supplémentaire. La stratégie adoptée par la direction des éditions électroniques en ce qui concerne le site de *L'Express* [visait] à faire face à deux difficultés majeures. D'une part la nécessité de faire fonctionner le site à un moindre coût, afin de minimiser les pertes qu'il [engendrait], et, d'autre part, celle de se différencier de la concurrence des quotidiens en ligne. C'est la raison pour laquelle l'accent [était] davantage mis sur des thématiques magazine et plus particulièrement

⁷⁵ « La toute première réaction, ça a été une sorte de perplexité quand on a vu que, à ce qui était l'unique support, à savoir l'hebdomadaire papier qui s'appelait *L'Express*, s'est ajouté soudain un site web, qui certes s'appelait *Lexpress.fr*, mais qui était malgré tout un autre support, et qui reproduisait systématiquement nos papiers. Ma première réaction a été de penser que d'une part, soudain, nous écrivions pour deux supports, presque malgré nous, car cela s'était fait sans qu'on demande notre avis. Nous avons craint que ça affaiblisse l'hebdo, car nos lecteurs avaient un accès gratuit à une partie du contenu. Au lieu de payer trois euros cinquante comme tout le monde pour avoir un accès à ce contenu, il y avait un support, *Lexpress.fr*, qui donnait un accès gratuit. » Marc Epstein – Rédacteur en chef du service Monde de *L'Express* (Notre entretien)

les sujets de société plutôt que sur une couverture exhaustive de l'actualité quotidienne. Néanmoins, des efforts [furent] consentis dans cette direction avec, par exemple, l'incitation faite aux journalistes de l'hebdomadaire de contribuer à une production des contenus multimédia en direction du site internet. Ceci en effectuant de prises des vue photo ou vidéo lors de leurs reportages quotidiens. Globalement, en ce qui concerne *lexpress.fr*, il y a [eu] un glissement vers un fonctionnement sur une base quotidienne, puisque la formule hebdomadaire ne [semblait] pas être adaptée à la concurrence sur l'internet. » (Smyrnaio, 2005, p. 387).

Le rachat par Roularta et l'arrivée de Christophe Barbier à la tête de *L'Express* ont marqué une accélération de ce glissement, ainsi qu'une tentative de rapprochement de la rédaction web de celle du papier. Il s'agissait de réunir les deux rédactions autour d'une même identité, et ce malgré les différences dans leur temporalité de publication. La stratégie consistait, pour parler dans les termes de l'ANT, d'*intéresser* la rédaction de l'hebdomadaire au fonctionnement du site web, et de les faire participer à son élaboration, afin que les journalistes *s'enrôlent* et se *mobilisent* dans un même réseau stabilisé autour de la marque *L'Express*⁷⁶.

Christophe Barbier décida que la conférence hebdomadaire de la rédaction papier aurait lieu à l'étage de la rédaction web. Ainsi, les journalistes de la version papier, qui sont au quatrième étage, descendent au troisième étage une fois par semaine et se réunissent au milieu des journalistes de la rédaction web. Mais ce rapprochement est perçu par les journalistes de la rédaction web comme factice. Ceux-là continuent en effet de travailler pendant qu'au milieu d'eux se tient la conférence. Quant aux deux journalistes de la version papier que nous avons interrogés, ils nous ont dit regretter que les journalistes du web ne prennent pas assez la parole lors de ces conférences.

En 2012, le « .fr » est abandonné dans le nom du site : « *lexpress.fr* » devient « *L'Express* ». Le but est d'afficher la continuité entre le papier et le web, toujours dans cette même démarche de fortification du réseau autour d'une identité commune.

⁷⁶ « C'est l'arrivée de Christophe Barbier à la tête du journal qui a vraiment rapproché le site web de l'hebdo. Auparavant, il existait un site web. On constatait sur nos écrans d'ordinateur qu'il reprenait nos papiers, mais on n'avait pas de contact avec l'équipe web. Christophe Barbier a institué des conférences où il était question à la fois de l'hebdo et du web, et on a connu une époque où en réalité nous nous retrouvions à parler davantage de nos propositions pour le site web que nous n'évoquions le contenu de l'hebdo. Ça nous a amenés à regarder le site de beaucoup plus près, parce qu'on était associés au processus. » Marc Epstein – Rédacteur en chef du service Monde de L'Express (Notre entretien)

Cependant, les journalistes de la version papier nous ont dit que la ligne éditoriale du site web n'était pas selon eux la même que celle de la version papier. Selon l'un d'entre eux en particulier, certains sujets sont traités par la rédaction web dans le seul but de capter du trafic, alors qu'ils ne correspondent pas à la ligne éditoriale⁷⁷. Il existe ainsi une mise en tension entre les lignes éditoriales du papier et du web qui peuvent diverger étant donné des modes de diffusion et des stratégies commerciales différentes.

Par ailleurs, en 2012, alors même que nous faisons nos entretiens, un nouveau format a été inauguré, appelé « Express Yourself », qui consiste à publier sur le site web des articles ou des chroniques écrits par des experts ou des amateurs, extérieurs au site. Alors qu'il présentait ce nouveau format, Eric Mettout citait en exemple le *Huffington Post* et le *Post*, tous deux connus pour faire écrire des amateurs, tout en disant bien que *L'express.fr* n'avait pas pour vocation de changer sa ligne éditoriale, tournée vers la politique, l'actualité internationale et les sujets de société. L'idée, selon lui, consisterait à appliquer les recettes du *Huffington Post* et du *Post* « notamment techniques mais aussi éditoriales à un matériau différent »⁷⁸. Dans la même interview, Christophe Barbier disait quant à lui qu'il y avait « un vrai rapport de force entre des prestataires qui fournissent des tuyaux et des éditeurs qui fournissent des contenus, il y a une richesse à se partager. Il faut que chacun soit juste. Dans les débuts d'internet, on n'a peut être pas assez été vigilant. Il faut qu'on discute de toute cela. A mon avis, *L'Express* ne peut être que dans le cocktail gagnant parce que qu'est-ce qu'une offre de médias d'information complète serait sans *L'Express* ? ». Le message envoyé ici à Google et aux autres infomédiaires est clair : ils ont besoin de *L'Express* autant que *L'Express* a besoin d'eux et, dès lors, il sera nécessaire de trouver des terrains d'entente.

⁷⁷ « C'est sûr que dans la rédaction web, ils traitent certains sujets qu'on ne traite pas. Cela est dû au fait qu'ils ne sont pas limités, comme nous, par le nombre de pages. Mais cela est aussi dû au fait qu'Eric Mettout vise une audience très large pour le site, à mon avis beaucoup plus large que celle du papier, et que par conséquent il est prêt à traiter des sujets qu'on ne traitera pas sur le papier, simplement pour récupérer une audience sur Google ou sur Yahoo qui n'est pas fidélisée, et qui ne correspond pas forcément au cœur de notre cible, mais qui fera de la page vue et qui rapportera de l'argent. Du coup, c'est difficile de voir clairement quelle est la ligne éditoriale du site. En tout cas, c'est sûr, ce n'est pas exactement la même que celle du papier. » Julien Bordier – Rédacteur en chef adjoint au service Arts & Spectacles (Notre entretien)

⁷⁸ http://www.dailymotion.com/video/zo7eva_eric-mettout-et-christophe-barbier-l-express_tech?start=4#.UVxG7aXlQs

3.3.4 Chiffres et perspectives

En 2012, le site Lexpress.fr⁷⁹ comptait une vingtaine de journalistes, une trentaine avec ceux de la rédaction web de *L'Expansion*. Son audience était de 7 711 000 visiteurs uniques par mois en janvier 2013 selon les chiffres Médiamétrie. Selon le classement de l'OJD, c'était en nombre de visites total en février 2013, parmi les sites d'information politique et générale adossés à un titre papier, le 8^{ème} (16 530 119), derrière 20minutes.fr (18 988 665), Liberation.fr (20 235 481), Lepoint.fr (20 470 405), Nouvelobs.com (27 742 146), Leparisien.fr (31 712 100), Lefigaro.fr (49 584 072) et Lemonde.fr (60 183 505). Son ratio pages vues / nombre de visites est quant à lui de 3,54 (soit plus de 58 millions de pages vues).

La structure d'audience nous a été décrite lors de nos entretiens⁸⁰ comme étant la suivante : 25% provenant de Google Actualités, 25% provenant de Google Search, 20% d'accès directs, 15% de sites influents, 7% de partenariats avec Yahoo et Orange, environ 2-3% *Facebook* et moins de 1% *Twitter*. La part d'audience apportée par Google est donc de 50% et nous a été présentée par l'Editrice Sophie Gohier comme « l'enjeu n°1 » pour le modèle économique de Lexpress.fr.

Le fondateur de l'Express Jean-Jacques Servan-Schreiber disait déjà en son temps qu'« un journal ne saurait demeurer longtemps libre et prospère s'il ne respire pas sur le même rythme que les entreprises industrielles de son temps » (Cayrol, 1966, p. 495). Or, quel meilleur exemple que Google peut désigner une entreprise de notre temps, dont l'action pourrait contribuer à faire de *L'Express* un leader d'opinion ? Et dans ce cas, que veut dire au juste « respirer sur le même rythme » ? Cela conduit-il vraiment, comme le disait Servan-Schreiber, à « demeurer libre et prospère » ? Quel genre de liberté alors ? Quelle prospérité ? Enfin, quel journalisme ? Et quelle traduction en ligne de l'identité du titre papier ?

⁷⁹ Même si le site s'appelle désormais « L'Express », nous continuerons à le nommer Lexpress.fr, dans le but de bien le différencier du titre papier dans notre compte-rendu.

⁸⁰ Etant donné le caractère stratégique de cette information, nous n'avons pu obtenir de chiffres exacts. Cependant, plusieurs des acteurs interrogés nous ont présenté les chiffres comme étant « à peu près, selon les mois » ceux que nous présentons ici.

CONCLUSION PARTIE 1

Dans cette première partie, nous avons présenté de quelle manière notre objet de recherche avait été construit et situé, puis nous avons mis notre problématique en perspective grâce à une approche sociohistorique de la recherche d'information et de la distribution-diffusion de la production des journalistes. Après avoir passé au tamis des SIC, de l'ANT et du concept de dispositif les questions motivant ce projet, puis présenté nos terrains d'investigation et nos méthodes, nous avons rendu compte de la succession des configurations ayant abouti à celle que nous étudions.

Nous avons vu que dès lors que celui qui émet une information souhaite que celle-ci soit la mieux communiquée possible, il lui faut penser à la manière dont on pourra y accéder. Pour cela, il doit notamment penser aux intermédiaires auprès desquels ceux à qui il s'adresse viendront s'informer. Il peut ainsi choisir un certain support, un certain style d'écriture, un certain format, et assortir sa production de méta-informations dans le but que son document soit repéré par les intermédiaires en question et mis en valeur.

Dans un contexte où le moteur algorithmique est une des principales voies d'accès à l'information sur le web, un éditeur désireux de maximiser son trafic pourra essayer d'apprendre comment fonctionnent les crawlers, l'indexeur et l'algorithme d'un moteur. Cela s'avère particulièrement difficile étant donné l'opacité des mécanismes qui prévalent au fonctionnement du dispositif. Google, en effet, ne divulgue pas dans le détail de quelle manière (par quelles *médiations*) l'appariement entre une requête et un résultat est effectué. En conséquence, l'éditeur peut faire confiance à Google et se conformer à ce qui lui est recommandé, ou bien effectuer certains tests quant aux sujets traités, aux mots choisis, aux formats, à la structure de son site, aux méta-information et au maillage hypertexte. En effet, même si le fonctionnement du moteur est opaque, l'éditeur peut tout de même constater ce qu'il se passe à l'issue de l'appariement, essayer d'identifier les symboles d'entrée retenus et tenter de maximiser l'émission de ces symboles.

Partie 2 – Configuration négociée et négociable : déploiement du dispositif

« Faire un lien vers un autre site, ne pas faire un lien, ou supprimer un lien peut être respectivement vu comme un acte d'association, de non-association ou de dissociation. »

(Rogers, 2010)

Dans cette partie, nous sonderons de quoi sont faites les relations entre Google et les éditeurs : de quels possibles, quelles opportunités, quelles négociations, quelles ententes la liste générée par Google est-elle le résultat ? Notre but est ainsi de rendre compte de ce qui se trouve *derrière* l'appellation « référencement naturel » utilisée par Google pour désigner le fonctionnement de ses services d'indexation gratuit pour les éditeurs. Car en effet, comme le soulignent Joëlle Farchy et Cécile Méadel, une telle appellation « comme souvent avec les technologies numériques, relègue dans l'arrière-boutique nébuleuse de la machine tout le travail intermédiaire de production des résultats, ici de repérage des contenus, en mettant en avant une assomption de neutralité » (Farchy et Méadel, 2013). Nous procéderons d'une part à une analyse concrète des moyens de communication entre les documents produits par les éditeurs et les dispositifs Google Search et Google Actualités (chapitre 4). Nous pourrons ainsi comprendre de quels « pouvoir faire » et « pouvoir de faire » les éditeurs peuvent jouir théoriquement, et le contrôle qu'une telle capacité d'action peut leur permettre d'exercer sur la publicisation de leurs contenus. Nous verrons également de quelle manière Google tente de « conduire la conduite » des éditeurs et se réserve le droit de les récompenser ou de les sanctionner. Dans un deuxième temps, nous nous intéresserons à la valeur économique des liens hypertextes générés par les moteurs, ainsi qu'aux incitations qu'elle induit (chapitre 5). Nous verrons que le dispositif crée des opportunités qui peuvent inciter Google ou les éditeurs à agir d'une certaine manière. Enfin, nous nous intéresserons au flou juridique qui nimbe l'activité du dispositif, et aux négociations qui se tiennent entre Google et les éditeurs (chapitre 6).

Le but de cette partie sera ainsi de déployer le dispositif : quel fonctionnement, quel champs de manœuvres, quels systèmes de production et d'appropriation, quelles incitations, quelles intentions, quel type de contractualisation, quelles frictions, quels partenariats ? Une fois que nous aurons apporté des éléments de réponse à chacune de ces questions et articulé ces réponses les unes avec les autres, nous serons à même d'analyser dans la partie suivante la manière dont les éditeurs se saisissent de leurs possibilités d'action, la façon dont les incitations et les négociations peuvent peser sur la nature et le résultat du processus d'infomédiation, et l'influence d'une telle situation sur l'organisation du travail, l'énonciation de l'actualité et le quotidien des journalistes.

CHAPITRE 4. LE MOTEUR COMME CHAMP DE MANŒUVRES

« Algorithm = Logic + Control »

(Kowalski, 1979, p. 424)

Ce chapitre a pour but de décrire et de comprendre de quoi est constitué le processus ayant lieu lorsque le contenu produit par les éditeurs est rendu accessible *via* Google Search et/ou Google Actualités, étant donné ce que l'on nomme le « référencement naturel ». Nous sondons la couche architextuelle⁸¹ qui ne modifie « pas seulement l'image du texte elle-même, mais les conditions d'accès à cette image » (Jeanneret et Souchier, 2005, p. 12), et donc dans notre cas les conditions d'accès à l'information d'actualité. Nous pourrions ainsi comprendre dans quel champ de manœuvres évoluent les éditeurs. Pour cela, nous décryptons le programme d'action attaché à l'ensemble du dispositif, puis à Google Actualités en particulier — programme d'action « qui se veut prédétermination des mises en scène que les [éditeurs⁸²] sont appelés à imaginer à partir du dispositif technique et des prescriptions (notices, contrats, conseils) » (Akrich, 1987, p. 208). Nous pourrions ainsi déterminer quels sont les degrés d'enrôlement possibles pour les éditeurs, c'est-à-dire les éléments techniques qui peuvent localiser, spécifier ou renforcer l'association de leur action à l'action de Google. Cela sera encore incomplet, car « tant qu'il ne se présente pas d'acteurs pour incarner les rôles prévus par le concepteur (ou en inventer d'autres), son projet reste à l'état de chimère : seule la confrontation réalise ou ir-réalise l'objet technique. » (*ibid.*). A la fin de ce chapitre, nous aurons donc des éléments de réponse à la seule question du pouvoir que Google souhaite exercer et peut exercer, et non du pouvoir que la firme exerce *effectivement*. De la même manière, nous aurons obtenu des éléments quant aux différentes manières dont les éditeurs peuvent agir, et non pas au sujet de la manière dont ils agissent *effectivement*. Il s'agit en somme de décrire ce que Google voudrait que les éditeurs

⁸¹ « Nous nommons *architextes* (de arché, origine et commandement), les outils qui permettent l'existence de l'écrit à l'écran et qui, non contents de représenter la structure du texte, en commandent l'exécution et la réalisation. Autrement dit, le texte naît de l'architexte qui en balise l'écriture. » (Jeanneret et Souchier 1999, p. 103)

⁸² Etant donné notre objet, ceux qu'Akrich désigne ici comme des « utilisateurs » représentent les éditeurs de presse en ligne, en cela qu'ils « utilisent » le dispositif pour rendre visibles leurs contenus. C'est pourquoi nous avons remplacé le mot « utilisateur » par « éditeur ».

fassent et ce que les éditeurs peuvent faire pour maximiser leur trafic, plutôt que de décrire ce que les éditeurs font.

Bernhard Rieder (2005) a tracé une morphologie du moteur de recherche autour des quatre questions auxquelles doit répondre son concepteur : crawling des données (scope, structure, pertinence, extraction), indexation (critères, structure, mise à jour), appariement (requête, critères de pertinence, hiérarchisation des résultats) et design (interface). « Considérées ensemble, ces quatre dimensions renseignent à propos de la morphologie d'un moteur de recherche – les séries de choix effectués par le développeur aussi bien que les séries de questions qui se posent pour celui qui utilise le moteur [et les éditeurs qui y sont référencés]. Chaque moteur de recherche fournit [...] une réponse particulière à chacune des quatre questions. Le succès commercial autant que l'impact politique en dépendent »⁸³ (Rieder, 2005, p. 28). Nous nous servons de ces quatre dimensions comme d'un plan pour déployer le champ de manœuvres des éditeurs.

4.1 Crawling

« Chercher » est issu du latin *circare*, qui signifie « faire le tour de, parcourir pour examiner », et induit une notion de déplacement : pour trouver, il faut avoir parcouru. C'est ce que fait un bibliothécaire lorsqu'il consulte les sommaires et les tranches des livres, puis leur contenu, afin de les ranger de manière à pouvoir les mobiliser. Et c'est ce que font sur le web les crawlers de Google.

4.1.1 Fonctionnement, fonction, enjeux

Les premiers concepteurs d'annuaires naviguaient eux-mêmes de page en page dans le but d'identifier le contenu qui intéresserait leurs publics. Ils pouvaient de cette façon manquer certaines informations. Leur action était limitée dans le temps, les employés étant dans l'incapacité de consulter le contenu de chacune des pages, dont le nombre augmentait de façon exponentielle. La chance jouait un rôle non

⁸³ « Taken together, these four layers trace the morphology of a search engine – a series of choices for the developer as well as a series of questions for the investigator. Every search engine gives a particular answer on each one of those levels. Commercial success as well as political impact depends on it. » (Rieder, 2005, p.28, notre traduction).

négligeable. Il s'agissait de « surfer » afin de découvrir les pages présentant un intérêt pour finalement produire une *inscription* hypertexte permettant de retrouver les pages en question. C'est en cherchant à optimiser cette activité qu'on l'a automatisée. Des logiciels « user-agents » furent mis au point, appelés « robots scanneurs », « robots d'indexation », « crawleurs » ou « spiders », conçus afin de visiter le web à la place de l'être humain.

Les crawleurs peuvent fonctionner de différentes façons. Ils peuvent par exemple copier la totalité du contenu d'une page ou bien copier le contenu qui leur semble le plus pertinent, comme les adresses URL ou les adresses courriel présentes sur la page en question. Dans tous les cas, ils se déplacent grâce aux liens hypertextes reliant les pages entre elles. Une page qui n'est pas accessible par des hyperliens sera donc hors de leur portée. Tout comme les êtres humains, ces robots ne peuvent deviner l'adresse des pages qu'ils doivent visiter. Ils ont besoin de moyens de transport (les liens) et d'indications quant à leur destination (les ancres).

Le crawler doit être en mesure, en partant d'une page prédéfinie par son programmeur⁸⁴, de visiter les pages dont le contenu est de la meilleure qualité possible (Najork et Wiener, 2001), de visiter les pages dont le contenu est spécialisé (Chakrabarti *et al.*, 1999) et le format particulier (Diligenti *et al.*, 2000), puis de revisiter les pages régulièrement (Cho et Garcia-Molina, 2000) et, enfin, de rendre compte de l'avancement de son travail (Talim *et al.*, 2001).

Vladislav Shkapenyuk et Torsten Suel (2002, p.6) énumèrent les éléments essentiels à la création d'un crawler efficace : la flexibilité, l'optimisation coût/performance, la solidité, le respect des standards, la perfectibilité et la possibilité de reconfigurer les paramètres. Edwards *et al.* précisent quant à eux que le concepteur d'un crawler se pose les questions suivantes : Quelle doit être la fréquence des visites du robot sur une page ? Dans quel ordre les pages doivent-elles être visitées ? La stratégie doit-elle être basée sur l'importance des pages ou bien sur la fréquence de leur actualisation ? (Edwards *et al.*, 2001)

⁸⁴ Google utilise l'annuaire communautaire et ouvert *Open Directory Project* pour définir le point de départ de ses robots-crawlers.

Chaque moteur déploie plusieurs crawlers qui parcourent la toile simultanément, et visitent et revisitent un document donné à des fréquences variées. Ils laissent des traces numériques⁸⁵ à chacun de leur passage, dont les éditeurs peuvent prendre connaissance grâce aux fichiers appelés « logs »⁸⁶.

4.1.2 Googlebots

Le crawler principal de Google s'appelle « Googlebot ». Les Googlebots peuvent explorer la plupart des types de page et de fichier⁸⁷. Toutefois, Google explique sur son Centre d'Aide aux Webmasters (CAW) être davantage « axé sur du contenu textuel » et demande aux éditeurs de fournir un équivalent en format texte de leurs contenus non-textuels (contenus *Rich Media* : *Flash*, *Silverlight*, fichiers vidéo). Il est précisé que « Google n'accepte aucun paiement pour explorer un site plus fréquemment ». Quant aux paramètres exacts des crawlers, la firme ne les publie pas, expliquant simplement :

« Googlebot est conçu pour être réparti sur plusieurs machines pour améliorer les performances et évoluer avec la croissance du Web. En outre, pour limiter l'utilisation de la bande passante, nous exécutons de nombreux robots d'exploration sur des machines situées à proximité des sites à indexer sur le réseau. Par conséquent, vos fichiers journaux peuvent indiquer des visites de plusieurs machines sur google.com, toutes avec le user-agent Googlebot. Notre objectif est d'explorer autant de pages de votre site que possible à chaque visite, sans surcharger la bande passante de votre serveur. » (CAW)

Comme le souligne Bernhard Rieder, « nous ne connaissons pas les limitations pratiques des crawlers de Google, mais les analyses empiriques et les spéculations dans la communauté des chercheurs pointent une étroite relation entre le PageRank d'une page, i.e. son degré de centralité topographique, et l'importance que Google lui accordera dans sa politique d'exploration. Une page considérée comme centrale est

⁸⁵ Lorsque Brin et Page ont activé les robots-crawlers de Google, ils racontent que certains webmasters n'étant pas familiers avec l'outil ont vu la trace laissée par le robot dans leurs fichiers logs, et ont cru que Brin et Page avaient visité eux-mêmes la totalité des pages du site. Certains d'entre eux ont même contacté Brin et Page pour leur demander ce qu'ils avaient pensé de leur site. Il y avait confusion alors entre le robot-crawler et l'être humain, confusion entre la machine et son concepteur (Brin et Page, 1998).

⁸⁶ Les fichiers logs, ou « fichiers journaux » enregistrent de manière automatique, et par ordre chronologique, les actions exécutées par un serveur ou une application informatique.

⁸⁷ Pour une liste complète des types de fichiers indexés : <http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=35287> (dernière visite le 21 août 2013)

traitée de manière plus complète, les crawleurs la pénètrent plus profondément et les mises à jour sont plus fréquentes. Le terrain de recherche est donc délimité de façon dynamique, en prenant en compte la structure de ses contenus. » (Rieder, 2006, p.120).

La fréquence de crawling de Googlebot varie selon les éditeurs, selon les documents et selon leur emplacement dans l'arborescence du site (profondeur) et la topologie du web (centralité). Dans le cas d'un éditeur de presse en ligne, la fréquence peut être de plusieurs passages par minute. Cela implique pour l'éditeur d'avoir des serveurs capables de supporter une telle fréquence. En effet, « lorsqu'un crawleur visite une page web, il consomme des ressources [...] appartenant au propriétaire de la page en question »⁸⁸ (Langville et Meyer, 2006, p.16-17).

Google permet aux éditeurs d'avoir en partie la main sur le taux de renouvellement du passage du crawleur, en forçant une augmentation ou une diminution de la vitesse d'exploration. L'augmentation est fortement déconseillée par la firme qui ne recommande d'utiliser cette fonctionnalité que pour diminuer la vitesse de crawling en cas de surcharge des serveurs, et non pas pour l'accélérer, même si cela est techniquement possible.

En plus du Googlebot, il existe d'autres types de crawleurs employés par Google, chacun permettant de procéder à une association en particulier:

- AdsBot scanne les pages des annonceurs du réseau publicitaire de Google et calcule leur « quality score » et leur « degré de convivialité ». (Association annonceur-Google.)
- Feedfetcher scanne les sites désignés par un utilisateur paramétré les lecteurs de flux RSS de sa page d'accueil iGoogle ou de Google Reader (Association Google-internaute.)
- Imagebot scanne les méta-informations qui accompagnent les images : titre, légende, date, thématique du site. (Association Google-éditeur.)
- Googlebot-Mobile traduit le contenu des fichiers HTML en un code lisible par les mobiles. (Association Google-éditeur.)
- Mediapartners, ou Mediabot, crawle les sites du réseau AdSense dans le but de déterminer quelles sont les publicités qu'il est préférable d'afficher sur le module. (Association Google-éditeur.)

⁸⁸ « when a spider visits a webpage, it consumes resources [...] belonging to the page's host » (Langville et Meyer, 2006, pp.16-17, notre traduction)

Chaque crawler contribue à rendre possible un type d'association en particulier. Les éditeurs de presse en ligne peuvent potentiellement être concernés par chacun d'entre eux, selon la nature et le degré d'implication dans le dispositif. Le passage de certains crawlers, contrairement à Googlebot, est conditionné. Par exemples, seul un annonceur partenaire du service AdWords recevra sur ses pages la visite d'AdsBot, et seul un éditeur partenaire du réseau de syndication AdSense recevra Mediabot.

4.1.3 Affaiblir, rompre ou localiser

Les éditeurs peuvent échapper à Googlebot d'une manière très simple, grâce au protocole d'exclusion initié par Martijn Koster en 1994. En implémentant la ressource textuelle « robots.txt », un webmaster peut signaler qu'il ne désire pas que son site, ou qu'une partie de son site, soit scanné⁸⁹. Ce protocole est un accord que les principaux acteurs du web, dont Google, se sont engagés à respecter, sans qu'il fasse pour autant l'objet d'une quelconque obligation légale. Par convention, le robots.txt, implémenté à la racine d'un site, est le premier élément visité par le crawler. Plusieurs informations peuvent s'y trouver, chacune donnant une indication à propos de ce que le webmaster autorise à faire de son contenu⁹⁰ :

- Tout indexer⁹¹ :
*User-agent: **
Disallow:
- Ne rien indexer :
*User-agent: **
Disallow: /
- Ne rien indexer sur une page, à part un seul fichier :
*User-agent: **
Disallow: /page/
Allow: /page/nom_du_fichier.html

⁸⁹ Il existe d'autres moyens de désindexer le contenu, comme le fichier « .htaccess », mais le plus répandu demeure le protocole « robots.txt » sur lequel nous nous concentrerons en particulier ici. Dans le chapitre 6, nous reviendrons sur les débats dont le robots.txt fait l'objet.

⁹⁰ Nous ne citons ici que les principales opérations possibles, mais il en existe de d'autres.

⁹¹ Si, à chaque fois que nous avons mis le symbole « * », nous spécifions « Googlebot », alors la seule association concernée sera celle de Google, les autres robots-crawlers étant autorisés à « faire ce qu'ils veulent ». En revanche, avec « * », c'est à l'ensemble des robots-crawlers que l'éditeur s'adresse.

- Ne pas indexer les pages du site contenant dans l'adresse URL « monde » et « societe » correspondant aux rubriques « Monde » et « Société » :

*User-agent: **
Disallow: /monde
Disallow: /societe

Pour l'éditeur qui ne voudrait pas implémenter le « robots.txt » à la racine de son site, mais simplement donner des indications à propos d'une page en particulier, il existe des balises « méta-tags » à implémenter dans le code source de la page. C'est le cas si l'on veut :

- Que la page ne soit pas indexée et que le crawler ne puisse pas suivre les liens qui y figurent :

`<meta name="robots" content="noindex, nofollow">`
ou `<meta name="robots" content="none">`

- Que la page ne soit pas indexée et que le crawler puisse suivre les liens qui y figurent :

`<meta name="robots" content="noindex, follow">`

C'est à l'éditeur de se prémunir des crawlers : agir pour choisir s'il veut rompre, interrompre, atténuer ou spécifier l'association sur l'ensemble de son site ou sur une page en particulier. S'il n'agit pas, l'association pourra être établie par défaut. Googlebot ne demande pas de permission : seule une interdiction peut lui être opposée. L'action représente un coût supplémentaire, lié au temps de travail, aussi court soit-il, et à la compétence, aussi facile à acquérir soit-elle, qu'il faut à l'éditeur pour prémunir son site.

Nous comprenons dès lors pourquoi l'observateur Sullivan conseille au géant des médias Rupert Murdoch, lorsqu'il se plaint de la vampirisation par Google de ses contenus, de mettre un « préservatif robots.txt » pour protéger ses sites (Sullivan, 2009a). Si Murdoch n'implémente pas ce protocole, c'est, selon Sullivan, parce qu'il est trop *intéressé* par l'apport d'audience potentiel, c'est-à-dire qu'il *s'engage* par son absence d'action.

4.1.4 Fortifier ou spécifier

Les éditeurs peuvent tenter de lisser leur communication avec les crawlers en leur destinant certaines métadonnées spécifiques. C'est ce qu'on appelle le « White

Hat SEO » : l'optimisation dans le but de faciliter le fonctionnement du dispositif comme il a été prévu par ses concepteurs, et, ainsi, pour l'éditeur, d'agir dans l'intérêt de Google (dont la satisfaction est immédiate) en même temps que dans le sien propre (dont la satisfaction est potentielle). De la bonne entente du contenant et du crawler dépendront en partie l'indexation du contenu et l'importance du trafic capté par l'éditeur.

Les principales métadonnées sont des balises « méta-tags ». Googlebot ne les prend pas toutes en considération mais donne⁹² des indications à propos de celles qui revêtent à ses yeux un intérêt particulier (CAW) :

- `<Title>Titre de la page</Title>`
Cette balise permet d'indiquer à Google quel est le titre de la page.
- `<Meta name="description" Content="Description de la page">`
Cette balise permet de donner une courte description de la page, qui sera ensuite reprise sur les résultats en dessous du titre, dans ce qu'on appelle le « snippet », c'est-à-dire le texte figurant en dessous du lien hypertexte généré par le moteur.
- `<Meta name="googlebot" Content="...", ...">`
Cette balise permet d'indiquer au crawler ce qu'il doit faire du contenu d'une page en particulier. Dans notre exemple, le fait que nous ayons écrit « googlebot » permet de s'adresser au seul moteur Google, et non aux autres moteurs, auquel cas nous aurions écrit « robots ». La balise peut être spécifiée par plusieurs indications séparées par des virgules. Elle peut par exemple être spécifiée par un « noindex, nofollow » pour que la page ne soit pas indexée et que les liens qu'elle contient ne soient pas suivis. Par défaut, si la balise n'est pas implémentée, Googlebot référence la page et suit les liens. D'autres spécifications existent, dont voici les plus utilisées :
 - `nosnippet` : Google n'affichera pas de description de la page sous le lien hypertexte généré par le moteur.
 - `noarchive` : Google ne doit pas proposer de version « en cache » de cette page, c'est-à-dire qu'il ne doit pas proposer aux internautes d'accéder au contenu mis en ligne sur ses propres serveurs.
 - `unavailable_after: 8-Dec-2013 10:00:00 GMT` : Google n'affichera plus de liens pointant vers cette page sans ses listes de résultats après la date et l'heure spécifiées (dans notre exemple le 8 décembre 2013 à dix heures).

⁹² <http://googlewebmastercentral.blogspot.fr/2007/12/answering-more-popular-picks-méta-tags.html> (dernière visite le 21 août 2013)

- `<Meta Name="google" Value="notranslate">`
Cette balise permet à un éditeur de dire qu'il ne veut pas qu'une traduction soit proposée par Google pour le contenu de sa page, car en effet Google peut proposer de traduire une page dans la langue de l'internaute qui utilise le moteur.
- `<Meta Http-Equiv="Content-Type" Content="...; charset=...">`
Cette balise permet de spécifier le code des caractères utilisés sur la page. Elle permet à l'éditeur de s'assurer que le contenu apparaisse selon les modalités d'affichage qu'il a paramétrées⁹³.
- `<Meta Http-Equiv="refresh" Content="...;url=...">`
Cette balise peut être utilisée pour que l'internaute soit automatiquement redirigé vers une autre page après un certain délai.

Etant donné le champ de manœuvres que nous avons commencé à décrire, nous comprenons dès à présent qu'un contrôle partiel est exercé par l'éditeur sur l'infomédiation de ses contenus, dans la mesure où les méta-tags lui permettent de spécifier la nature et la durée de l'association de son action à l'action de Google. Ne pas se saisir de ces modalités, c'est laisser à Google la possibilité de définir l'action des crawlers : lui *déléguer* la capacité de faire certains choix dont dépendra *in fine* le traitement effectué du contenu publié par l'éditeur.

Les méta-tags ne sont pas seuls à permettre ce type de communication. Il existe d'autres moyens, qui cette fois n'ont pas pour but de spécifier l'association, mais de la renforcer en la fluidifiant : faciliter l'action des crawlers en leur faisant gagner de la bande passante. Les éditeurs peuvent notamment implémenter un « *Sitemap XML* », dont l'usage est fortement conseillé par Google⁹⁴.

Le *Sitemap* est une extension du « robots.txt » proposée initialement par Google et désormais adoptée par l'ensemble des moteurs. Il vient « en complément du crawl habituel ; il n'est donc pas indispensable d'utiliser ce système pour être

⁹³ En France, les types de codage les plus utilisés sont UTF-8 ou iso-8859-1.

⁹⁴ « [...] un Sitemap XML (appelé couramment Sitemap) est une liste des pages de votre site Web. La création et l'envoi d'un Sitemap vous permet de vous assurer que Google connaît toutes les pages de votre site, y compris celles qui pourraient ne pas être détectées par le processus d'exploration standard de Google. [...] Vous pouvez également utiliser les Sitemaps pour transmettre à Google des métadonnées relatives à certains types de contenu spécifiques présents sur votre site, y compris des vidéos, des images, du contenu pour mobile et des actualités. Par exemple, une entrée de Sitemap pour vidéos peut indiquer la durée d'une vidéo, sa catégorie et le public qu'elle cible (tous publics, par exemple). Une entrée de Sitemap pour images peut indiquer des informations concernant l'objet, le type ou la licence d'une image. Vous pouvez également utiliser un Sitemap pour fournir à Google des informations supplémentaires sur votre site, telles que la date de la dernière modification et la fréquence à laquelle vous pensez qu'une page sera modifiée. » (CAW)

indexé dans Google, par contre il peut éventuellement aider à mieux indexer les sites » (Duffez, 2005). Le *Sitemap* se présente sous forme de sommaire hypertexte permettant au crawler de savoir exactement sur quelles pages il devra se rendre. Le fichier, une fois créé, est soumis à Google. Soit il lui est envoyé via les « Outils Google pour les webmasters », soit son emplacement est signalé à la racine du site via le « robots.txt ». Le fichier est ensuite mis à jour automatiquement dès qu'une page est créée ou réactualisée, et soumis de nouveau *via* une requête HTTP⁹⁵.

Google préconise⁹⁶ de prévoir au moins deux cartes du site : l'une pour les internautes, accessible en général par un lien en pied de page, et l'autre, le *Sitemap*, destinée spécifiquement à Google : afin que ses crawlers puissent s'y retrouver aisément. Les crawlers sont ainsi comparés à des visiteurs d'un genre particulier, auxquels on s'adresse d'une façon différente de celle dont on s'adresse aux humains. Les crawlers, autant que les internautes, sont des *actants* du dispositif.

Il est sous-entendu par Google que l'utilisation du *Sitemap* peut permettre aux contenus de l'éditeur d'être mieux référencés, sans préciser à quel point et en leur assurant qu'ils ne pourront en aucun cas être pénalisés⁹⁷. Google tente ainsi de jouer sur le mécanisme d'intéressement/enrôlement afin de convaincre les éditeurs qu'ils auront raison de se mobiliser.

L'utilisation du *Sitemap* par les éditeurs va également dans le sens des intérêts de Google. Interrogé en 2005, l'ingénieur en charge de ce projet chez Google, Shiva Shivakumar, expliquait : « Nous voulons travailler en collaboration avec les webmasters afin d'obtenir une carte de toutes les URL que nous devrions crawler, et à quelle fréquence nous devrions les recrawler. Finalement, cela bénéficiera à nos usagers en augmentant l'étendue et la fraîcheur de notre index⁹⁸. » Le porte-parole utilise ici le mode conditionnel : il demande aux éditeurs de bien vouloir implémenter

⁹⁵ Une requête http est un ensemble de lignes de codes envoyées à un serveur par un navigateur.

⁹⁶ Google, Guide d'optimisation, p.11.

⁹⁷ [...] nous utiliserons les données incluses dans votre Sitemap pour en savoir plus sur la structure de votre site Web. Nous pourrions ainsi mieux planifier les explorations futures de votre site, lesquelles seront également plus performantes. Les Sitemaps se révèlent généralement utiles pour les webmasters et ne peuvent en aucun cas les pénaliser. » (CAW)

⁹⁸ « We want to work collaboratively with webmasters to get a big picture of all the URLs we should be crawling, and how often they should be recrawled. Ultimately this benefits our users by increasing the coverage and freshness of our index. » Extrait de l'interview effectuée par *Search Engine Watch* (notre traduction), disponible ici : <http://searchenginewatch.com/article/2061916/New-Google-Sitemaps-Web-Page-Feed-Program> (dernière visite le 21 août 2013)

le *Sitemap* pour savoir quelles URL, à quelle fréquence, il *devrait* crawler ; mais il ne leur garantit pas que ce sera le cas.

Ainsi, le *Sitemap* permet à Google de mieux connaître le site d'un éditeur et donc, potentiellement, d'être mieux à même d'identifier quand un des contenus correspond à une requête. C'est ce qui transparaît dans les propos de Shiva Shivakumar : « Ce programme est un complément au, et non un remplacement du, crawling régulier. Cependant, nous espérons que les données que vous inscrirez dans le *Sitemap* nous aiderons à mieux faire notre travail que le crawling régulier ne nous le permet. »⁹⁹

Voici les caractéristiques des balises d'un Sitemap généraliste¹⁰⁰ :

<urlset>	obligatoire	Contient toutes les informations sur l'ensemble d'URL inclus dans le Sitemap.
<url>	obligatoire	Contient toutes les informations concernant une URL spécifique.
<loc>	obligatoire	Indique l'URL. Indique pour les images et les vidéos la page de destination (page de lecture, page de provenance). Vous devez indiquer une URL unique.
<lastmod>	facultatif	Date de la dernière modification de l'URL au format AAAA-MM-JJThh :mmTZD (l'heure est facultative).
<changefreq>	facultatif	Donne une indication sur la fréquence de modification de la page. Les valeurs acceptées sont les suivantes : • always (Utilisez cette valeur pour les pages qui sont modifiées à chaque ouverture.) • hourly (chaque heure) • daily (chaque jour) • weekly (chaque semaine) • monthly (chaque mois) • yearly (chaque année) • never (Utilisez cette valeur pour les URL archivées.)
<priority>	facultatif	Décrit la priorité d'une URL par rapport aux autres URL du site. Ce niveau de priorité doit être compris entre 1.0 (importance élevée) et 0.1 (sans importance). Il n'affecte pas le classement de votre site dans les résultats de recherche Google. Cette valeur étant associée aux autres pages de votre site, le fait de définir un haut niveau de priorité (ou de spécifier le même niveau de priorité pour toutes les URL) n'améliore pas le classement de votre site dans les résultats de recherche. De même, l'attribution du même niveau de priorité à toutes vos pages n'a aucune incidence.

⁹⁹ « « This program is a complement to, not a replacement of, the regular crawl. However, we hope that the hints you offer in the Sitemap will help us do a better job than the regular crawl ». (*ibid.*, notre traduction)

¹⁰⁰ Il existe également des Sitemaps spécialisés, et notamment un Sitemap « Actualités » que nous décrirons plus tard. Les informations que nous donnons ici proviennent de <http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=183668&topic=8476&ctx=topic> (dernière visite le 21 août 2013)

Les informations données par l'éditeur sont spatiales (où se situent les documents ?) et temporelles (quand ont-ils été modifiés et quelle est la fréquence de leur modification ?), cela pour aider le crawler à définir au mieux où agir et à quel rythme. Par ailleurs, un éditeur a la possibilité d'indiquer ce qu'il estime être la priorité parmi ses documents. Google dit que cela n'a aucun impact sur le classement. Pourtant, si cela a un impact sur le crawler, nous ne voyons pas comment cela pourrait ne pas avoir d'impact sur la formation de la liste de résultats, dont le crawling constitue la première étape. En effet, Google essaye ici de dire aux éditeurs qu'il y aura un impact technique mais pas informationnel, comme si le déroulement du crawling n'avait pas d'influence sur le processus qu'il sous-tend.

Il existe d'autres manières pour l'éditeur de « faciliter la vie » des crawlers. L'une d'elles consiste par exemple à utiliser le protocole « robots.txt » pour exclure certaines zones trop volumineuses et/ou non pertinentes du champ de crawling. Cela permet à l'éditeur de diminuer les coûts en bande passante tout en augmentant la vitesse du crawling et donc, supposément, la fréquence de passage, dans l'espoir d'optimiser la captation de trafic. En effet, si le Googlebot identifie qu'un site lui coûte cher en temps et en bande-passante, il y passera moins souvent.

Nous voyons ici pourquoi se conformer aux préconisations de Google n'est pas forcément synonyme pour l'éditeur d'une absence de liberté mais peut être au contraire l'occasion de contrôler partiellement les modalités de crawling. En effectuant le déplacement que Google préconise, l'éditeur défend ses propres intérêts tout en servant ceux de l'acteur-réseau Google, auquel il est associé d'autant plus. Cela permet à la firme de fluidifier le fonctionnement du dispositif, selon ce que les concepteurs considèrent comme étant un fonctionnement adéquat, et ainsi de fortifier l'association de son action à l'action des internautes, puisque ces derniers recevront potentiellement de meilleures réponses à leurs requêtes.

4.1.5 Optimiser, sonder et contrôler

Ce que nous connaissons désormais du crawling nous a appris plusieurs points fondamentaux à propos du champ de manœuvres. D'une part, si les éditeurs utilisent un format HTML, leurs pages pourront être indexées par Google Search sans qu'ils

n'aient rien à faire. Par ailleurs, si un éditeur souhaite qu'une partie ou la totalité de son contenu soit désindexée, il devra implémenter le protocole « robots.txt » ou un méta-tag approprié. Il pourra ainsi rompre l'association ou bien la diminuer en la localisant sur une zone spécifique. Enfin, l'éditeur désireux de s'assurer que l'indexation ait lieu conformément aux souhaits des concepteurs du dispositif pourra s'impliquer davantage. Il lui faudra pour cela renseigner des méta-tags, implémenter une *Sitemap*, ou bien encore soulager le crawleur en excluant de son champ les fichiers trop volumineux. Ainsi, en fortifiant et en spécifiant l'association, il pourra potentiellement augmenter la fréquence du crawling et, peut-être, l'apport de trafic.

L'éditeur dispose de certains outils pour étudier la manière dont ses documents communiquent avec Googlebot — outils dont certains lui sont fournis par la firme elle-même. Il ne s'agit plus cette fois de spécifier, de réduire, de renforcer ou d'interrompre l'association, mais de la sonder et de la contrôler. Ainsi, l'outil « Explorer comme Googlebot » permet de « voir les pages comme Googlebot les voit »¹⁰¹. Google communique également les « Erreurs d'explorations » : liste des adresses URL des pages que Googlebot n'a pas réussi à explorer. Enfin, Google fournit des « suggestions HTML », c'est-à-dire des rapports à propos des difficultés rencontrées lors de l'exploration de certaines pages, que Google suggère alors à l'éditeur de modifier. Avec ces outils, l'éditeur peut procéder à un audit de ses pages pour essayer de déterminer ce qui peut expliquer la rupture d'association et tenter de réparer ses documents. Il peut ensuite demander à Googlebot de procéder à un réexamen de son site et de ses documents.

Enfin, les éditeurs peuvent se renseigner à propos des autres crawleurs de Google. Certains webmasters prétendent par exemple que le MediaPartners peut avoir une influence positive sur le référencement naturel, même si cela fait l'objet d'un démenti de la part des porte-parole de Google¹⁰². Quoi qu'il en soit, un éditeur désireux de mettre toutes les chances de son côté pourra devenir un client d'AdWords

¹⁰¹ Le navigateur Lynx, qui permet également de ne voir que l'information textuelle d'une page, est très utilisé par les référenceurs professionnels.

¹⁰² On trouvera par exemple un des nombreux articles mentionnant ce point et le débat dont il fait l'objet. Par exemple : <http://www.searchenginejournal.com/google-adsense-mediabot-is-crawling-sites-says-matt-cutts/3290/> (dernière visite le 21 août 2013)

et un partenaire d'AdSense simplement pour que son site soit crawlé le plus souvent possible, par différents crawleurs (cf. 8.3.1).

4.2 Indexation

Une fois qu'il connaît l'information, le bibliothécaire doit la ranger d'une manière qui lui permettra, ensuite, de la mobiliser. Pour cela, il répertorie dans un même fichier chacun des documents — son thème, sa date, son auteur et certains mots-clés décrivant les thématiques — pour être à même de retrouver plus tard un document en particulier ou bien tous les documents traitant d'un même thème. Pour un moteur, la création de ce fichier est appelée « indexation », et l'outil qui s'en charge « indexeur ». Là encore, le concepteur fait des choix : faut-il indexer le document intégralement ou bien seulement son titre ? sa description ? les méta-informations ? (Halavais, 2009, p. 17).

« [...] L'index forme la base sur laquelle s'exercera le traitement algorithmique qui suit, ses caractéristiques déterminent la capacité et la facilité qu'auront les algorithmes à trouver des résultats intéressants. Cependant, la richesse de l'index est proportionnelle à son "poids" et un index qui fournit beaucoup d'informations pour chaque page qu'il référence prend non seulement plus d'espace de stockage, mais devient aussi plus lourd à traiter par la suite. *Google* se vante (Brin et Page, 1998) d'établir un index très complet. Il contient le document entier pour les pages importantes ; une matrice recueillant les traits qui seront pris en compte par les algorithmes de recherche : par exemple une liste des mots uniques ; une table des hyperliens sortant de la page ou encore des indications sur la structure du document. En ce sens, les choix de la méthode d'indexation déterminent la *représentation interne* du terrain de recherche. Elle est nécessairement sélective. » (Rieder, 2006, p. 120)

Sans index, il faudrait que le moteur, pour chaque requête, effectue un audit de l'ensemble des documents en mémoire sur ses serveurs, ce qui lui coûterait extrêmement cher en temps et en énergie. Pour optimiser leur référencement, les éditeurs doivent faire attention aux mots employés dans les documents et à leur position (titre, sous-titre, chapeau, début du texte, milieu, fin). Car cela sera

déterminant pour l'indexeur. Ainsi, pour fortifier l'association, il faudra que les documents produits soient les plus explicites possible : éviter les sous-entendus, les jeux de mots et les métaphores.

Le système de tags permettra également à l'éditeur de mettre en avant certains mots pertinents. Cela constitue à nouveau une manière d'aider le dispositif à fonctionner, et donc de renforcer l'association. En revanche, certains éditeurs moins intéressés pourront faire des jeux de mots ou des sous-entendus, quitte à empêcher leurs documents d'être indexés par Google de manière optimale. Ils affaiblissent alors l'association, sans la rompre.

En même temps qu'il procède à l'indexation textuelle, l'indexeur enregistre la topologie du web dans le fichier « ancrs » (*anchors*), qui est une sorte d'index des liens hypertextes. Les crawlers rapportent en effet, en plus des informations à propos du document, les informations liées à sa position. Le dispositif identifie ainsi où sont les liens sortants, les liens entrants et les ancrs (les mots) sur lesquelles chaque lien est positionné. Pour qu'un document soit indexé au mieux, les liens qui pointent vers lui doivent être positionnés sur des ancrs le décrivant le mieux possible. Or un éditeur ne peut pas obliger les autres éditeurs qui feront des liens vers ses documents à les positionner sur des ancrs pertinentes. En cela, le bon déroulement de l'indexation d'un document dépend également des actions des autres éditeurs, en plus des actions de Google et de celui qui l'a mis en ligne.

Dans une étude récente (Richard, 2011), des indications sont données à propos des différentes manières dont un éditeur peut optimiser l'indexation. Il y est notamment conseillé de veiller au¹⁰³ :

- Poids des mots-clés selon leur place :
 - a. Présence dans le texte du document.
 - b. Présence dans les éléments de navigation : « Afin de séparer les contenus originaux d'une page des éléments de navigation communs à toutes les pages d'un site, Google détecte les éléments de texte et de code revenant

¹⁰³ Les différents éléments de cette partie proviennent du dossier consacré à l'ouvrage de Raphaël Richard par le Journal Du Net : <http://www.journaldunet.com/solutions/seo-referencement/google-decode-les-300-criteres-de-classement/> (dernière visite le 21 août 2013)

systématiquement aux mêmes endroits de la page. La présence d'un mot clé dans ces éléments est prise en compte mais son poids est minoré »¹⁰⁴.

- c. Présence du mot-clé dans le nom de domaine.
- d. Présence du mot-clé dans l'URL.
- e. Présence du mot clé dans l'attribut « Alt » des images.
- Poids des mots-clés selon les balises :
 - f. Présence du mot-clé ou d'un synonyme très proche les balises H1, H2, H3, H4, H5, H6 (mettre un mot-clé dans le titre et les sous-titres de la page, quels que soient leurs niveaux).
 - g. Proximité du mot-clé du début de la balise <Title> (mettre le mot-clé le plus tôt possible dans le titre).

Selon Raphaël Richard (*op. cit.*), il ne faut pas hésiter à répéter un mot-clé plusieurs fois dans le texte, à utiliser ses versions déclinées (féminin, pluriel), et à utiliser des synonymes et un champ lexical proche du mot-clé principal. Ces paramètres influencent, à des degrés variés (et inconnus), l'indexation et l'identification, et donc, possiblement, l'apport de trafic.

4.3 Appariement

Des visiteurs viennent voir le bibliothécaire et l'interrogent. En fonction de ce qu'ils disent et de ce qu'il sait lui-même, celui-ci leur fait des recommandations. Il apparie une requête et une liste de résultats. Pour Google Search, c'est le moment dont dépend l'utilité perçue par l'internaute et l'apport de trafic des éditeurs et des annonceurs. Aucun acteur ne peut savoir, avant que l'interaction ait eu lieu, si oui ou non Google aura effectivement représenté ses intérêts. C'est ici que l'algorithme, l'infrastructure, le crawler et l'indexeur inter-agissent : de ce qu'ils sont (leurs paramètres) et de ce qu'ils peuvent faire (leurs capacités) dépendent les

¹⁰⁴ *Ibid.*

caractéristiques de la liste de résultats et le trafic que les éditeurs capteront (ce qu'ils auront fait *effectivement*)¹⁰⁵.

Le résultat de cette interaction est le fruit de la procédure algorithmique appliquée aux signaux d'entrée récoltés par le crawler et traités par l'indexeur. Le nombre de critères de l'algorithme, ainsi que leurs pondérations, sont inconnus et régulièrement mis à jour par les concepteurs de Google. Ils sont 200, selon la firme (CAW), 300 selon certains observateurs (Richard, 2011). Parmi eux, certains sont tout de même connus, et il nous paraît intéressant de nous y intéresser¹⁰⁶, dans le but de voir ce qu'ils peuvent nous apprendre au sujet du script des concepteurs et du champ de manœuvres des éditeurs.

4.3.1 PageRank

Ce que nous apprend le PageRank (cf. 2.2.4) dans le cas d'un éditeur est en premier lieu que la publicisation de ses contenus ne dépend pas seulement de ses actions, des actions des internautes ou des actions des employés de Google, mais également des actions des autres éditeurs. Les liens que ces derniers effectuent vers ses contenus augmenteront en effet sa pertinence supposée. Ces votes augmenteront également l'impact des liens qu'il fera lui-même vers les contenus d'autres éditeurs. Ainsi, la capacité d'un éditeur en matière de stockage et de déploiement de l'autorité liée au PageRank dépend *aussi* des actions de ses homologues.

Plus un contenu est visible sur Google Search et plus il sera consulté et cité. En outre, plus un éditeur jouit d'une image de marque forte et plus il y aura de probabilités que ses contenus soient cités. Ainsi, sur Google Search, ceux qui sont visibles deviennent plus visibles : les riches sont plus riches (Rieder, 2012 ; Cardon,

¹⁰⁵ « Une fois l'index mis en place, les développeurs sont confrontés à une autre question : *comment chercher dedans*. Les différentes informations et caractéristiques représentées dans l'index constituent la *matière première* que va traiter l'algorithme de recherche. Un niveau de décision supplémentaire s'impose : quels traits prendre en compte et comment les traiter ? Question difficile car l'algorithme résultant de la réponse à cette question est censé fournir des résultats structurés, dans la plupart des cas hiérarchisés, dont on puisse dire qu'ils sont « pertinents » sur le plan de la « qualité » de l'information qu'ils véhiculent. De manière abstraite, on pourrait dire que la problématique consiste à savoir comment réduire le plus possible la « distance » entre la requête formulée par un usager et le document qui répondra le mieux à ses attentes. La réponse technique doit venir du traitement algorithmique de l'index élaboré à partir du terrain d'information visé. *Google* doit en très grande partie son succès à son moteur de classement dont les résultats se distinguaient à un moment assez fortement par leur qualité – telle qu'elle était perçue par l'ensemble des internautes – de ceux de ses concurrents. » (Rieder, 2006, p.120-121).

¹⁰⁶ Pour une liste des critères hypothétiques et de leurs effets probables, voir par exemple : <http://www.leonard-rodriguez.com/blog/search-marketing/seo-search-marketing/liste-complete-des-facteurs-en-referencement-naturel-seo-384#Offpage2> (dernière visite le 21 août 2013)

2013). L'éditeur redistribue son autorité. Il pourra faire des liens vers les contenus d'autres éditeurs ou bien vers ses propres pages. Plus il fera de liens et moins chaque lien sera porteur de valeur. La stratégie consistant à canaliser le PageRank aux bons endroits dans le but d'optimiser le référencement des documents s'appelle « *PageRank sculpting* ». Cette stratégie est difficile à mettre en œuvre et peut être contre-intuitive, comme nous le verrons en particulier dans la Partie 3 (chapitre 8).

4.3.2 Taux de clics

Le taux de clics (nombre de clics sur un lien divisé par le nombre d'affichages dans une liste de résultats) est intégré à l'algorithme pour définir la pertinence d'un document. Là encore, selon ce principe, les plus visibles deviennent plus visibles. Un éditeur dont le site se trouve en haut des résultats reçoit potentiellement plus de trafic qu'un éditeur qui se retrouve en bas de la même liste, ce qui augmente son taux de clics. A l'inverse, le classement de l'éditeur dont le lien est en bas de la même page de résultats peut diminuer, et avec lui le trafic reçu par l'éditeur en question.

4.3.3 Fraîcheur des documents

Le critère de fraîcheur contrebalance en partie le PageRank et le taux de clics, en cela qu'il permet de valoriser les documents mis en ligne récemment. « Le besoin de fraîcheur diffère selon les types de requêtes, explique un porte-parole de Google. Cet amélioration de l'algorithme a été faite pour mieux comprendre comment différencier ces types de requêtes et leurs besoins de fraîcheur respectifs »¹⁰⁷. Il n'était pas en place le 11 septembre 2001, et c'est la raison pour laquelle les internautes qui formulaient la requête « *world trade center* » n'étaient pas dirigés vers des documents mentionnant la catastrophe en cours (cf. Introduction). Ainsi, ce critère revêt une importance particulière pour l'information d'actualité.

¹⁰⁷ « Different searches have different freshness needs. This algorithmic improvement is designed to better understand how to differentiate between these kinds of searches » <http://insidesearch.blogspot.fr/2011/11/giving-you-fresher-more-recent-search.html> (Notre traduction, dernière visite le 21 août 2013))

4.3.4 Efficacité de l'infrastructure

Google prend en compte l'efficacité de l'infrastructure, et notamment la vitesse de chargement de chaque page depuis février 2009. La firme « conseille vivement aux webmasters de surveiller régulièrement les performances de leur site en utilisant *Page Speed*, *YSlow*, *WebPagetest* ou d'autres outils » (CAW). Elle justifie la prise en compte de ce critère en expliquant que la firme a toujours elle-même été obsédée par la vitesse¹⁰⁸. Ainsi, selon Google, un bon site est avant tout un site efficace. La firme fait de la *performance* un critère de *pertinence*. La « couche logique », permettant la traduction du langage humain en langage technique, *et vice versa*, se base alors sur l'analyse de la « couche physique » pour organiser et hiérarchiser la « couche des contenus ».

Les éditeurs désireux de maximiser leur trafic devront donc investir dans des infrastructures afin de réduire la vitesse de chargement de leurs sites, tout en s'assurant que les documents soient dans un format qui ne sera pas trop gourmand. Cela représente des coûts et a par conséquent tendance à avantager les plus riches. Un éditeur pourra par exemple investir dans des infrastructures de « *Content Delivery Network* » (CDN) afin que les requêtes de ses lecteurs soient servies, grâce à un mécanisme de routage, par le « nœud » du réseau le plus proche, optimisant ainsi la transmission/réception. Au sujet du coût que représente une telle opération, les prix pratiqués par exemple par l'entreprise *OVH* (spécialiste de solutions de CDN), laquelle présente d'ailleurs cette solution comme susceptible d'améliorer considérablement le trafic provenant des moteurs de recherche, varient de 9,99 € à 599,99 € euros par mois (hors taxe).

¹⁰⁸ « Vous avez probablement entendu dire que, chez Google, nous sommes obsédés par la vitesse, pour ce qui est de nos produits et du web en général. Pour aller dans ce sens, nous incluons aujourd'hui un nouveau signal dans nos algorithmes de classement : la vitesse du site. La vitesse du site reflète à quel point un site web répond rapidement aux requêtes. Accélérer le chargement des sites est important — pas seulement pour les propriétaires des sites, mais pour tous les utilisateurs d'Internet. Des sites plus rapides créent des utilisateurs heureux et nous avons vu lors d'études internes que lorsqu'un site répond lentement, les visiteurs y passent moins de temps. En outre, des sites plus rapides ne font pas qu'améliorer l'expérience utilisateur, certaines données récentes montrent en effet qu'améliorer la vitesse d'un site diminue également les coûts de l'opérateur. Comme nous, nos utilisateurs valorisent considérablement la vitesse — c'est pourquoi nous avons décidé de la prendre en compte lors de nos hiérarchisations. » — CAW, notre traduction, disponible ici : <http://googlewebmastercentral.blogspot.fr/2010/04/using-site-speed-in-web-search-ranking.html> (dernière visite le 21 août 2013)

4.3.5 Qualité éditoriale

Les porte-parole de Google indiquent que les contenus de qualité seront avantagés, tout en restant flous sur ce que la « qualité » d'un contenu signifie, et sur la manière dont elle peut véritablement influencer le classement : « Fournissez un contenu de haute qualité sur vos pages, et spécialement sur votre page d'accueil. C'est la seule chose importante à faire. Si vos pages contiennent des informations utiles, leur contenu attirera de nombreux visiteurs et incitera les webmasters à effectuer des liens vers votre site. Pour créer un site utile et riche, rédigez vos pages de manière à décrire clairement et exactement le sujet dont il est question. » (CAW)¹⁰⁹

Il s'agit d'une montée en généralité, qui ne s'appuie à notre connaissance sur aucun critère technique, si ce n'est l'impact que peut avoir la qualité d'un contenu sur le taux de clics. Google se fait ici le porte-parole des internautes, que la firme représente auprès des éditeurs. Pour intéresser les éditeurs et essayer qu'ils se mobilisent, Google leur explique qu'un contenu de qualité leur permettra d'attirer davantage de visiteurs, tout en leur indiquant que d'autres éditeurs pourront eux-mêmes effectuer des liens vers leurs contenus, ce qui augmentera leur PageRank. Par ailleurs, Google défend ses propres intérêts en demandant aux éditeurs de décrire clairement le sujet de leurs pages. Cela fluidifiera le fonctionnement du dispositif et augmentera les chances de Google de répondre au mieux aux requêtes.

Même si Google a clamé le contraire pendant longtemps, il a finalement été révélé qu'un jugement humain était donné sur la qualité des contenus éditoriaux. Des travailleurs à domicile, employés par les entreprises *Leapforce*, *Butler Hill* et *Lionbridge*, jouent en effet un rôle dans le processus de sélection et de hiérarchisation. En 2011, certains observateurs ont réussi à se procurer une copie du manuel de 160 pages qui leur était transmis¹¹⁰. Il y est demandé aux évaluateurs de donner une des appréciations suivantes à chacun des sites visités : « *Vital* »,

¹⁰⁹ « Provide high-quality content on your pages, especially your homepage. This is the single most important thing to do. If your pages contain useful information, their content will attract many visitors and entice webmasters to link to your site. In creating a helpful, information-rich site, write pages that clearly and accurately describe your topic. », notre traduction, disponible ici : <http://support.google.com/webmasters/bin/answer.py?hl=en&answer=40349>

¹¹⁰ <http://searchengineland.com/download-the-latest-google-search-quality-rating-guidelines-97391> (dernière visite le 21 août 2013)

« *Useful* », « *Relevant* », « *Slightly Relevant* », « *Off-Topic or Useless* » ou « *Unratable* ». Un contenu qui a plus de quatre mois ne peut pas être considéré comme « *Vital* »¹¹¹. Ainsi, la fraîcheur est là encore un critère permettant à Google de juger de la pertinence de ce document. En plus de ceux-là, un autre indicateur est renseigné manuellement : le « *Page Quality* ». Pour cet indicateur, les travailleurs doivent juger de la réputation du site : bonne, mauvaise, ambiguë, OK, malicieuse ou impossible à déterminer. Nous ignorons quel est l'impact de ce travail manuel sur la liste de résultats. Et nous ignorons également si un tel travail manuel a lieu en France. Cependant, il est intéressant de considérer que cette révélation quant au travail manuel nourrit la réflexion selon laquelle le moteur peut être conduit à faire « exactement le contraire de son programme d'action » en limitant « l'automatisme de l'algorithme » (Farchy et Méadel, 2013).

4.3.6 AuthorRank

Depuis juin 2011, Google calcule un critère appelé *AuthorRank*, mesurant la popularité et la pertinence supposée d'un auteur en particulier, et ce même si cet auteur publie des documents sur des sites différents. Ainsi, grâce à la balise « rel=author » implémentée dans les métadonnées du document, et reliée à un compte Google+, ou bien encore grâce à un « badge Google+ », le dispositif identifie les documents qui sont le fait du même auteur. Il s'agit d'un dispositif de « signature numérique » reliée à une « plate-forme d'indentification » (Kohn, 2012).

Google explique que cet indicateur est un moyen « d'aider les internautes à trouver des contenus provenant de bons auteurs » (CAW). L'*AuthorRank* n'a pas pour vocation de remplacer PageRank, mais de le préciser. S'ajoutant aux critères de l'algorithme, selon une pondération et des modalités dont le détail n'est pas communiqué, il ajoute un symbole d'entrée au calcul d'appariement. Selon les spécialistes, qui ont étudié ce que pouvait leur apprendre le brevet déposé par Google à ce sujet en 2005¹¹², les critères pris en compte pourraient être (Arnesen, 2012) :

¹¹¹ http://www.theregister.co.uk/2012/11/27/google_raters_manual/ (dernière visite le 21 août 2013)

¹¹² <http://www.google.com/patents/US20070033168> (dernière visite le 21 août 2013)

- Le PageRank moyen des pages signées par l’auteur
- Le nombre de « +1 » (l’équivalent du « *Like* » de Facebook) et de partages dans Google + dont les pages signées par l’auteur ont fait l’objet.
- Le nombre de cercles de Google+ où l’auteur figure.
- Les connections réciproques entre l’auteur et d’autres auteurs dont l’*AuthorRank* est élevé.
- Le nombre et l’autorité des sites où l’auteur publie du contenu.
- L’activité de l’auteur sur Google+ (le nombre de « *posts* »)
- L’engagement de l’auteur auprès des lecteurs : le nombre de commentaires que les articles d’un auteur reçoivent en moyenne, et le nombre de fois où l’auteur a répondu à ces commentaires.
- Les indicateurs extérieurs d’autorité (par exemple, le fait qu’une page soit consacrée à l’auteur sur Wikipédia).
- Les travaux publiés par l’auteur et référencés sur Google Books et Google Scholar.
- Le fait que l’auteur ait un compte sur Youtube et qu’il y soit actif.
- Les signaux sociaux provenant des réseaux Twitter, Facebook, LinkedIn, SlideShare, Quora, etc. (Réseaux dans lesquels Google a « confiance » et auxquels les robots-crawlers ont, au moins partiellement, accès.)

Pour les éditeurs de presse, cela signifie qu’ils ont tout intérêt à recruter des journalistes actifs sur le web, dont les contenus sont souvent partagés, possédant et utilisant un compte sur le réseau social Google+, lequel aura été ajouté par de nombreux internautes dans leurs « cercles », présents et actifs sur d’autres réseaux sociaux, prêts à interagir avec les internautes qui commenteront leurs articles, ayant éventuellement publié des livres, des articles scientifiques et possédant une page à leur sujet sur Wikipédia. De tels journalistes pourront, quant à eux, monnayer leur autorité auprès de leur employeur en utilisant ce critère, et son impact présumé sur l’audience du site.

Google se sert ainsi de sa situation dominante sur le marché des moteurs pour mettre en avant ses services Google+ et Youtube. En quelque sorte, la firme dit aux éditeurs : « Si vous vous voulez maximiser l’audience que vous apportera Google (*intéressement*), vous avez intérêt à vous investir sur Google+ et Youtube (*enrôlement et mobilisation*). »

4.3.7 Signaux sociaux

Les « posts » sur les réseaux sociaux comme *Facebook*, *Twitter* et *Google+*, ainsi que les recommandations effectuées par les internautes via des dispositifs de

« *Like* » ou de « *+1* », sont également pris en compte pour le classement des documents (SearchMetrics, 2012). Dans une interview de janvier 2012, Amit Singhal, porte-parole régulier de la firme Google, justifie l'attention portée aux signaux sociaux de la manière suivante : « Un bon produit ne peut être construit qu'à condition que nous comprenions qui est qui, et qui est relié à qui. Les relations sont également importantes au sein même du document. [...] Car fondamentalement, il ne s'agit pas que de contenu. Il s'agit d'identité, de relations et de contenus. »¹¹³

Cette fois, ce sont les internautes qui agissent et participent à la publicisation des contenus qu'ils ont appréciés. Nous retrouvons l'idée d'une « échelle de visibilité collective » (Cardon, 2010), construite conjointement par les éditeurs et les internautes. Pour l'éditeur, l'attention portée aux signaux sociaux l'invite à implémenter sur ses pages les dispositifs de recommandation, et notamment le « *+1* » de Google, tout en étant présent sur les réseaux sociaux dans l'espoir de multiplier les interactions dont son contenu est l'objet. Même si sa présence sur les réseaux n'est pas directement rentable (puisqu'il ne peut y afficher de publicité), elle peut l'être indirectement en influençant la visibilité sur le moteur Google Search, et en aboutissant à un apport de trafic, qui, lui, sera monétisé.

4.3.8 Tendances de Recherche

Google demande aux webmasters d'imaginer les termes susceptibles d'être utilisés par les internautes pour trouver leur site et d'insérer ces termes dans leurs pages. Google propose pour cela l'outil « Tendances de recherche », mis en ligne en 2006 puis, dans sa version améliorée, en 2008. Ce service renseigne quant aux volumes liés à des termes de requêtes, dans des pays et durant des périodes que l'on peut spécifier. Il est également possible d'effectuer des comparaisons afin de voir quels sujets et quels mots, dans quels pays, quand, ont été les plus prisés. L'outil donne également, pour certains termes, les prévisions de recherche pour les mois à venir. Il est ainsi possible de connaître quelles actualités font l'objet, à un moment

¹¹³ « A good product can only be built where we understand who's who and who is related to whom. Relationships are also important alongside content. To build a good product, we have to do all types of processing. But fundamentally, it's not just about content » (Sullivan, 2012, Notre traduction)

donné, sur une zone géographique définie, de requêtes sur le moteur. En revanche, l'information délivrée est incomplète :

« Les nombres du graphique indiquent la quantité de recherches ayant été effectuées pour un terme donné, par rapport au nombre total de recherches effectuées sur Google au cours de la même période. Ils ne représentent pas le volume de recherche en valeur absolue, car les données sont normalisées et présentées sur une échelle allant de 0 à 100. Chaque point du graphique est divisé par le point le plus élevé ou par 100 » (CAW¹¹⁴).

Un éditeur pourra donc savoir que le traitement d'un certain sujet pourrait lui apporter de l'audience en provenance de Google sans jamais savoir quelle est la taille, en valeur absolue, de l'audience en question. En publiant ainsi ce que les usagers de son moteur cherchent, Google parle en leur nom. Dès lors, le résultat qui apparaît sur l'interface est une *inscription* au sens de l'ANT, dont les éditeurs pourront se saisir dans le but de renforcer l'association de leur cours d'action à celui de Google, quitte pour cela à modifier leur stratégie éditoriale. « Tendances de Recherche » a ainsi pour fonction de *traduire* (interpréter et déplacer) les actions effectuées par les internautes sur Google Search, à partir de quoi il génère un énoncé susceptible de performer la réalité dont le moteur rend compte.

L'éditeur désireux de maximiser l'apport d'audience dispose de deux solutions. D'une part, il peut se positionner sur les mots et les sujets qui, effectivement, sont les plus en vogue. Mais il risque alors de faire face à une concurrence très forte et de se trouver, malgré ses efforts, derrière des sites dont les documents auront un PageRank, une vitesse de chargement, etc., plus élevés que les siens. D'autre part, « Tendances de Recherche » peut servir à savoir quels sujets un éditeur ne traitera pas et quels mots il n'emploiera pas : sa stratégie consistera alors à se positionner sur la « longue traîne » (Anderson, 2006 ; Smyrnaioi *et al.*, 2010), c'est-à-dire sur des mots et des sujets à faible potentiel, dans l'espoir de récupérer un trafic faisant l'objet d'une concurrence moindre.

¹¹⁴ <http://support.google.com/insights/bin/answer.py?hl=fr&answer=87285> (dernière visite le 21 août 2013)

4.4 Interface

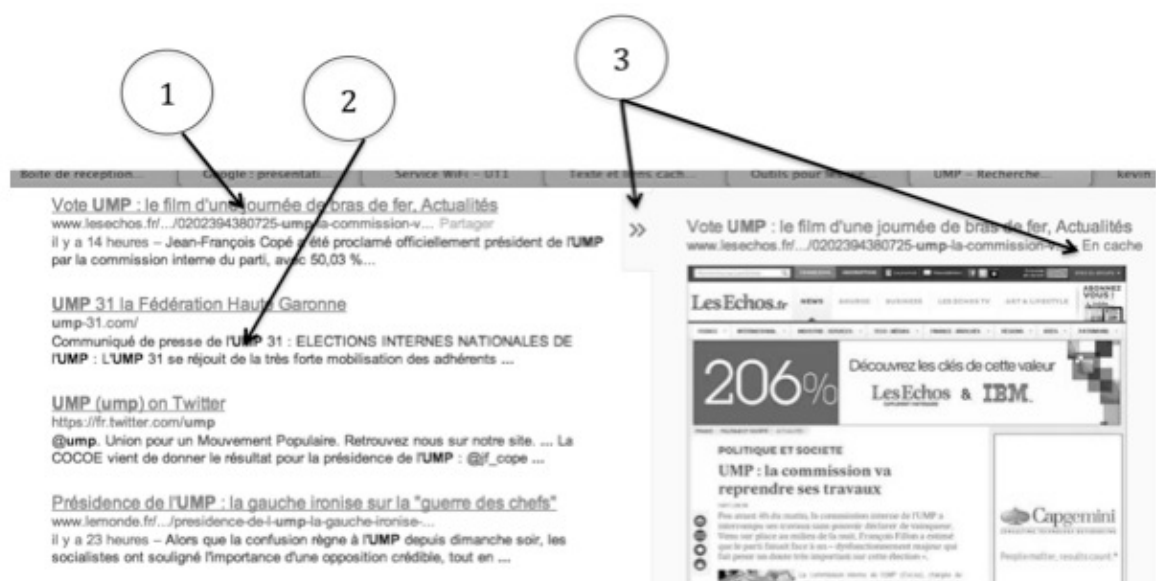
La dernière étape de notre description correspond à la dernière batterie de questions que doit se poser le concepteur d'un moteur, concernant les modalités d'affichage de la liste de liens hypertextes. La liste comprend dix résultats naturels sur la première page, ainsi que des liens sponsorisés à droite, en haut et parfois en bas de la page. Pour les résultats naturels, nous distinguons quatre types de liens : les liens simples, les liens enrichis, les liens « moteur verticaux » et les liens de reformulation.

4.4.1 Liens simples

Le mardi 20 novembre 2012, la victoire de Jean-François Copé sur François Fillon était décrétée dans le cadre de l'élection à la présidence du parti UMP, après deux jours d'hésitations pendant lesquels les voix des militants avaient été comptées et recomptées, et les victoires de l'un et de l'autre candidat annoncées successivement par leurs partisans. Nous avons effectué ce jour-là une requête « UMP » sur Google Search et obtenu des liens sous la forme suivante :

Figure 1. Capture d'écran : liste de liens sur Google Search.

(Requête « UMP » - 20/11/2012)



L'internaute voit apparaître les liens simples (1) sous forme de titres cliquables, et de snippets, non cliquables. On y trouve également l'adresse URL de la page et la fraîcheur. Les mots qui apparaissent en gras dans le « snippet » (2) sont ceux qui ont été inscrits par l'internaute dans Google Search lors de sa requête. Ils indiquent qu'il y a bien appariement entre la requête et les résultats.

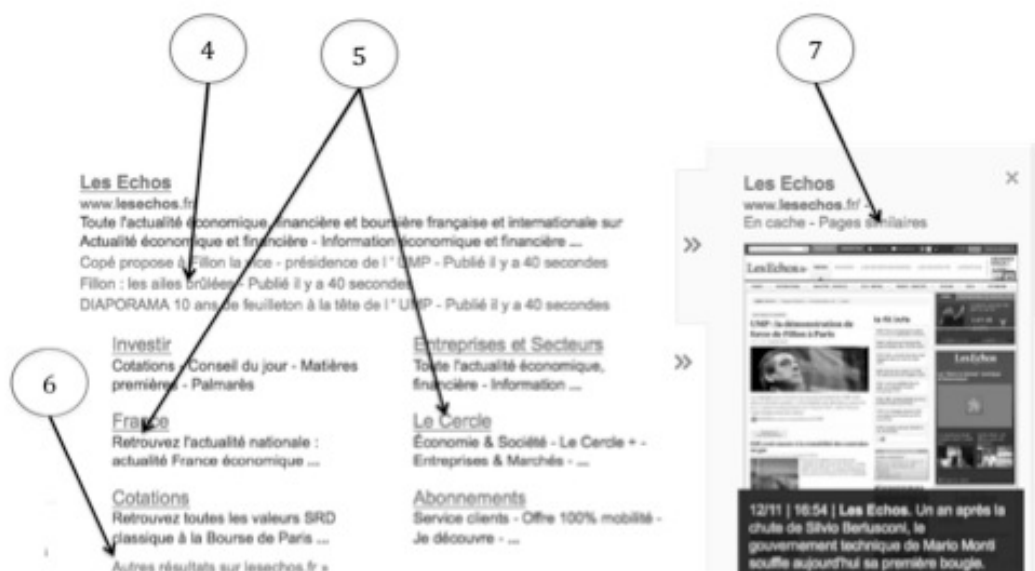
Un système de prévisualisation (3) permet à l'internaute de voir la page sans avoir cliqué sur le lien. Il faut pour cela passer le curseur sur la double flèche à droite du lien. L'internaute qui aurait seulement pré-visualisé la page n'est pas compté comme un visiteur et ne peut donc pas être monnayé par l'éditeur auprès de son réseau d'annonceurs, et ce même si les publicités apparaissent en effet (ici une bannière *CapGemini* et une publicité pour un partenariat entre *Les Echos* et IBM).

Dans l'onglet de prévisualisation, le lien « en cache » permet d'accéder non pas au contenu directement mais au contenu enregistré par les crawleurs et mémorisé sur les serveurs de Google. Dans le cas d'une « Erreur 404 », due par exemple à un retrait par l'éditeur de la page indexée par Google, le « cache » permet ainsi à l'internaute d'accéder quand même au contenu, comme il a été crawlé pour la dernière fois. Nous verrons que cela a été l'objet d'une importante controverse entre Google et les éditeurs de presse belges (cf. 6.1.4).

4.4.2 Liens enrichis

Si nous tapons « Les Echos » dans Google Search, nous tombons cette fois sur un lien enrichi en première position, suivi de liens simples.

Figure 2. Capture d'écran : liste de liens sur Google Search.
 (Requête « Les Echos » - 20/11/2012)



Le lien est considéré par Google comme un seul et même résultat. Par conséquent, l'éditeur est pointé par plusieurs liens au lieu d'un seul (ici dix au total) et occupe davantage d'espace sur la page. Google n'affiche des liens enrichis que lorsque la requête est explicite, par exemple quand le nom de l'éditeur correspond exactement à la requête formulée (ici « *Les Echos* »). Nous retrouvons dans les liens enrichis les caractéristiques du lien simple, auquel sont ajoutés certains liens secondaires. Parmi eux, les liens « Actualités » (4) indiquent quels sont les derniers articles mis en ligne. Ils sont accompagnés de l'heure de publication et permettent de proposer un « lien profond » et ce même si l'internaute s'est servi du moteur comme d'un « *bookmark* » pour accéder à un site d'information, et non à une information en particulier. Les liens « rubriques » (5) sont au nombre de huit au maximum et indiquent les rubriques du site de l'éditeur, permettant ainsi à l'internaute de se rendre directement sur la page d'accueil d'une rubrique. Si les noms sont trop longs, il peut arriver que les titres des rubriques soient reformulés ou écourtés automatiquement par le dispositif. L'éditeur peut choisir, *via* l'interface « Webmasters Tools » mise à sa disposition par Google (pouvoir de faire), de bloquer certaines rubriques, qui n'apparaîtront pas alors dans le lien enrichi. En revanche,

l'éditeur ne pourra pas choisir l'ordre d'apparition des rubriques. Cela empêchera par exemple un éditeur de mettre en avant sur la page d'accueil de son site la rubrique « Politique » et sur Google Search la rubrique « People », simplement parce qu'il sait que les utilisateurs du moteur cliquent plus facilement sur cette dernière.

Les « Autres résultats » (6) permettent à l'internaute de préciser à Google Search que la requête ne doit être effectuée que sur ce site¹¹⁵. Une nouvelle requête sera effectuée et, dans notre exemple, seuls des liens vers des pages du SIPI adossé au journal *Les Echos* seront proposés. Un clic sur le lien « Pages similaires » (7) relancera quant à lui une requête sur Google, où seront affichés, sous forme de liens simples, l'ensemble des sites des éditeurs identifiés par le dispositif comme ayant la même spécialité que l'éditeur pré-visualisé. Dans notre exemple, nous obtiendrons une liste de liens vers les pages d'accueil des éditeurs plus ou moins concurrents du site web *Les Echos*, c'est-à-dire par exemple : Latribune.fr, Lefigaro.fr, Lexpress.fr, lemonde.fr, Libération.fr, Lepoint.fr, Alternatives-économiques.fr, Challenges.fr, Boursorama.com. Les éditeurs ne peuvent pas demander à Google de figurer dans une liste en particulier.

Lorsque la balise « rel=author » est utilisée, ou le « badge Google+ », reliés au compte Google+ d'un journaliste, le nom de l'auteur, ainsi que sa photo, le nombre de « cercles » auxquels l'auteur appartient et le nombre de commentaire dont son « post » sur Google+ a fait l'objet, peuvent apparaître dans la liste de résultats sous forme de lien enrichi :

Figure 3. Capture d'écran : lien vers un article de Danny Sullivan sur Google Search.



Nous retrouvons ici l'idée selon laquelle « l'identité personnelle se donne davantage comme un processus que comme un état, une activité plutôt qu'un statut. La fabrication identitaire épouse étroitement les technologies numériques, dont la

¹¹⁵ Il s'agit de l'équivalent de la commande « site: » utilisée dans la barre du moteur. Il s'agit de l'équivalent de la commande « site: » utilisée dans la barre du moteur.

plasticité et l'interactivité favorisent ce travail de subjectivation, entendu comme un processus continu de (re)création de soi » (Cardon, 2010, p. 60). L'implémentation du dispositif de signature numérique « rel=author » peut en effet être considéré comme une stratégie d'exposition de soi. L'intérêt pour l'auteur est de mieux valoriser ses liens au sein de la liste de résultats et d'espérer capter davantage d'audience, notamment grâce à la prise en compte de *l'AuthorRank*. L'auteur devient sa propre marque. Selon le site *CatalystSearchMarketing.com*, le recours à ce dispositif a augmenté son taux de clic de 150%¹¹⁶.

Google se réserve toutefois le droit de considérer ou non l'indication « rel=author »¹¹⁷. Ainsi, même si un « pouvoir de faire » est attribué aux éditeurs, il n'en demeure pas moins que le « pouvoir sur » demeure du côté de Google. Nous avons trouvé un exemple de non considération de l'indication dans un article consacré à ce sujet par *Search Engine Land*. Le journaliste y écrit : « J'allais utiliser un de mes articles comme exemple, mais Google a annulé la prise en compte de mon statut rel=author pour des raisons que j'ignore ».¹¹⁸

Enfin, l'affichage des liens générés par Google peut également être enrichi par les éditeurs grâce à l'utilisation de « données structurées » (ou balises sémantiques). Depuis 2011, les données structurées font l'objet d'un protocole nommé « Schema.org ». L'éditeur qui les implémente influencera la visualisation de ses pages sur tous les moteurs. Par exemple, pour une recette de cuisine, il est possible de faire apparaître le temps de préparation ainsi que la notation des internautes sous forme d'étoiles :

Figure 4. Capture d'écran : lien vers le site « Thaifood » sur Google Search.



¹¹⁶ <http://www.emarketinglicious.fr/webmarketing/autorship-authorrank-rel-author-pourquoi-vous-en-preoccuper-des-maintenant> (dernière visite le 21 août 2013)

¹¹⁷ « Nous ne pouvons pas garantir que les informations relatives à l'auteur s'afficheront dans les résultats de la Recherche sur le Web Google ou de Google Actualités » (CS).

¹¹⁸ I was going to use one of my articles as an example, but Google has turned off my rel=author status for some reason. » Notre traduction <http://searchengineland.com/google-tweaks-its-relauthor-display-promotes-google-in-search-results-98972> (dernière visite le 21 août 2013)

Avant la mise en place de ce protocole, il existait trois formats de données structurées : les microdonnées, les microformats et les RDFa. Le protocole Schema.org vise à ce que le format « microdonnée » soit employé de préférence aux deux autres. Google supporte toujours les deux autres formats mais plaide en faveur d'une utilisation des microdonnées¹¹⁹.

Les éditeurs de presse en ligne sont concernés en particulier par les microdonnées relatives aux articles de presse, appelées « *NewsArticle* ». Ainsi, ils pourront à la fois spécifier et renforcer leur association à l'action de Google en contrôlant et en enrichissant les modalités d'affichage de leurs liens, sans pour autant, à aucun moment, être assuré que Google suivra les indications qui lui sont données. Le champ de manœuvres permet notamment à l'éditeur d'indiquer au moteur :

- La personne physique ou morale légalement responsable du contenu ;
- L'auteur du contenu ;
- Un autre auteur pour le cas où l'article serait co-signé ;
- Les récompenses remportées par le contenu ou par son auteur (le prix Pulitzer par exemple) ;
- La zone dédiée aux commentaires des internautes ;
- La localisation du contenu (utile pour la PQR) ;
- Le propriétaire du *Copyright* du contenu ;
- L'année de *Copyright* ;
- La date de création ;
- La date de la dernière modification ;
- La date de publication ;
- Le nom de la personne ayant édité ou revu le contenu ;
- Le langage dans lequel est rédigé le contenu ;
- Le genre du contenu (article, chronique, édito...) ;
- Une image (*thumbnail*) associée à ce contenu ;
- La rubrique du contenu (Monde, Société, Sport...)
- Le nombre de mots du contenu ;
- Le nombre de colonnes que l'article occupe dans la version papier ;
- A quelle page figure ce contenu dans la version papier ;
- Dans quelle rubrique de la version papier le contenu apparaît.

¹¹⁹ <http://googlewebmastercentral.blogspot.fr/2011/06/introducing-schemaorg-search-engines.html> (dernière visite le 21 août 2013)

Nous voyons avec ces trois derniers éléments que le Schema.org permet de faire le lien entre la version imprimée d'un article et la version mise en ligne, pour le cas où l'éditeur serait à la fois détenteur d'un titre imprimé et d'un site internet. Cela nous intéresse particulièrement dans la mesure où notre terrain de Partie 3 est composé de sites issus de la presse imprimée.

4.4.3 Liens moteurs verticaux

Des transferts de requêtes peuvent être proposés par le moteur généraliste vers des moteurs verticaux, c'est-à-dire des moteurs dont le champ de crawling est restreint à certains sites (Google Actualités) ou à certains formats de documents (Google Images). Ces propositions de transferts ont pour but de permettre à l'internaute de préciser le type de contenu auquel il s'intéresse en particulier.

Figure 5. Capture d'écran : Transfert de requête vers Google Actualités
(Requête « UMP » - 20/11/2012)



Dans la capture ci-dessus (Fig. 5), un clic sur « Actualités correspondant à UMP » conduira le dispositif à reformuler la requête sur le moteur Google Actualités, c'est-à-dire à restreindre le champ de recherche dans le but de le préciser et d'affiner les résultats. Un clic sur un des trois autres liens, ou sur l'image, conduira directement sur le site de l'éditeur. Nous voyons que seuls les éditeurs les mieux classés sur Google Actualités apparaissent sur les trois liens affichés par Google Search.

Pour la même requête, un autre transfert de requête était proposé, cette fois vers le moteur vertical « Google vidéo », avec, lui aussi, trois liens pointant directement vers les sites des éditeurs :

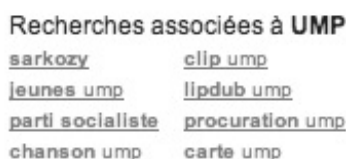
Figure 6. Capture d'écran : Transfert de requête vers Google vidéos
(Requête « UMP » - 20/11/2012)



4.4.4 Liens de reformulation

Toujours pour cette même requête, nous avons, tout en bas de la première page de résultats, des propositions de « recherche associée » :

Figure 7. Capture d'écran : Proposition de reformulation
(Requête « UMP » - 20/11/2012)



Ces requêtes permettent à l'internaute de préciser ce qu'il a voulu dire, en fonction de ce que Google lui indique être des « requêtes apparentés ». Une nouvelle requête sera alors générée par le moteur, susceptible de donner des résultats différents de ceux qui se sont affichés d'abord.

4.5 Google Actualités

Nous avons décrit l'association de l'action des éditeurs, et ce quels que soient le type de contenus qu'ils mettent en ligne, à l'action de Google, comme elle est rendue possible étant donné des caractéristiques inhérentes au web (pouvoir faire) et comme elle a été rendue possible par les concepteurs du moteur (pouvoir de faire). Les éléments présentés ci-avant concernent donc le moteur généraliste Google Search, ainsi que l'ensemble des moteurs verticaux. Cependant, chacun des moteurs verticaux est également doté de spécificités. C'est pourquoi nous procéderons

ci-après à une description des lignes d'action concernant en particulier Google Actualités.

4.5.1 Intégrer

Google définit un certain nombre d'exigences techniques et de préconisations éditoriales qui doivent être respectées par un éditeur désireux d'intégrer le champ de crawling de Google Actualités. Cependant, il est stipulé à l'éditeur qui respecterait ces recommandations qu'il ne sera pas pour autant assuré du résultat. Google se réserve ainsi le droit de choisir si, oui ou non, un site fera partie du champ de crawling, ainsi que le droit d'exclure dès qu'il le voudra un site autrefois inclus. Ce « pouvoir sur » peut constituer un « pouvoir de faire faire » dès lors que certains éditeurs, dans le but d'être inclus, ou par crainte d'être expulsés, se conforment aux recommandations techniques et éditoriales qui leur sont données. Ainsi, nous nous intéresserons en premier lieu ici aux règles que tous les sites de presse en ligne doivent théoriquement respecter pour être référencés sur Google Actualités.

A. Adresse URL

Les URL de chacune des pages doivent comporter au moins trois chiffres. Ce critère permet de rendre explicite la volonté de l'éditeur d'être indexé. En quelque sorte, il indique qu'il a pris connaissance des règles et « montre patte blanche »¹²⁰.

Par ailleurs, les URL doivent être uniques et permanentes afin que le système identifie quels sont les nouveaux articles, et quels sont ceux qui ont été mis à jour. Il est également mentionné sur le centre d'aide aux éditeurs d'actualités (CAA) que : « chaque article doit contenir un lien redirigeant vers une page exclusivement dédiée à l'article en question. »

B. Formats

Le crawler de Google Actualités est conçu uniquement pour le format HTML. Cela est dû à la légèreté du format, qui permet de scanner à une fréquence élevée sans surchauffer les serveurs. Ainsi, contrairement à Google Search, qui supporte la

¹²⁰ Cette règle ne concerne pas les sites qui ont opté pour un *Sitemap* pour Google Actualités, qui ont indiqué autrement leur volonté de collaborer (voir ci-après).

plupart des formats, Google Actualités demande aux éditeurs souhaitant y être référencés de se cantonner au seul HTML. La contrainte est plus grande à leur endroit.

Les liens « JavaScript », les liens graphiques et les liens figurant dans des cadres, ainsi que les menus déroulants, ne sont pas analysés, mais leur usage n'empêche pas l'indexation de la source. La raison donnée pour le JavaScript est intéressante : « Google Actualités ne prend pas en charge les articles intégrés dans JavaScript, car le contenu accessible aux utilisateurs (qui lisent le texte basé sur JavaScript) diffère parfois du contenu indexé par les moteurs de recherche (qui, eux, tiennent compte du texte basé sur script) » (CAA). Google empêche ainsi l'éditeur d'utiliser une technique qui lui permettrait de s'adresser au crawleur autrement qu'à l'internaute.

Pour les images, Google indique un certain nombre de consignes à respecter, comme une taille minimum (60x60 pixels) et une préférence pour le standard « .jpg ». Au milieu des critères techniques, la firme demande également aux éditeurs de veiller à utiliser « des images dont les proportions sont raisonnables » sans préciser ce que « proportions raisonnables » signifie.

« Nous ne pouvons pas garantir que nous serons en mesure d'explorer les images de votre site ni d'inclure toutes vos images. Toutefois, si vous respectez ces consignes, vous augmentez les chances de voir vos images apparaître sur Google Actualités » (CAA).

Le pouvoir de décision reste ainsi aux mains de Google, qui suggère aux éditeurs de suivre ces règles non pas pour figurer dans l'index au même titre que leurs concurrents, mais pour augmenter leurs chances d'y figurer.

Pour les contenus audio, Google indique qu'il n'est pas en mesure de les indexer sur Google Actualités. Il s'agit cette fois d'une contrainte technique due à une incapacité de traiter de tels contenus. Il faut donc que l'éditeur rédige un texte accompagnant ses contenus audio. Google donne comme exemple sur son CAA le site de *France Info* pour la radio, et le site *NPR.org* pour la vidéo.

Pour les vidéos, l'éditeur qui voudrait les indexer n'a pas d'autre choix que de créer un compte Youtube, et d'encaster les modules sur son site. Google utilise ici sa position dominante sur le marché des moteurs pour mettre en avant ses autres

services, en l'occurrence ici Youtube, support d'affichage publicitaire et générateur de revenus pour la firme.

Les cadres sont déconseillés sans être bannis, pour des raisons techniques, l'usage de cadre aboutissant à la présence de plusieurs URL dans une même page (une par cadre), ce qui empêche l'indexation de s'effectuer correctement. Idem pour la langue : Google Actualités est optimisé pour les sites ne comprenant qu'une seule langue, car la présence de plusieurs langues nuirait au fonctionnement optimal de l'indexeur.

Les forums et les espaces de commentaires, quant à eux, doivent être clairement indiqués par les balises méta-tags : /board/ ou /boards/ ; /forum/ ou /forums/ /messageboard ; /showthread ; ?threadid= ou &threaded. Google exclut donc de son champ de crawling les commentaires à propos de l'actualité. Ainsi, seule l'actualité produite sous forme d'article, par des journalistes ou sous leur contrôle, ou bien par des contributeurs officiels, est scannée.

C. Le cas des sites payants

A propos des sites fonctionnant par abonnement ou par inscription, Google explique : « En général, nos internautes (qui deviendront peut-être un jour vos visiteurs) préfèrent ne pas être redirigés vers une page d'inscription ou d'abonnement lorsqu'ils cliquent sur un lien dans Google Actualités » (CAA).

Nous voyons dans cette citation comment Google se fait le porte-parole des utilisateurs en représentant leurs intérêts auprès des éditeurs, tout en tirant la corde de *l'intéressement*. Il est en effet remémoré que ceux qui sont « déjà » les visiteurs de Google deviendront « peut-être un jour » ceux des éditeurs. En outre, l'apport de trafic apparaît ici comme un argument clé en faveur d'un accès gratuit à l'information, et donc d'un modèle économique qui fonctionnerait avant tout grâce à la publicité en ligne – marché sur lequel Google est un acteur dominant. Cette fois, la tentative de « pouvoir de faire faire » vise le modèle économique des éditeurs, plaidant pour du gratuit plutôt que du payant ou « *freemium* » — modèle dont, via AdSense et DoubleClick (cf. chap. 5), Google peut potentiellement bénéficier. Ce sont donc ses propres intérêts que la firme représente, en même temps que ceux des utilisateurs, et en tentant d'en faire les intérêts des éditeurs.

Pour les éditeurs qui tiendraient absolument à faire payer leurs visiteurs, Google leur propose d'accepter de mettre en place le « Premier clic gratuit » qui consiste, pour l'internaute qui proviendra de Google Actualités, à bénéficier d'un accès à l'article gratuitement et sans avoir à s'inscrire, tandis qu'il devra payer ou s'inscrire s'il veut consulter ensuite d'autres articles. « Ceci permet à nos utilisateurs, explique Google, de consulter l'article qui les intéresse, tout en leur présentant votre site et en les encourageant à s'y abonner » (CAA). Ceci permet également à Google de s'assurer que l'ensemble des articles agrégés sur Google Actualités soient gratuits, et donc d'offrir un service d'autant plus valorisé par les internautes qu'il ne leur coûtera rien. Cela permet enfin, une fois que l'internaute aura lu l'article, de l'inviter à revenir sur Google Actualités, où il pourra de nouveau avoir accès à des contenus gratuitement, pour lesquels il payerait s'il y accédait autrement que par l'intermédiaire de Google.

« Nous préférons cette solution car elle profite autant à nos éditeurs qu'à nos utilisateurs. Elle permet à Googlebot d'indexer entièrement votre contenu, ce qui augmente alors la probabilité selon laquelle les internautes viendront visiter votre site. Par ailleurs, vous pouvez limiter le nombre d'articles gratuits auxquels un lecteur de Google Actualités peut accéder via le premier clic gratuit. Toutefois, un utilisateur provenant du domaine [*.google.*] doit pouvoir consulter un minimum de cinq articles par jour, à défaut de quoi le site sera traité comme un site exigeant un abonnement et non comme un site de type premier clic gratuit. » (CAA)

Cette citation est intéressante dans la mesure où il s'agit d'une des seules fois où nous avons trouvé qu'il était fait mention par Google de « nos éditeurs » au même titre que « nos utilisateurs ». Il s'agit d'une tentative de *représentation*, l'adjectif « nos » marquant l'attachement, l'association, le lien, entre les éditeurs et la firme. D'autre part, nous retrouvons à nouveau une tentative d'intéressement/enrôlement lorsque Google prétend que l'utilisation du « Premier Clic Gratuit » augmentera pour l'éditeur la probabilité de recevoir un trafic substantiel sur ses pages.

Google explique que, si la solution du premier clic gratuit n'est pas envisageable, un logo « *subscription* » apparaîtra à côté du nom des publications provenant de sources qui accueillent les internautes par un formulaire d'inscription ou d'abonnement ; avant de préciser :

« Si vous préférez cette solution, affichez un extrait de votre article (80 mots au minimum). Il peut s'agir du résumé de l'article ou d'une citation. Nous n'autorisons pas le "cloaking". Cette pratique consiste à montrer à Googlebot la version complète d'un article tout en affichant la version d'inscription ou d'abonnement pour les internautes. Nous explorons et affichons votre contenu sur la seule base des extraits d'articles que vous nous fournissez. [...] REMARQUE : Si vous dissimulez du contenu à Googlebot, votre site risque d'être sanctionné pour non-respect des consignes aux webmasters. » (CAA)

Google suggère enfin aux éditeurs dont certains contenus sont payants, mais pas tous (modèle *freemium*), d'extraire les contenus payants de l'index de Google Actualités grâce au protocole « robots.txt » ou aux méta-tags appropriés. Comme l'avait déjà souligné Nikos Smyrnaiois dans sa thèse (Smyrnaiois, 2005, p. 302-303), cette manière qu'a Google de privilégier les modèles économiques gratuits peut être considérée par certains éditeurs comme une tentative d'ingérence vis-à-vis de leur stratégie, et donner lieu à des frictions, notamment si les éditeurs dont les contenus sont payants mais qui font autorité refusent de se plier aux règles édictées par Google. En punissant les éditeurs, la firme userait de son « pouvoir sur », étant donné l'échec de son « pouvoir de faire faire », mais risquerait alors de dégrader la qualité de son service aux yeux des internautes, dont certains cherchent un contenu de qualité pour lequel ils peuvent éventuellement être prêts à payer.

D. Formulaire

Une fois qu'un site respecte les règles édictées par Google, l'éditeur doit en demander l'inclusion via un formulaire en ligne. Il répond d'abord par oui ou par non à un ensemble de questions. Dès lors qu'une réponse est jugée incorrecte par Google, l'éditeur ne peut pas poursuivre sa demande. Nous citerons ici certaines de ces questions et ce qui apparaît une fois que l'une ou l'autre des réponses a été cochée.

Google demande si les pages respectent bien les règles édictées et si elles comportent du contenu relatant des événements et des points de vue récents. Si nous répondons « non » à cette question, nous obtenons :

« Les sites figurant dans Google Actualités doivent proposer en temps opportun des articles relatifs à des sujets importants ou susceptibles d'intéresser nos utilisateurs. En règle générale, nous n'incluons pas les didacticiels, les éditoriaux, les offres d'emploi, ni les contenus exclusivement informatifs (prévisions météo, informations boursières, etc.) »

Il est intéressant de constater ici que les éditoriaux, qui sont pourtant des points de vue, et qui sont parmi les formats classiques de la production journalistique, sont exclus et placés entre les didacticiels et les offres d'emploi. Il faut également que le sujet soit « important ». Google demande ensuite si le site comporte des contenus originaux, rédigés et publiés par l'organisation. Si nous répondons « non » à cette question, nous obtenons :

« Les reportages originaux et la citation des sources font depuis longtemps partie de l'éthique journalistique. Pour le moment, Google Actualités n'accepte pas les sites principalement axés autour de l'agrégation de contenus déjà publiés par ailleurs. »

Il est question ici de « l'éthique des journalistes », dont le respect consisterait à produire des reportages originaux et à citer ses sources. La firme sous-entend ainsi qu'un jugement pourra être apporté par un de ses employés quant à l'éthique des rédacteurs employés par un éditeur.

Enfin, Google demande à l'éditeur si les sections qu'il souhaite faire apparaître dans Google Actualités sont « exemptes de toute publicité, annonce ou autre offre de produit ». Si nous répondons « non » à cette question, nous obtenons :

« Google Actualités n'est pas une solution de marketing. Nous ne voulons pas rediriger les internautes vers des sites principalement consacrés à la promotion d'un produit ou d'une entreprise. Si ce type de contenu figure sur votre site aux côtés des actualités, Google Actualités ne prendra probablement pas en compte votre site. »

Voici un critère qui est davantage éditorial que technique, et qui fera l'objet d'un jugement par ceux qui sont chargés de décider chez Google (manuellement) de l'intégration d'un site au champ de crawling de Google Actualités. Ce critère empêche théoriquement les éditeurs indexés de revendre à des annonceurs, sous forme publi-rédactionnelle, c'est-à-dire en leur proposant de créer des documents consacrés exclusivement à leurs produits, la visibilité qu'ils ont grâce à leur présence dans le champ du moteur vertical.

*

Nous avons énuméré les différentes règles à respecter pour être éligible à l'indexation sur Google Actualités. Nous avons également vu comment la firme Google, grâce au « pouvoir sur » dont elle jouit étant donné le choix qu'elle peut faire d'accorder ou non la « grâce » de l'indexation, pouvait exercer un « pouvoir de faire faire », notamment en faisant la promotion d'un modèle de gratuité plutôt qu'un modèle payant. Il y a ici une forme de censure possible de la part du diffuseur, dès lors qu'il peut choisir les sources dont proviendront les contenus vers lesquels il générera des hyperliens. Dans le cas d'un refus, l'éditeur n'aura aucun recours lui permettant d'exiger que ses contenus soient indexés sur Google Actualités, cela car il n'y a pas d'équivalent de la Loi Bichet pour le web (cf. 3.2.1).

4.5.2 Communiquer

Nous proposons à présent de considérer comment, et à quel point, une fois indexés, les éditeurs peuvent faire en sorte d'établir une communication entre leurs documents et les crawlers de Google Actualités, et dans quelles mesures ils peuvent avoir la main sur la publicisation par ce biais, étant donné leur champ de manœuvres.

A. Sitemap Actualités

Les sites des éditeurs de presse en ligne ont leur propre *Sitemap*, spécialisé, dont la principale différence avec le *Sitemap* généraliste est que celui-ci est destiné à prendre en compte les particularités de leur activité, notamment le traitement en temps réel. Comme avec le *Sitemap* généraliste, l'idée est d'établir une communication entre le contenu et le dispositif dans le but de maximiser l'apport de trafic et de maîtriser les modalités de publicisation. Le spécialiste du référencement Andrew Shotland explique à son sujet : « Je suis toujours surpris quand je découvre qu'un site de presse en ligne n'utilise pas [le *Sitemap* Actualités]. J'ai constaté de grandes différences en termes de crawling, d'indexation et de classement, dès lors que ce *Sitemap* était implémenté. Ce n'est pas particulièrement difficile à mettre en œuvre. Qu'attendez-vous ? » : »¹²¹

¹²¹ « I am always surprised when I find a news site not using these. I have seen dramatic differences in crawling, indexing and rankings as a result of using a Google Actualités Sitemap. It's not particularly hard to implement. What are you waiting

Pour implémenter le Sitemap, l'éditeur doit renseigner les champs suivants :

Figure 8. Sitemap Actualités : exemple

```
<url>
<loc>http://www.example.org/business/article55.html</loc>
<news:news>
  <news:publication>
    <news:name>The Example Times</news:name>
    <news:language>en</news:language>
  </news:publication>
  <news:access>Subscription</news:access>
  <news:genres>PressRelease, Blog</news:genres>
  <news:publication_date>2008-12-23</news:publication_date>
  <news:title>Companies A, B in Merger Talks</news:title>
  <news:keywords>business, merger, acquisition, A, B</news:keywords>
  <news:stock_tickers>NASDAQ:A, NASDAQ:B</news:stock_tickers>
</news:news>
</url>
</urlset>
```

En plus de demander le titre, la langue et la date de publication, Google demande que soient renseignés les modalités d'accès (inscription ou abonnement, ce qui est le cas ici) et le genre. L'éditeur indique pour chaque URL, s'il s'agit d'un communiqué de presse (PressRelease), d'un article annonçant une opinion (OpEd ou Opinion), d'un contenu amateur (UserGenerated), d'un contenu satirique (Satire) ou d'un blog (Blog). Dans l'exemple ci-dessus, deux genres sont renseignés : communiqué de presse et blog. Si aucun de ces termes n'était renseigné, Google considérerait qu'il s'agit d'un article qui n'est ni une opinion ni une satire, produit par un journaliste, et qu'il ne s'agit pas d'un blog. Il s'agit d'une politique éditoriale, ou méta-éditoriale, consistant à définir un rubriquage *a priori* susceptible de se substituer au rubriquage de l'éditeur. Cette politique se base sur des critères relevant de l'avis des dirigeants de Google et des ingénieurs¹²².

for? » Disponible sur : <http://searchengineland.com/top-10-negative-google-news-ranking-factors-95012> (Notre traduction, dernière visite le 21 août 2013)

¹²² Au passage, il est légitime de se poser la question de ce que peut être l'impact de cette balise sur le référencement. Un communiqué de presse, ou un blog, une opinion, une satire, seront-ils moins visibles sur le dispositif qu'un article factuel ? (Nous n'avons hélas pas la réponse à cette question.)

La balise `<news:keywords>` est principalement adressée à l'indexeur. Elle permet d'optimiser la pertinence des résultats dans les cas où ces mots clefs apparaîtraient dans la requête. Les éditeurs peuvent consulter « Tendances de Recherche » pour choisir les mots qu'ils renseigneront. Enfin, la balise « `stock_tickers` » concerne en particulier les entités financières citées dans l'article ou simplement reliées au sujet. Elle aide Google à pouvoir faire le lien lorsque les entités sont cotées en bourse avec le flux d'informations Google Finance. La firme peut ainsi se servir de l'implication des éditeurs dans le dispositif Google Actualités pour optimiser le fonctionnement de Google Finance.

B. Méta-tags spécialisés

De nombreux éditeurs reprochent à Google Search et Google Actualités de ne pas faire remonter leur contenu quand il s'agit d'une information exclusive et originale, notamment parce que les documents les plus récents remontent en premier dans les grappes de liens. Si vous êtes le premier à avoir traité une information et que d'autres l'ont traité ensuite, vous risquez ainsi d'être « au fond » du cluster, et donc de recevoir peu de visiteurs en provenance de Google. Pour tenter de remédier à cette imperfection, Google a mis au point deux méta-tags permettant d'identifier la source originale d'un document. Le premier – « *syndication source meta-tag* »¹²³ – permet à un éditeur syndiquant un contenu qu'il n'a pas produit d'indiquer au crawleur où le contenu original se trouve. Ainsi, Google peut considérer ce tag pour faire remonter le site du producteur de l'article de façon prioritaire par rapport à ceux qui l'ont repris, sous forme de « *cross-posting* » ou d'agrégation. Bien sûr, on peut se demander dans quelle mesure un éditeur a intérêt à implémenter ce méta-tag étant donné qu'il risque de se trouver lui-même relégué au fond de la grappe.

Le méta-tag « *standout* »¹²⁴, quant à lui, qui a remplacé en septembre 2011 le méta-tag « *original source* », permet à un éditeur d'indiquer au crawleur que c'est lui l'auteur original du document et qu'il considère que ce contenu est important.

¹²³ `<meta name="syndication-source" content="http://www.example.com/article-original-pouvant-etre-publie-ailleurs.html">`

¹²⁴ `<link rel="standout" href="http://www.example.com/scoop_article_2.html" />`

« Si vous mettez ce tag dans l'entête HTML d'un de vos articles, Google Actualités pourra montrer l'article accompagné du label "original" sur la page d'accueil de Google Actualités ou sur les requêtes qui y seront effectuées » (Google News Blog)¹²⁵.

Le *standout* vise directement à attribuer un label aux journalistes, visible sur Google Actualités et Google Search : ce fameux signe de reconnaissance dont parle Dominique Wolton comme d'un « impératif catégorique » pour identifier « l'information-presse par rapport à toutes les autres » (Wolton, 2003, p.14, cf. 1.1.2). Voici comment Google présente ses préconisations :

« Si votre rédaction traite un scoop, ou publie un travail de journaliste hors du commun, vous pouvez l'indiquer en utilisant le tag *standout*. Quand il s'agira de déterminer s'il est judicieux ou non d'utiliser ce tag pour votre article, considérez les critères suivants : / Votre article est la source originale. / Votre entreprise a investi des ressources significatives dans la production et le report de cette information. / Votre article mérite une reconnaissance spéciale. » (CAA)¹²⁶

Google propose ainsi à l'éditeur d'être lui-même le juge de la qualité de son travail. L'usage de cette balise ne suffit cependant pas à ce qu'un contenu en particulier se trouve en haut du cluster. D'autres critères algorithmiques sont à l'œuvre¹²⁷. Mais son utilisation permet à Google d'avoir une indication supplémentaire quant à la qualité de l'article en question, et à l'éditeur de reprendre partiellement la main sur la hiérarchisation des contenus opérée par le dispositif. C'est une manière pour les éditeurs de contribuer à restructurer en partie leur ligne éditoriale défaite par les « liens profonds » et le rangement en « grappes ».

Dès lors, il paraît évident que la tentation est grande pour les éditeurs d'insérer cette balise dans le code de chaque nouvelle page mise en ligne. C'est pourquoi

¹²⁵ « If you put the tag in the HTML header of one of your articles, Google News may show the article with a 'Featured' label on the Google News homepage and News Search results », disponible ici : <http://googlenewsblog.blogspot.com/2011/09/recognizing-publishers-standout-content.html> (Notre traduction)

¹²⁶ « If your news organization breaks a big story, or publishes an extraordinary work of journalism, you can indicate this by using the standout tag. When determining whether to use this tag for your own article, consider whether that article meets the following criteria: / Your article is an original source for the story. / Your organization invested significant resources in reporting or producing the article. / The article deserves special recognition. », disponible ici : <http://support.google.com/news/publisher/bin/answer.py?hl=en&answer=191283> (Notre traduction)

¹²⁷ « Standout tags are just one signal among the many signals that algorithmically determine prominence on Google News. », disponible ici : <http://googlenewsblog.blogspot.com/2011/09/recognizing-publishers-standout-content.html>

Google limite l'usage du méta-tag à sept articles par semaine¹²⁸. Sept est un chiffre arbitraire, donné sans véritable explication. Google exerce ici son « pouvoir sur » afin que les éditeurs n'abusent pas du « pouvoir de faire » qui leur a été laissé. Cette limite lui permet de s'assurer qu'il n'y ait pas *trahison*, mais toujours *traduction* du côté des éditeurs. La limitation permet en effet d'imposer aux éditeurs d'utiliser le méta-tag comme cela a été prévu par les concepteurs du dispositif.

Le même méta-tag peut également être utilisé dans le code HTML d'un article n°1 pour désigner un article n°2, original et à forte valeur ajoutée, présent soit sur le site de l'éditeur de l'article n°1 soit sur le site d'un autre éditeur. Ainsi, non seulement le journaliste cite ses sources via les liens hypertextes, mais en plus il peut signaler au robot quelles sont les sources qui, selon ses critères à lui, sont des travaux originaux à forte valeur ajoutée. Les porte-parole de Google encouragent fortement cette pratique de désignation¹²⁹, qui n'intervient pas dans la limite des sept articles par semaine (mais qui, en revanche, peut diluer le PageRank de l'éditeur n°1 s'il cite le document d'un éditeur n°2).

Le méta-tag « *standout* » constitue un excellent exemple de la tentative d'une traduction par les employés de Google d'un principe déontologique propres aux journalistes, à partir de ce que les employés de Google pensent à leur sujet. En effet, les porte-parole de Google donnent des recommandations aux journalistes concernant leur honnêteté professionnelle, en plaidant pour un enrôlement qui irait de pair avec une certaine forme d'éthique dès lors qu'ils s'appuieraient sur des sources ayant fourni l'essentiel du travail de collecte d'information¹³⁰. Par ailleurs, il est important de souligner que la balise *standout* n'est pas simplement un outil visant à servir

¹²⁸ « At this point, we ask news organizations to use the Standout tag to cite their own content at most seven times in each calendar week. If a site exceeds that limit, it may find that its tags are less recognized, or ignored altogether. », disponible ici : <http://googlenewsblog.blogspot.com/2011/09/recognizing-publishers-standout-content.html> (dernière visite le 21 août 2013)

¹²⁹ « Standout Content tags work best when news publishers recognize not just their own quality content, but also the original journalistic contributions of others when your stories draw from the standout efforts of other publications. Linking out to other sites is well recognized as a best practice on the web, and we believe that citing others' standout content is important for earning trust as you also promote your own standout work. » Google News Blog, <http://googlenewsblog.blogspot.com/2011/09/recognizing-publishers-standout-content.html> (dernière visite le 21 août 2013). « We strongly recommend citing standout articles from other publishers when your own article draws from that standout piece of journalism. When determining whether to use this tag to cite the work of others, consider the following criteria: / The publisher's article was the original source for the story you are now reporting. / The original source invested significant resources in reporting or producing the article. / You know that the original article deserves special recognition. If a piece draws on multiple pieces of standout work, you can use the standout tag multiple times in the header of the article. Also, citing standout articles from other publishers does not count against the limit of seven self-citations per week. » (CAA, <http://support.google.com/news/publisher/bin/answer.py?hl=en&answer=191283>, dernière visite le 21 août 2013)

¹³⁰ Voir la note de bas de page précédente.

l'intérêt des journalistes, cela car une utilisation du tag conforme à celle qui a été prévue permettra potentiellement à Google d'identifier plus facilement les contenus « de qualité », non plus selon des critères techniques et un procédé automatisé mais d'après une indication manuelle effectuée du côté des éditeurs. Google a donc intérêt à ce que le *standout* soit massivement utilisé, pour ensuite accroître la qualité du fonctionnement du dispositif qu'il a conçu. Le « pouvoir de faire » laissé aux journalistes permettrait ainsi à Google d'inclure une expertise journalistique à son mécanisme automatisé de hiérarchisation.

4.5.3 Optimiser

A présent que nous avons mentionné les mises en œuvre qui permettaient aux éditeurs d'être éligibles à l'indexation, puis les mises en œuvre visant à communiquer avec les crawlers et l'indexeur, il nous reste à considérer comment, techniquement, l'éditeur peut tenter de maximiser le trafic apporté par le dispositif.

A. *SourceRank*

En plus du PageRank, un des critères considérés comme les plus importants par les référenceurs dans le cas de Google Actualités est appelé « *SourceRank* ». Il s'agit d'attribuer une note de pertinence non pas seulement au document mais également au site lui-même. Nikos Smyrnaio et Franck Rebillard ont présenté le SourceRank à partir d'une étude des brevets déposés en 2005. Ils classent dans leur analyse les treize critères proposés par Google en deux catégories, dont la première est centrée sur la source et la deuxième sur la production. Dans le premier cas, expliquent-ils, « l'objectif semble être la traduction en langage algorithmique du "capital journalistique" d'un média, "correspondant à une reconnaissance qui s'obtient auprès de ses pairs, mais aussi auprès du public et de pouvoirs externes" (Duval, 2004, p.107) » (Smyrnaio et Rebillard, 2009, p.102). Nous retrouvons dans cette citation l'idée d'une *traduction* (interprétation et déplacement) opérée par le langage algorithmique de certains éléments comme le capital journalistique. Dans cette première catégorie, Nikos Smyrnaio et Franck Rebillard énumèrent des critères comme la prise en compte du nombre de liens entrant (PageRank), le trafic global, et aussi la taille de la rédaction, le nombre de bureaux dont elle dispose à l'étranger, les

prix reçus par les journalistes et l'opinion des internautes (qui peut être mesurée par sondage). Ils précisent toutefois qu'à part le PageRank, rien n'indique pour le moment que les autres critères cités dans le brevet soient effectivement pris en compte par Google Actualités.

Les deux chercheurs expliquent ensuite que la deuxième catégorie de critères, centrés sur la production, semble en revanche avoir été privilégiée par Google. Dans cette catégorie, on trouve le nombre d'articles produits par la source (lié à la quantité davantage qu'à la qualité), le respect des règles d'orthographe et de grammaire, le *story size* et le *breaking score*. Le *story size* « consiste à évaluer l'importance d'un sujet en fonction du nombre de sources qui le traitent » et « tend à privilégier les sujets d'actualité qui font l'objet d'un traitement médiatique par un grand nombre de sources au détriment des informations peu reprises » (*op. cit.*, p. 104), ce que d'ailleurs Franck Rebillard et Nikos Smyrnaio, avec d'autres, ont confirmé empiriquement trois ans plus tard (Marty *et al.*, 2012). Le *breaking score*, quant à lui, « représente la capacité d'une source à produire un article peu de temps après que l'événement traité s'est produit. Autrement dit, plus l'apparition d'une information est rapide sur un site (en comparaison des autres sites), mieux celui-ci sera noté. Il s'agit là d'évaluer la réactivité d'une rédaction en fonction des événements. La satisfaction de ce critère implique une activité de veille considérable pour les journalistes et une organisation du travail permettant de diffuser un contenu extrêmement rapidement. *In fine* ce critère tend à privilégier la source qui se trouve à l'origine d'une information puisque, par définition, le moment de la première publication sur un sujet donné constitue la référence à partir de laquelle est calculé le *breaking score* des articles suivants » (Smyrnaio et Rebillard, *op. cit.*, p. 104). Un autre critère centré sur la production est lié à la diversité des thèmes abordés par une même source, et « tend à privilégier les sources généralistes au détriment des sites d'information spécialisée » (*ibid.*). Quant au dernier critère, appelé *reporting*, il concerne l'originalité de la production. Celui-ci est mesuré par exemple grâce au nombre de noms propres, qui indiquerait que l'article apporte des informations supplémentaire à un sujet (*ibid.*). Nous verrons que c'est à cause de ce critère, couplé avec le *breaking score*, que les éditeurs ne se contentent pas de reprendre les

dépêches AFP, mais les « bâtonnent » dans le but de « faire croire » au dispositif qu'ils ont produit un contenu original (cf. 9.1.6).

A l'analyse du SourceRank effectuée par Nikos Smyrnaio et Franck Rebillard, nous pouvons ajouter plusieurs critères qui entrent selon les spécialistes, et notamment selon une réflexion menée par une dizaine d'entre eux en 2011¹³¹, dans le calcul de la pertinence de la source et des documents :

- L'autorité du site sur un sujet précis : Google Actualités doit identifier un site comme spécialiste d'un domaine en particulier, ce qui tendrait à désavantager les sites de presse trop généralistes.
- La présence de mots-clés dans le titre de la page, critère qui, comme pour Google Search, est primordial pour Google Actualités.
- Les signaux sociaux : comme pour Google Search, le nombre de fois qu'un article a été partagé et commenté sur les réseaux sociaux influence son classement dans Google Actualités.
- Le taux de clics dans Google Actualités ainsi que dans Google Search : plus un site reçoit de visiteurs lorsqu'un de ses liens s'affiche, et plus ses liens s'afficheront, entraînant alors une spirale au sein de laquelle « les plus riches deviennent plus riches ». Par conséquent, un site « neuf », dont l'historique du taux de clic est court, aurait, selon les spécialistes, très peu de chance de recevoir un trafic significatif *via* Google Actualités.
- L'utilisation d'un Sitemap : l'utilisation de cet outil permet d'augmenter significativement l'apport d'audience en provenance du dispositif.

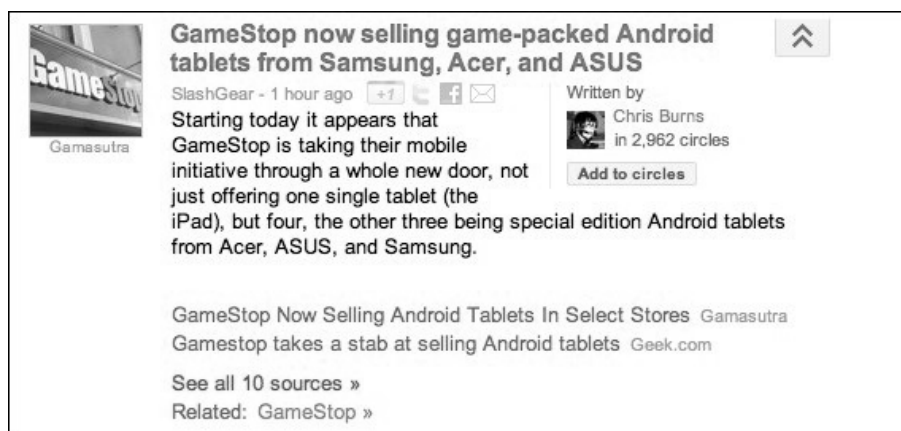
B. AuthorRank spécifique

Le dispositif mis en place pour la reconnaissance des auteurs est semblable à celui de Google Search, à quelques détails près. Le 2 novembre 2011, voici comment Eric Weigle, ingénieur chez Google, en présentait l'intérêt : « Le bon journalisme tient à davantage d'éléments qu'aux faits et aux chiffres — il nécessite les compétences de journalistes qui sauront les tricoter ensemble, et ainsi raconter une

¹³¹ Tous les résultats de cette étude sont disponibles ici : <http://googlenewsrankingfactors.com/> (dernière visite le 22 août 2013)

histoire. Savoir qui a écrit un article peut ainsi aider les lecteurs à comprendre quel est le contexte de l'article et en supposer la qualité, accéder à d'autres articles de ce même auteur, et même interagir directement avec lui. »¹³². L'ingénieur explique ici ce qu'est selon lui le « bon journalisme » et effectue une montée en généralité afin de justifier ses choix tout en invitant les journalistes à se saisir d'un « pouvoir de faire » mis à leur disposition, et en œuvrant aux intérêts du réseau social Google+. Les journalistes auront ainsi leur propre réputation vis-à-vis de l'algorithme de Google Actualités, et une visibilité qui ne sera pas la même selon que cette réputation est forte ou faible. A notre connaissance, les critères de mesure de l'AuthorRank sont les mêmes que ceux de l'AuthorRank de Google Search, déjà mentionnés. L'affichage de l'auteur, quant à lui, diffère quelque peu :

Figure 9. Capture d'écran : Author Rank Google Actualités



On retrouve cette fois, un seul lien vers Google+. Le nom de la source – ici, *SlashGear* – est écrit en toutes lettres, sans être mêlé à l'adresse URL ni véritablement mis en retrait par rapport au nom de l'auteur, comme c'est le cas sur Google Search.

¹³² « Great journalism takes more than facts and figures -- it takes skilled reporters to knit together compelling stories. Knowing who wrote an article can help readers understand the article's context and quality, see more articles by that person, and even interact directly with them », disponible ici : <http://googlenewsblog.blogspot.com/2011/11/highlighting-journalists-on-google-news.html> (Notre traduction, dernière visite le 22 août 2013)

C. Suggestions

Les éditeurs et les internautes sont invités sur une page dédiée¹³³ à donner leur avis à propos des améliorations possibles. Le système fonctionne comme une sorte de pétition, où Google suggère des changements pour lesquels les éditeurs et les internautes peuvent voter sans savoir jamais combien d'autres personnes ont voté ni les conditions requises pour qu'une implémentation soit réellement envisagée. Par ailleurs, une suggestion libre peut être effectuée sans qu'on puisse savoir non plus ce qui en sera fait : un message de remerciement apparaît simplement sur l'écran une fois la suggestion effectuée.

4.5.4 Préconisations éditoriales

Il est clairement stipulé par Google que le champ d'indexation de Google Actualités n'inclura pas de « sites qui ne disposent pas d'un processus formel de révision éditoriale » (CAA). En outre, un certain nombre de conseils sont donnés au sujet de la politique éditoriale considérée comme « idéale » par les concepteurs du dispositif.

A. Longueur du texte

Google invite ceux qui voudraient recevoir un trafic significatif à écrire un article qui ne soit ni trop long ni trop court, sans expliquer ce que « trop long » ou « nombre de mots trop faibles » signifient. Lorsqu'un article est jugé « trop long pour être considéré comme un article d'information, il est possible que [le] robot d'exploration ne le reconnaisse pas en tant que tel. Cela peut se produire lorsque, sous l'article, s'ajoutent des commentaires envoyés par les utilisateurs, ou lorsque des mises en page HTML contiennent des éléments autres que l'article en lui-même » (CAA). Google demande également de s'assurer « que le contenu des articles est composé de phrases et que le nombre de balises intégrées dans les paragraphes reste raisonnable », toujours sans préciser ce que « raisonnable » signifie.

¹³³ <https://support.google.com/news/suggestions/21796?hl=fr> (dernière visite le 28 août 2013)

B. Titre

Voici les consignes que Google donne pour le titre :

- « Placez votre titre bien en vue au-dessus du corps de l'article, par exemple dans une balise <h1>.
- Veillez à ce que le titre de la page contenant l'article (balise HTML <title>) soit le même que celui de votre article (balise <h1> ou une balise équivalente).
- Veillez à ce que le texte d'ancrage pointant vers votre article sur vos pages corresponde au titre de l'article et à celui de la page.
- Sur la page contenant votre article, évitez d'utiliser le titre ou le sous-titre de l'article comme lien hypertexte.
- Le titre de votre article ne doit pas comporter de date ni d'heure.
- Assurez-vous que le titre de votre article comprenne au moins dix caractères et entre 2 et 22 mots.
- Créez et envoyez un Sitemap pour Google Actualités.
- Pour garantir un affichage correct du titre de votre article sur des mobiles, n'incluez pas de préfixe (qui peut parfois correspondre à une clé d'accès) dans le texte d'ancrage. » (CAA)

Nous verrons dans la Partie 3 que les considérations liées au titre sont très présentes chez les éditeurs de presse en ligne. En effet, le titre est souvent considéré comme le principal moyen d'optimiser le référencement et peut poser problème aux journalistes, parmi lesquels certains estiment leur créativité bridée. Nous verrons également que ces contraintes donnent lieu à certaines innovations, lesquelles, ouvrant des *lignes de fuites*, visent à retrouver la liberté de ton perdue.

Sur le CAA, on trouve également ce paragraphe : « Gardez cependant à l'esprit que, si les suggestions ci-dessus augmentent les chances d'extraction de titres corrects à partir des articles, nous ne pouvons garantir l'exploration des articles de votre site, ni l'inclusion de tous vos articles » (CAA). Encore une fois, l'éditeur apprend qu'il ne peut qu'augmenter ses chances de figurer dans le champ, mais que cela n'oblige en rien Google à lui garantir quoi que ce soit.

C. Date

« Si notre robot d'exploration rencontre des difficultés pour trouver la date d'un article ou si la date détectée par notre robot indique que l'article date de plus de deux jours, il se peut que ce dernier ne soit pas indexé dans Google Actualités » (CAA).

Ainsi, l'actualité selon Google est datée de moins de deux jours. Cela exclue *a priori* les traitements de types magazine hebdomadaire ou mensuel. Google opte de cette façon pour un temps court et un traitement en continu. Mais ce n'est pas obligatoire non plus, car, dans la citation ci-avant, « il se peut que ce dernier ne soit pas indexé » ne signifie pas que ce dernier « ne sera pas indexé du tout ».

D. Contenu idéal

Certaines préconisations signalent aux éditeurs quel genre de contenus Google souhaite indexer idéalement, et, ce, en dehors de toute justification liée aux techniques de crawling, d'indexation et d'appariement. Lorsque Google donne ces conseils, c'est en tant que porte-parole des internautes, et étant donné la représentation que ses dirigeants et ses employés se font de ce que les internautes estimeront être des contenus « de qualité ».

« Rédigez des articles sur les sujets que vous maîtrisez. Les meilleurs sites d'actualités font preuve d'une compétence et d'une expertise indéniables » (CAA)¹³⁴. Cette recommandation n'a pas de matérialité technique véritable. C'est une vision du journaliste en tant que spécialiste, qui pourrait faire l'objet d'un débat, dans la mesure où de nombreux journalistes sont des généralistes dont le métier est d'interroger des spécialistes à propos de divers sujets, plutôt que véritablement des spécialistes.

Google mentionne également la langue comme critère :

« Des articles rédigés de façon claire, et irréprochables sur le plan de l'orthographe et de la grammaire, contribuent également à améliorer l'expérience utilisateur » (CAA)¹³⁵.

Là encore, un site qui serait identifié comme employant des auteurs dont la langue est mal maîtrisée — la syntaxe, l'orthographe — pourrait ne pas être indexé. Cela rejoint le critère de qualité orthographique identifié par Nikos Smyrnaio et Frank Rebillard (2009) dans le brevet relatif au SourceRank.

¹³⁴ <http://support.google.com/news/publisher/bin/answer.py?hl=fr&answer=40787> (dernière visite le 21 août 2013)

¹³⁵ *Ibid.*

4.5.5 Choix des rédactions

Lancé par Google en octobre 2011, le « Choix des rédactions » permet à un éditeur dont le contenu est indexé sur Google Actualités de mettre en avant cinq liens pointant vers son propre site, dans un ordre hiérarchique défini par lui. Ces liens seront repris sous le logo de l'éditeur dans un encart à droite de la page d'accueil de Google Actualités, assortis du nom des auteurs des pages pointés. Seuls certains éditeurs, parmi ceux qui sont intégrés au champ d'indexation, ont accès à ce dispositif. Le choix est fait manuellement par les employés de Google, après que les éditeurs leur ont soumis un formulaire.

Figure 10. Capture d'écran : Le choix des rédactions



Ce dispositif permet notamment aux éditeurs de faire remonter des contenus qui, étant donné leurs caractéristiques techniques et éditoriales, n'auraient pas forcément été mis en avant par le dispositif : une vidéo par exemple. La firme limite le nombre de liens à cinq, afin de circonscrire le « pouvoir de faire » attribué aux éditeurs.

Les critères que Google demande aux éditeurs de respecter sont :

- Pointer vers des contenus d'actualité (pas de liens vers une offre d'abonnement, prévisions météo, petites annonces, indices boursiers...)
- Logo sur arrière-plan transparent, format .png, sans espaces autour, respectant les critères de tailles de l'option 1 ou 2 :
 - Option 1 : largeur de 250 pixels, hauteur comprise entre 20 et 40 pixels
 - Option 2 : hauteur de 40 pixels, largeur comprise entre 125 et 250 pixels
- Titre de moins de 75 caractères
- Obligation d'associer chaque article à un nom d'auteur. (S'il n'y a pas d'auteur, les éditeurs peuvent signaler « AFP » ou « L'Equipe », mais ils doivent impérativement renseigner ce champ.)
- Intégrer au moins un nouvel article toutes les 48 heures, même si l'idéal, selon Google, « serait que le flux soit actualisé une ou deux fois par jour »¹³⁶. On peut penser, dès lors, qu'en respectant cet « idéal », l'éditeur augmente ses chances de maximiser son audience.
- Mettre toujours au minimum trois articles.

Ces deux dernières conditions n'ont pas de justification technique. Elles peuvent être attribuées à la politique éditoriale de Google. Par ailleurs, une fois encore, Google se réserve le droit de désindexer un flux RSS :

« Si l'examen du flux RSS révèle que nos consignes n'ont pas été respectées, nous désactiverons le flux dans notre système. Sachez également que nous pouvons, à tout moment, réexaminer ou désactiver automatiquement votre flux RSS, sur la base d'indicateurs de qualité définis par nos soins ou de commentaires provenant des utilisateurs » (CAW).

Les indicateurs de qualité cités ici ne sont détaillés nulle part à notre connaissance.

Dans le communiqué qui annonçait le lancement du dispositif en France, le porte-parole de Google Madhav Chinnappa, Directeur des partenariats stratégiques en

¹³⁶ <http://support.google.com/news/publisher/bin/answer.py?hl=fr&answer=1407682> (dernière visite le 21 août 2013)

Europe, expliquait : « Le choix des rédactions offre aux éditeurs un espace pour rassembler le meilleur du journalisme traditionnel et du journalisme en ligne, en valorisant des articles de fond tout en expérimentant de nouveaux formats. » Il semble ainsi que le dispositif ait été conçu en partie dans l'espoir de contenter ceux qui jusqu'alors se plaignaient du manque de diversité sur Google Actualités, pointant du doigt l'absence de traitements et de sujets originaux, ce qui a d'ailleurs été démontré à l'issue du projet de recherche *Ipri* (Marty *et al.*, 2012).

Même si les éditeurs peuvent choisir ce qu'ils mettent en ligne sans passer par la moulinette algorithmique, un calcul de pertinence a tout de même lieu. Lorsqu'un internaute se rend sur Google Actualités, une seule source s'affichera dans l'encart « Choix des rédactions », déterminée en fonction de critères tels que l'autorité de la source et l'historique de navigation. Si l'internaute rafraîchit la page, la source changera. Ainsi, même si les éditeurs ont la main sur la hiérarchisation dans cet encart, ils ne sont pas assurés d'y apparaître. Nous voyons également dans notre capture (Fig. 11) que Google propose à l'internaute, grâce à un curseur, de renseigner la fréquence à laquelle il souhaite voir apparaître du contenu en provenance de cette source en particulier. Le dispositif redonne ainsi la possibilité à l'internaute de choisir la source et de *déléguer* la fonction « choix du sujet » à des journalistes, et non plus à un algorithme. En cela, « Le choix des rédactions » est clairement un revirement par rapport à la reconfiguration des voies d'accès à l'origine de la création de Google Actualités (cf. Introduction).

A propos de ce dispositif, il est d'ailleurs dit sur le CAA :

« Le choix des rédactions permet aux lecteurs de suivre les actualités issues de leurs sources favorites et de découvrir d'autres sources intéressantes jusqu'alors inconnues. Les internautes peuvent facilement indiquer leurs préférences (sources à privilégier ou, au contraire, à éviter) et nouer ainsi des relations plus directes avec les éditeurs » (CAA).

Google présente ainsi cette innovation comme une tentative d'effacement partiel de son infomédiation.

Le « Choix des rédactions » est un dispositif encore trop jeune pour présumer de ce qu'il deviendra, mais nous pouvons clairement y voir une tentative de pallier certains défauts de Google Actualités. Il s'agit en particulier d'une manière de rendre

partiellement la main aux éditeurs sur la hiérarchisation des articles et le choix des sujets, tout en donnant partiellement la main aux internautes sur le choix des sources. Le pouvoir performatif de l'éditeur et de l'internaute est accru, dès lors que ce sont eux qui « prennent la parole » et suscitent (performent) la réalité qu'ils énoncent. Par ailleurs, il est évident que la firme a elle aussi un intérêt dans la mise en place de ce dispositif : en lui indiquant quels contenus ils souhaitent mettre en avant, les éditeurs lui donnent un nouveau « signal d'entrée » à propos de ce qu'ils tiennent à valoriser. Ce signal est susceptible d'aider l'algorithme à faire remonter des pages pertinentes dans Google Search et Google Actualités, notamment dans le cas de certaines pages difficiles à indexer — pages que Google encourage à mettre en avant précisément dans l'encart « Choix des rédactions ».

*

Maintenant que nous savons dans quel champ des possibles les éditeurs se déplacent, il nous faut décortiquer la manière dont les intérêts de Google et des éditeurs, et en particulier leurs intérêts économiques, peuvent être liés ou contradictoires. Nous saurons ainsi à quelles opportunités ces différents intérêts peuvent donner lieu, et de quelle façon certaines incitations peuvent peser sur la nature et le résultat du processus d'infomédiation, étant donné précisément les programmes d'action et les marges de manœuvre que nous venons de décrire.

CHAPITRE 5. LE MOTEUR COMME LIEU D'ÉCHANGE ET D'OPPORTUNITÉS

« Hyperlinks between websites have been a defining feature of the World Wide Web right from its inception. Nevertheless, their full economic implications on content ecosystems are not yet fully understood. Linking has fundamentally transformed the ways in which content sites must compete for attention and revenue. »

(Dellarocas *et al.*, 2010, p. 2)

Étant donné sa capacité à rassembler les consommateurs et les producteurs en un même réseau, où l'information circule rapidement, d'aucuns ont pu croire que le média web serait un lieu de concurrence pure et parfaite. Cependant, il apparaît au contraire que, comme ailleurs, des logiques d'économie d'échelle y sont déployées, des tendances au monopole et des stratégies de verrouillage (Farchy, 2006, p. 144).

Google, ainsi que nous l'avons expliqué dans le chapitre 2 (cf. 2.4), a damé le pion à ses concurrents, pour finalement se trouver en situation de quasi-monopole sur le marché des moteurs. Par ailleurs, tout en étant le champion de l'« économie de l'infomédiation » (Chantepie, 2009), la firme œuvre sur plusieurs autres marchés en fournissant ses services aux internautes, mais aussi aux éditeurs et aux annonceurs. Elle se livre ainsi à une stratégie de *bundling*, qualifiée de « stratégie du couteau suisse » (Girard, 2008), qui consiste à multiplier les offres de service autour d'un service principal : le moteur.

Dans ce chapitre, nous proposons de nous intéresser au moteur en tant que lieu de valorisation et d'échange de l'information d'actualité. Nous préciserons en particulier certaines des caractéristiques qui font de la publication et de la hiérarchisation de liens hypertextes un service dont la valeur et son destinataire sont extrêmement difficiles, voire impossibles, à déterminer. Puis, dans un second temps, nous positionnons la recherche d'information parmi les autres activités de Google, afin de décrire de quelle stratégie le moteur est l'épicentre. Nous nous intéressons en particulier aux opportunités de profits attachées à la liste de résultats, afin de localiser les incitations. Dans un troisième temps enfin, nous situons la production journalistique dans l'écosystème déployé.

5.1 Addition des parcours et somme des possibilités

Il s'agit ici de décrire certaines caractéristiques de ce qui est produit lorsque des liens sont générés et ordonnés par le moteur. Nous mettrons notamment en avant le fait que la valeur dépend des actions coordonnées de nombreux acteurs, dont chacun pourrait potentiellement en revendiquer une partie. Nous verrons également grâce à ce que nous appelons « l'arbre de navigation », qu'en plus d'être l'addition et l'interprétation des cours d'action l'ayant précédée, la liste est la somme d'opportunités futures.

5.1.1 Interaction / coproduction

Le moteur Google Search a pour vocation de réunir trois types d'agents : les internautes qui effectuent des requêtes, les éditeurs qui fournissent du contenu et qui bénéficient d'un apport de trafic gratuit (liens naturels) et les annonceurs¹³⁷ qui payent pour le trafic qu'ils reçoivent (liens sponsorisés). Lorsqu'un internaute effectue une requête, les cours d'action de ces acteurs convergent. Le dispositif produit un lien *entre* les actions : *inter*-actions. Et il est important de considérer que la liste n'est pas seulement le fruit des actions des seuls acteurs visibles, mais également d'éditeurs qui n'apparaissent pas sur l'interface. Par exemple, c'est parce que certains éditeurs ont effectué des liens vers les pages d'autres éditeurs que le PageRank de ces derniers a pu augmenter. L'internaute est également l'auteur de ce qui apparaît à l'écran, puisqu'il a choisi pour sa requête les termes dont l'appariement final dépend, et puisque Google utilise des mécanismes de personnalisation pour adapter le moteur à ses préférences (Pariser, 2011). De surcroît, l'historique des autres utilisateurs du moteur est également pris en considération pour attribuer des scores de qualité aux annonceurs, dont dépend également l'appariement (Levy, 2009). La liste en tant que produit éditorial (*output*) ne peut donc pas être seulement imputée à quelques éditeurs, quelques annonceurs et un seul internaute, mais à l'action de nombreux éditeurs, annonceurs et internautes, en plus de l'action des

¹³⁷ Dès lors qu'ils ont un site web, les annonceurs sont aussi des éditeurs. Ainsi, quand nous disons « annonceurs », nous faisons spécifiquement référence aux éditeurs qui reçoivent du trafic par l'intermédiaire des liens sponsorisés. Un même acteur peut donc être présenté comme un éditeur ou un annonceur dans notre compte-rendu, selon la provenance de son trafic.

dirigeants et des employés de Google (*inputs*), formant un même *réseau* d'action. Les acteurs coproduisent l'énoncé et le processus d'infomédiation présente dès lors chacune des propriétés de ce que Latour nomme une *interaction* (Latour, 2007) :

- Il n'est pas *isotropique*. La liste qui apparaît à l'écran est le résultat de différentes actions faites par différents acteurs/actants en différents lieux.
- Il n'est pas *synchronique*. Les actions dont la production de la liste découle ne sont pas effectuées au même moment, et il est même nécessaire que certaines actions aient précédé sa production : le *crawling*, l'indexation, le paramétrage de l'algorithme, la publication des contenus, la programmation des enchères.
- Il n'est pas *synoptique*. Seules quelques entités apparaissent à un moment donné. En effet, de nombreux éditeurs, internautes, annonceurs, ingénieurs, pourtant responsables en partie, et à des degrés divers, de ce qu'est la liste, n'apparaissent pas à l'écran.
- Il n'est pas *homogène*. « Les relais qui assurent le déroulement de l'action n'ont jamais la même qualité matérielle tout au long » (Latour, 2007, p. 294). Il y a, derrière la liste, des infrastructures techniques, des informations et des méta-informations, des logiciels, des documents dont les formats sont différents, des fournisseurs d'accès à internet, des formules mathématiques, etc.
- Il n'est pas *isobarique*, c'est-à-dire que les acteurs impliqués dans la production du méta-énoncé n'exercent pas tous le même pouvoir. « Certains participants imposent leur présence avec force, exigent d'être entendus, pris en considération, tandis que d'autres ne sont plus que des routines relayées par des habitudes [...] sans aucune visibilité » (*ibid.*, p. 295). L'internaute qui a formulé la requête n'exerce pas le même pouvoir que les autres utilisateurs du moteur, qui influencent pourtant eux-aussi le résultat. Les éditeurs n'ont pas la même autorité PageRank. Les annonceurs n'ont pas les mêmes moyens financiers ni le même score de qualité. Les concepteurs ont une plus grande influence sur les caractéristiques de la liste que le blogueur qui effectue un lien vers une page ou le fournisseur d'accès à Internet.

Dès lors que le processus d'infomédiation présente les caractéristiques mentionnées, une myriade d'acteurs pourraient prétendre à une part, même infime, de sa valeur. Car en effet, même si Google permet et cadre la rencontre des cours d'action, il n'en demeure pas moins que les internautes, les éditeurs et les annonceurs ont contribué à son élaboration. Cependant, pour qu'il y ait partage de la valeur, il faudrait que la valeur puisse être établie, or nous nous apprêtons à expliquer pourquoi dans notre cas la notion même de valeur est évanescence.

5.1.2 Bien public / positions privatives

Par essence, le nombre d'interactions possibles est illimité. Ainsi, effectuer une requête et générer une liste ne diminuera pas la quantité de listes disponibles. Le service rendu à l'internaute est donc *non-rival*. Dès lors, le coût marginal pour satisfaire un internaute supplémentaire est nul. « Le prix est donc nul [...], du moins si on le facture au coût marginal. Il est en effet inefficace de vouloir faire payer des consommateurs dont la disposition à payer est faible, puisqu'ils peuvent tirer un bénéfice de l'usage du bien sans gêner les autres. Mais la tarification à prix nul ne permet pas au producteur de couvrir ses coûts fixes » (Benhamou et Farchy, 2009, p. 8). Par ailleurs, tous ceux qui disposent d'une connexion Internet peuvent effectuer une requête sur le moteur. La liste est donc *non-exclusive*. Ces principes de non-rivalité et de non-exclusion lui confèrent le statut de *bien public pur* (Samuelson, 1964 ; Sonnac, 2009¹³⁸). Mais cela n'est vrai que dans la mesure où nous nous plaçons du côté de l'utilisateur, et si nous considérons que les éditeurs et les annonceurs, eux, ne consomment aucun service. Or, il y a bien consommation de leur côté : ce qu'Elizabeth Van Couvering nomme « *traffic commodity* » (Van Couvering, 2011) est en effet consommé gratuitement par l'éditeur et contre rétribution par l'annonceur. Et nous ne pouvons comprendre les caractéristiques de ce qu'ils consomment qu'en considérant les positions (*slots*) que leurs contenus occupent.

¹³⁸ « Paul Samuelson établit en 1964 une classification entre les biens, distinguant biens privés et biens publics. La *rivalité* et l'*exclusion* sont les deux principes sur lesquels s'appuie cette taxinomie : la rivalité est un principe en vertu duquel la consommation d'un bien par agent diminue la quantité disponible de ce même bien par un autre agent ; l'exclusion conduit à écarter de la consommation d'un bien un individu, qui ne pourrait pas ou ne voudrait pas payer, pour jouir de la consommation de ce bien. Un bien privé répond à ces deux principes, à l'inverse du bien public pur. » (Sonnac, 2009, p. 24)

Le nombre de positions disponibles dans les listes de résultats de Google est potentiellement illimité. Cependant, si nous ne considérons que la première page de résultats, qui reçoit plus de 90% des clics (Optify, 2011), le nombre de positions y est limité à dix pour les liens naturels et onze, donc davantage, pour les liens sponsorisés (3 au-dessus des liens naturels et 8 dans la colonne de droite). Occuper une position sur la première page, pour l'éditeur comme pour l'annonceur, signifie diminuer le nombre de positions disponibles pour les autres. Ainsi, de leurs points de vue, les *slots* sont *rivaux*. Même si ce n'est pas le cas pour la liste de liens naturels, il est possible de faire payer pour les *slots*, ce que démontre le fonctionnement de la liste de liens sponsorisés. Par ailleurs, Google peut exclure certains sites du champ d'indexation. C'est ce que la firme fait avec les éditeurs qui ne correspondent pas aux critères de Google Actualités, et ce qu'elle peut faire en plaçant certains sites sur liste noire¹³⁹. D'autres sites, s'ils ne sont pas exclus à proprement parler, sont exclus de la première page, soit parce qu'ils utilisent des formats non-appropriés pour les crawlers, soit parce que leur score de qualité est trop faible. Ces sites n'ont aucune chance de figurer en tête des résultats. Les *slots* sont donc *exclusifs*. Comprise ainsi, la liste est constituée cette fois par un ensemble de positions privatives. La valeur de ces positions étant extrêmement compliquée à établir *a priori* et de manière homogène pour toutes les requêtes possibles, Google laisse aux annonceurs le soin de paramétrer eux-mêmes leurs enchères, et les facturent au coût-par-clic¹⁴⁰ (Battelle, 2005 ; Levy, 2009). Ce sont donc les annonceurs qui attribuent une valeur à chaque position de la liste de liens sponsorisés. Quant à la valeur de la liste de liens naturels, il faudrait, pour la calculer, enquêter sur la valeur attribuée par les éditeurs à chaque position. Cependant, dès lors qu'ils ne payent rien, cette valeur demeurerait hypothétique, calculée sur la seule base des déclarations des éditeurs quant à leur propension à payer. En revanche, les éditeurs qui mettent en œuvre une stratégie de SEO payent une agence externe ou un employé pour optimiser leur position dans les classements. Une enquête approfondie auprès d'eux et sur ce marché tiers nous permettrait sans doute d'évaluer la propension à payer de ces éditeurs pour une

¹³⁹ Ce fut par exemple le cas du site de la société américaine *iAcquire*, après avoir acheté et vendus des liens pour augmenter le référencement de ses clients, ce qui contrevenait aux préconisations de Google.

¹⁴⁰ Cf. 5.2.1

« bonne position » dans les listes de résultats. Même si nous n'avons pas pu mener une telle enquête, et que personne ne l'a encore fait à notre connaissance, il y a tout lieu de croire que la valeur attribuée varierait considérablement selon les secteurs d'activité des éditeurs.

La valeur de la liste et ses propriétés intrinsèques sont donc perçues différemment par les acteurs qui en sont à la fois les coproducteurs et les consommateurs et, finalement, les annonceurs sont les seuls à payer et la firme Google la seule à profiter de ces paiements.

5.1.3 Liste expérientielle

Un bien d'expérience, ou *bien expérientiel*, est un bien dont on ne connaît la valeur qu'après l'avoir consommé (Caves, 2002). Du point de vue des internautes, des éditeurs et des annonceurs, la liste a les caractéristiques d'un tel bien. En effet, tant qu'il n'y a pas eu de clic de la part d'un internaute, aucun des acteurs impliqués ne peut savoir s'il bénéficiera ou non de cette liste en particulier. Ainsi, il réside une incertitude quant à la qualité du service jusqu'à ce que la liste ait été consommée, c'est-à-dire jusqu'à ce qu'elle ait disparu. L'internaute ne peut savoir qu'après avoir cliqué sur un lien s'il est pleinement satisfait, tandis que l'éditeur et l'annonceur ne sauront qu'à ce moment s'ils ont reçu du trafic. Etant donné ce caractère expérientiel de la liste hypertexte, les éditeurs, et dans notre cas les éditeurs de presse en ligne, développent « des procédures [...] de signalisation capables de susciter le désir d'expérience » (Sonnac, 2009, p. 26). Et nous verrons dans la Partie 3 comment ces procédures, rendues possibles par les champs d'action décrits dans le chapitre précédent, sont effectivement mises en place.

D'autre part, la liste est un bien expérientiel d'un type particulier, car elle n'est pas seulement un bien dont on connaîtra la valeur après l'avoir consommé, mais aussi, et surtout, un bien qui n'existe qu'à condition d'être expérimenté. En effet, tant qu'il n'y a pas eu de requête effectuée par l'internaute, et même si les contenus ont été mis en ligne, crawlés, indexés ; la liste n'existe pas. En ce sens, l'expérience effectuée par l'internaute (celle du moteur) lui permet d'effectuer d'autres expériences (celles des liens hypertextes) et de savoir alors si, oui ou non, il aura été

satisfait. Ainsi, la liste est un service intermédiaire qui découle de plusieurs actions et appelle d'autres actions, les premières ne pouvant être jugées satisfaisantes qu'à condition que les suivantes l'aient également été. L'*utilité*¹⁴¹ du moteur est fonction de l'*utilité* des documents vers lequel le moteur conduit l'internaute. Les éditeurs et les annonceurs, quant à eux, ont effectué un certain nombre d'actions visant à apparaître dans les résultats et à attirer du trafic (SEO, enchères). Lorsqu'ils apparaissent sur la première page, leur but est partiellement atteint, mais ils ne savent si cela a été bénéfique que lorsque l'internaute aura cliqué et la liste disparu, c'est-à-dire lorsqu'on sera sorti du dispositif.

5.1.4 Arbre de navigation

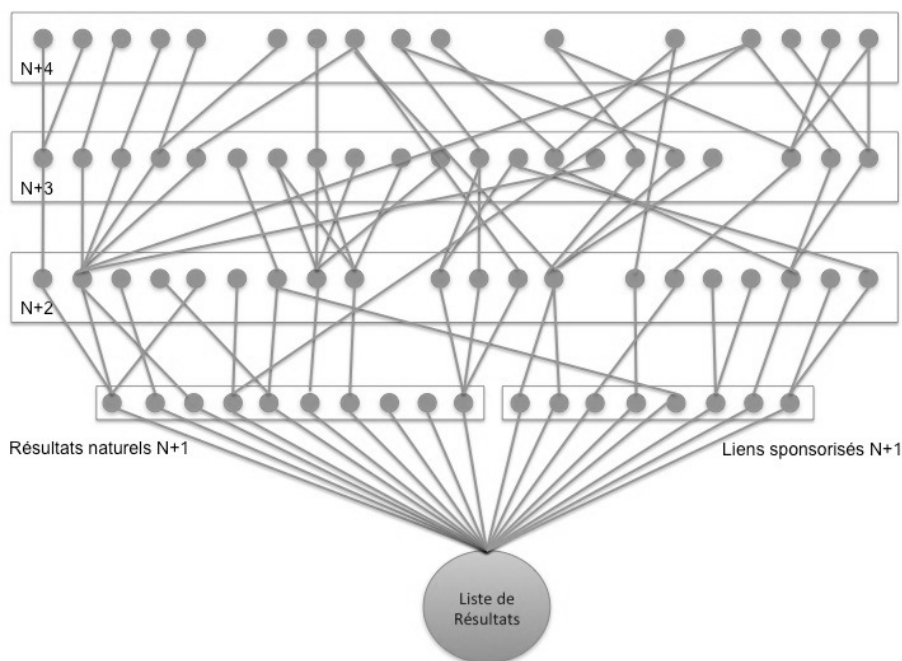
La liste n'est pas seulement un ensemble de liens pointant vers des pages. S'y limiter reviendrait à considérer que la navigation ne dépasse jamais le niveau « $n+1$ », i.e. qu'elle s'arrêtera aussitôt que l'internaute aura cliqué sur un hyperlien. Or, il n'y a pas seulement des pages derrière la liste de liens, mais aussi, sur ces pages, de nouveaux liens, pointant vers de nouvelles pages, contenant elles-mêmes de nouveaux liens, etc. La liste s'ouvre ainsi sur un ensemble de possibilités futures. Cliquer sur un lien, c'est faire un choix, et sortir du dispositif pour aller vers d'autres choix possibles et d'autres dispositifs, fuir pour intégrer, fermer pour ouvrir.

Nous avons représenté dans la Figure 11 quatre niveaux de navigation¹⁴². Les lignes représentent les hyperliens et les ronds des documents. Depuis la liste, l'internaute peut accéder à différentes ramifications de l'arbre. Cliquer sur un lien, c'est quitter un ensemble de navigations possibles pour ouvrir un nouvel ensemble de navigations possibles. Les ensembles peuvent se rejoindre : certains documents du niveau « $n+3$ » effectuent des liens vers des documents du « $n+1$ », accessibles alors par un chemin plus long que celui qui consistait à y accéder depuis la liste. Enfin, certaines pages ne pointent vers aucune autre page : les ensembles navigationnels peuvent être limités ou se croiser.

¹⁴¹ En économie, l'utilité est une mesure de la satisfaction d'un agent lorsqu'il consomme un bien ou un service en particulier. Dans le cas d'un internaute utilisant un moteur de recherche, elle est liée à la satisfaction du besoin d'information qu'il cherche à combler.

¹⁴² Cette représentation est un schéma simplifié, effectué à partir de la requête « chaussure ». Mais dans les faits, l'arbre navigationnel est beaucoup plus complexe.

Figure 11. Schéma simplifié : Arbre de navigation



Ce que nous avons représenté constitue très schématiquement ce que les économistes nomment en théorie des jeux¹⁴³ un « arbre de décision ». Il s'agit de la représentation de l'état présent du jeu, lorsque la liste est à l'écran. Ainsi, la liste, qui est le résultat objectivé et expérientiel de la rencontre d'une multitude de cours d'action, est aussi la racine d'un arbre de décision où chaque lien hypertexte constitue un choix possible, conduisant à d'autres choix possibles. Nous nommerons désormais cet arbre de décision : « arbre de navigation ». Du point de vue de l'internaute, cela rend plus évanescence encore l'idée d'*expérience* présentée ci-avant. Son utilité n'est en effet pas seulement fonction de la page qui vient après la liste, mais également de ses futures opportunités de navigation. Il ne s'agit plus pour Google de donner des résultats jugé pertinents mais de proposer des ensembles navigationnels jugés satisfaisants. Et du point de vue des annonceurs et des éditeurs, la stratégie de référencement peut consister non pas seulement à être mentionné sur la première page de résultats, mais également à être cité par un site référencé sur la première page,

¹⁴³ La théorie des jeux permet aux économistes d'étudier les différents choix possibles auquel un agent est confronté, et ce qu'il sera optimal pour lui de faire étant donné ces choix et les informations dont il dispose à propos du comportement de ses concurrents et de ses partenaires. Il s'agit pour lui d'anticiper, dans une situation où le plus souvent l'information dont il dispose est incomplète, et de faire le meilleur choix possible étant donné ses anticipations.

c'est-à-dire à se positionner au niveau « n+2 » de l'arbre de navigation en espérant recevoir un apport substantiel de trafic grâce au bon référencement d'un autre. Enfin, du point de vue de Google, nous allons voir dans la partie suivante en quoi l'arbre de navigation est le théâtre d'un ensemble d'opportunités de profits et d'incitations économiques, inévitablement liées au processus d'infomédiation.

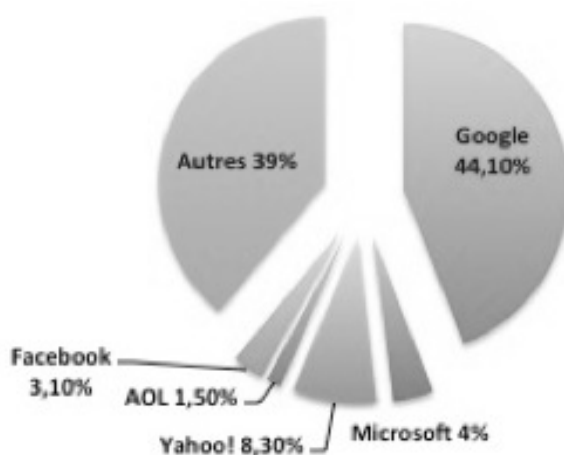
5.2 Opportunités et incitations

Cette sous-partie a pour but de décrire et d'analyser la stratégie globale de Google et de resituer la liste dans l'écosystème dont le moteur est l'épicentre. Nous décrirons en premier lieu la situation de Google sur le marché de la publicité en ligne. Puis nous décrypterons la stratégie de diversification de la firme, avant de représenter et d'analyser les effets de réseaux dont son modèle lui permet de bénéficier. Nous pourrions ainsi obtenir un panorama complet des dynamiques qui enserrant le dispositif, et identifier les incitations pesant sur la nature et le résultat du processus de communication qu'il sous-tend. Cela nous permettra de comprendre les tenants de la controverse dont le partage de la valeur entre Google et les éditeurs fait l'objet et, fort de cette compréhension, d'aborder le terrain de notre Partie 3.

5.2.1 Google et la publicité

Dans le domaine de la publicité en ligne, Google est en situation de position dominante, avec 44,1 % des parts du marché mondial en 2010 (ZenithOptimedia, 2011), sans pour autant être en situation de quasi-monopole comme c'est le cas sur le marché des moteurs. Représentée graphiquement, la situation décrite par ZenithOptimedia est la suivante :

Figure 12. Marché mondial de la publicité en ligne (2010)



Source : ZenithOptimedia, 2011

Même s'il n'y a pas ici de monopole à proprement parler, on voit sur ce diagramme qu'aucun acteur n'est détenteur d'une part de marché comparable à celle de la firme. Google est en position dominante sur le marché de la publicité, avec des concurrents importants mais loin derrière, ainsi qu'une multitude de petits concurrents. Les revenus provenant de la publicité représentaient 96% du chiffre d'affaires de la firme en 2011, qui était de 37,9 milliards de dollars pour un bénéfice net de 9,7 milliards de dollars (Google, rapport annuel 2011). L'activité publicitaire se répartit selon trois axes principaux : les liens sponsorisés (Adwords), la publicité « search » sur les sites web (AdSense) et la publicité display (DoubleClick).

A. Adwords

Les publicités du réseau AdWords, générant plus de deux tiers des revenus de Google¹⁴⁴, sont celles qui apparaissent sur le moteur Google Search, ainsi que sur certains des moteurs verticaux affiliés, à droite de la page de résultats (8 slots disponibles), parfois également en haut de cette même page dans un encart légèrement rosé (3 slots disponibles) et, plus rarement, en bas (2 slots disponibles¹⁴⁵). Elles ont la particularité de répondre à une requête, c'est-à-dire à l'expression d'un

¹⁴⁴ En 2011, Google a généré 26,1 milliards de dollars de chiffre d'affaires depuis ses propres sites (AdWords) et 10,3 milliards depuis son réseau de partenaires (AdSense et DoubleClick) (Google, rapport annuel 2011).

¹⁴⁵ Si les liens sponsorisés apparaissent en bas de la page, aucun lien sponsorisé n'apparaît en haut.

besoin par le consommateur (Röhle, 2009), et ont ainsi vocation, selon les porte-parole de Google, à prendre le contre-pied d'une publicité qui pourrait être intrusive et non-pertinente (Wojcicki, 2007).

Le principe d'Adwords est de proposer aux annonceurs une interface leur permettant d'indiquer l'adresse URL de leur site, leur secteur d'activité et les mots sur lesquels ils souhaitent se positionner. Pour chaque mot-clé, l'annonceur indique également s'il se positionne sur ce mot exclusivement, ou bien s'il souhaite que Google effectue un rapprochement de différents mots autour d'une même thématique¹⁴⁶. L'annonceur attribue également un prix à chaque clic fait par un internaute sur le lien lorsqu'il apparaîtra. Ce système d'enchères constitue une quadruple originalité :

- 1) Le coût par clic : ce n'est pas l'affichage qui est facturé à l'annonceur, comme dans le cas du *display*, mais l'action de l'internaute (CPC), et donc le fait que celui-ci ait exprimé un intérêt et qu'il soit effectivement arrivé sur la page de l'annonceur.
- 2) Le style d'enchères : au moment du lancement du service, les dirigeants de Google pensaient d'abord implémenter un système de « *first-price auction* », lequel consistait à facturer pour chaque mot-clé un prix $p_1 = b_1, p_2 = b_2, \dots, p_k = b_k$, (p pour prix facturé, b pour enchère) mais impliquait que chaque publicitaire x_n ait intérêt à baisser son offre b_n jusqu'à $p_{n-1} + d$ (où le paramètre d tend vers 0). Le publicitaire pourrait ainsi minimiser ses coûts sans que cela ne change son classement. Cependant, « le paramétrage constant du système risquait de surcharger les serveurs, aussi Google a-t-il finalement décidé de facturer par défaut un prix égal à l'offre immédiatement inférieure à celle de l'annonceur, puisque c'est ce qui aurait été fait dans tous les cas. [...] Il s'agissait d'abord d'une décision d'ingénieur » (Varian, 2007, p.184). Ainsi, les dirigeants de Google choisirent de paramétrer le système pour que, par défaut, p_n (le prix payé par l'annonceur x_n) soit égal à b_{n-1}

¹⁴⁶ Hal Varian, économiste en chef de Google, explique que les revenus sont supérieurs lorsque les liens sont survendus plutôt que sous-vendus (Varian, 2007). Les publicitaires, lorsqu'ils font leur offre, ont la possibilité de choisir un « appariement exact », ou bien un « broad match », qu'ils choisissent plus volontiers, et qui leur permettra de voir leur lien affiché si la requête est identifiée comme étant à propos du même thème, même si celui-ci n'a pas employé exactement le mot en question. Cette stratégie permet à Google d'avoir une plus grande marge de manœuvre, en même temps qu'il s'assure d'avoir toujours des liens survendus. Cela bénéficie, selon ce « porte-parole » de la firme, à l'utilisateur du moteur et aux publicitaires « car ils sont davantage de chance qu'un internaute clique sur leur annonce », mais cela bénéficie également à Google, comme l'indique la suite de la citation : « cela augmente aussi la compétition sur les enchères, ce qui fait monter les prix » (Varian, 2007, p. 188).

(l'enchère de l'annonceur x_{n-1}). C'est ce qu'on appelle le système d'enchères « *Generalized Second-Price auction* » (GSP). Nous voyons ici comment une réalité technique a pu décider la firme à innover en matière de modèle économique afin de ne pas risquer d'endommager les infrastructures et la qualité des services fournis. En principe, les deux dimensions sont séparées ; en pratique, elles sont liées. Ainsi, chaque annonceur dit potentiellement la vérité à propos de la valeur qu'il donne au clic et les serveurs ne tombent pas en panne, tandis que les dirigeants et les employés de Google procèdent à une *délégation* vis-à-vis de la plateforme Adwords, qui joue alors le rôle de « dispositif de libre-service » (Cochoy, 2010) capable de vendre continuellement et davantage que des employés humains ne l'auraient fait.

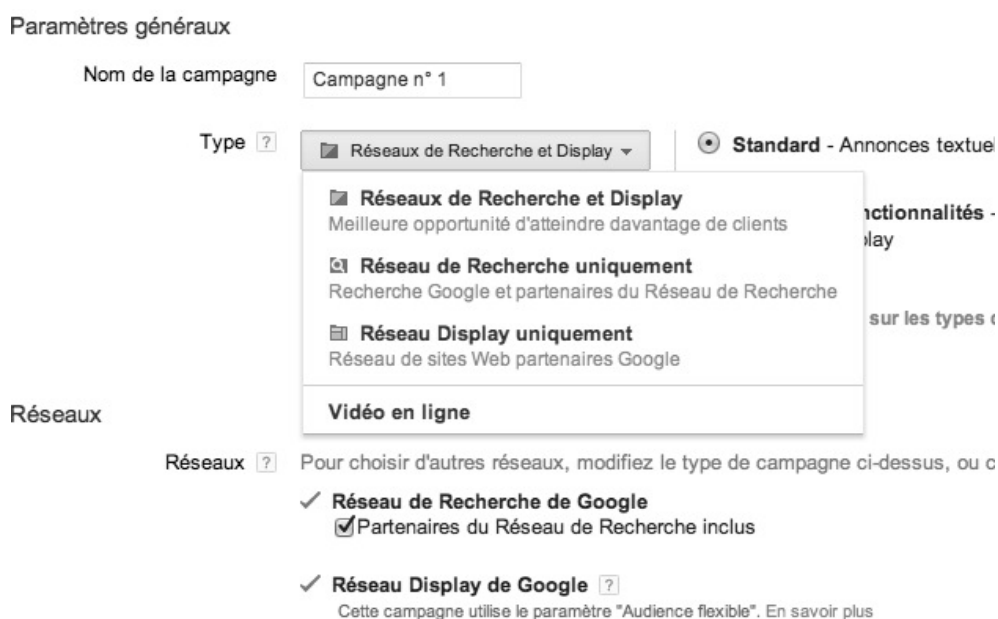
- 3) Le score de qualité : chaque annonceur reçoit un score qui pondère l'enchère, et peut aboutir à ce qu'un annonceur X dont l'enchère est inférieure à celle d'un annonceur Y se retrouve devant Y dans le classement, car sa publicité sera considérée comme ayant le plus fort potentiel de clics¹⁴⁷. On ne sait pas exactement comment ce score est calculé, mais l'on sait que l'historique du taux de clics et l'historique du taux de conversion sont considérés (Levy, 2009). Selon les économistes, il s'agit d'une solution plus efficiente que celle qui consiste à considérer seulement l'enchère (Liu et Chen, 2005 ; Feng *et al.*, 2005 ; Lahaie, 2006).
- 4) La séparation stricte : les porte-parole de Google répètent qu'aucune corrélation n'existe entre la liste de liens sponsorisés et la liste de liens naturels. En revanche, il continue d'exister des suspicions à ce sujet, étant donné l'apparente concurrence qu'il y a entre ces deux listes de résultats. En effet, quelqu'un qui cliquera sur un des résultats de la liste organique ne cliquera pas sur un lien de la liste Adwords : il y a *effet de substitution* (White, 2008).

¹⁴⁷ « L'idée était de positionner les publicités qui avait le plus fort revenu potentiel à la meilleure place, autrement dit à l'endroit où les liens étaient le plus susceptibles d'être cliqués. De la même manière qu'une firme se préoccupe du prix *multiplié par* la quantité vendue, le moteur de recherche doit se préoccuper du prix *multiplié par* le nombre de clics potentiels, qui constitue le revenu total lié à l'affichage de la publicité. Bien sûr, cela requiert qu'on puisse estimer la probabilité d'un clic, ce qui n'est pas une mince affaire » (Varian, 2007, p.183, notre traduction).

B. AdSense

Lorsqu'un annonceur prévoit une campagne sur AdWords, il indique s'il souhaite que ses publicités apparaissent sur le « Réseau de Recherche uniquement », le « Réseau Display uniquement », ou bien les deux, solution mise en avant par Google comme étant la « meilleure opportunité d'atteindre davantage de clients ». Le réseau *display*, dont par exemple Leparisien.fr fait partie, reçoit plus de 166 millions de visiteurs uniques par mois, c'est-à-dire l'équivalent du trafic de Google sur ses propres sites (Levêque, 2009). On y trouve les sites du réseau de syndication AdSense, service mis en place en 2003 permettant à un éditeur d'insérer sur ses pages un module à l'intérieur duquel apparaîtront des publicités textuelles sélectionnées de manière automatisée en fonction de ce que le dispositif aura identifié comme étant la thématique privilégiée. Ce dispositif est nommé « AdSense for content ».

Figure 13. Capture d'écran : Interface annonceurs



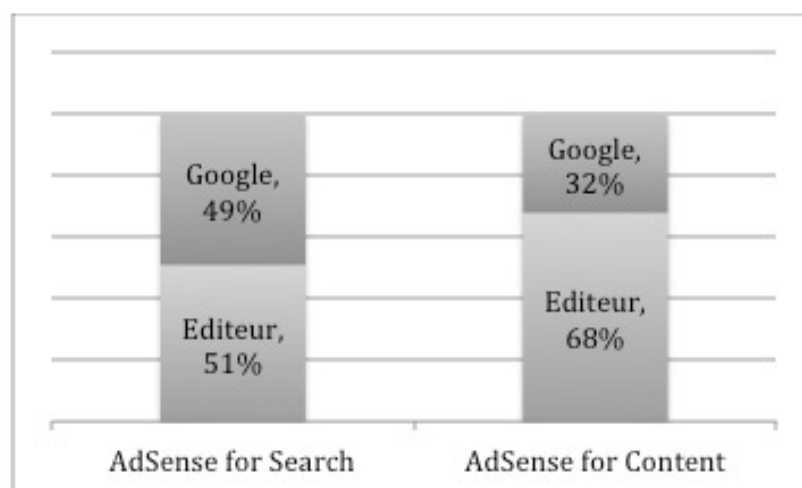
Il est extrêmement aisé d'implémenter AdSense. L'avantage, pour l'éditeur qui adhère, est qu'il bénéficie d'une part de la technique de ciblage publicitaire de Google et, d'autre part, de son réseau d'annonceurs. Ainsi, il n'a pas à prendre en charge la conception de bannières pour son site, ni à entreprendre de démarche pour trouver des annonceurs. Il *délègue* ces tâches au dispositif. En contrepartie, il doit

respecter le règlement édicté par Google : en plus de prohiber l'usage d'AdSense dans le cas de contenu pornographique, violent, lié au jeux d'argent, haineux ou contrevenant au droit d'auteur, le règlement stipule que le site doit présenter un « intérêt » pour les internautes.

Les éditeurs peuvent également mettre sur leur site un moteur de recherche Google (généralement dans la bannière d'entête), dont l'apparence peut être personnalisée, et qui permettra aux internautes d'effectuer une requête soit sur l'ensemble du web, soit sur le site en question, soit sur un ensemble de sites dont l'éditeur aura spécifié les adresses URL. Sur ce réseau, nommé « *AdSense for search* », des publicités apparaîtront dès qu'une requête sera effectuée, de la même manière que les publicités AdWords sur le site de Google Search : à droite et en haut de la page, clairement séparées des résultats naturels.

Les publicités du réseau AdSense peuvent être facturée au CPC, comme c'est le cas sur AdWords, mais également au coût par affichage (CPM). L'éditeur choisit la solution qui lui semble la plus appropriée. Le partage des revenus entre Google et l'éditeur se fait de la manière suivante :

Figure 14. Répartition des revenus Google/Editeur



Source : Google, rapport annuel 2011

Puisque Google jouit d'une situation de monopole sur le marché des moteurs, la firme peut proposer à l'éditeur du réseau « for Search » 51% des revenus sans trop craindre qu'un concurrent ne propose à l'éditeur un partage plus avantageux pour un

service équivalent. Dans le cas de « AdSense for Content », le contenu auprès duquel les publicités se trouvent est valorisé. Google fournit son réseau de syndication et sa technique de ciblage, mais cette fois c'est le contenu qui génère la visibilité des annonces. Il s'agit d'un domaine où la concurrence est plus rude, ce qui explique sans doute que la firme propose un partage des revenus plus intéressant pour l'éditeur.

C. Doubleclick

Google a acquis la régie Doubleclick en 2007 pour 3,1 milliards de dollars, une des plus importantes entreprises mondiales en matière de vente d'espaces publicitaires, d'infrastructures et de services pour les annonceurs. Pointant du doigt une forte concentration du marché au profit d'un seul acteur, de nombreuses protestations eurent lieu lors de ce rachat, à commencer par celle de Microsoft qui s'était également porté acquéreur. Malgré ces protestations, la *Federal Trade Commission* (FTC) et la Commission européenne (CE) donnèrent leur accord, la FTC estimant¹⁴⁸ que les produits vendus par Google et Doubleclick ne pouvaient pas être considérés comme des substituts, et la CE décrétant¹⁴⁹ que « Google et DoubleClick n'exerçaient pas de fortes pressions concurrentielles sur leurs activités mutuelles et que les deux entreprises ne pouvaient donc pas être considérées comme des concurrents »¹⁵⁰. La firme Google est ainsi présente à tous les étages du marché de la publicité en ligne, facturant aussi bien au CPC qu'au CPM, sur le marché du *search* et sur le marché du *display*, mais aussi sur le marché de la publicité vidéo via la plate-forme Youtube et sur le marché de la publicité pour mobiles et tablettes.

En 2008, en plus du rachat de DoubleClick, des changements dans la stratégie publicitaire de Google ont été opérés. Tandis que les publicités étaient créées par les annonceurs directement dans les interfaces de Google, la firme a ouvert l'infrastructure AdSense à des vendeurs tiers comme les agences de publicités. L'acquisition de DoubleClick a notamment permis d'appliquer cette logique à grande échelle, en proposant des services à la fois comme fournisseur d'infrastructure, via DoubleClick Ad Exchange et le réseau Google Display, et comme intermédiaire commercial faisant le lien entre éditeurs

¹⁴⁸ Federal Trade Commission (2007, p.3), « Google/DoubleClick, F.T.C. File No. 071-0170: Statement of Federal Trade Commission Concerning Google/DoubleClick »

¹⁴⁹ Commission Européenne, Bruxelles, 11 mars 2008, « Concentrations: la Commission autorise le projet de rachat de DoubleClick par Google », IP/08/426.

¹⁵⁰ Plusieurs économistes se sont étonnés des conclusions de ces enquêtes, arguant qu'en achetant DoubleClick, Google créait précisément le marché dont la FTC et la CE disaient qu'il n'existait pas (Devine, 2008).

et publicitaires. Tandis que la majorité des revenus proviennent toujours de la vente au clic et à l’affichage, Google a ainsi investi dans la fourniture de services techniques, logistiques et commerciaux aux agences de publicité détenant elles-mêmes de nombreux clients, et aux éditeurs propriétaires de leurs propres régies. Après avoir commencé avec des services adressés aux micro-annonceurs, la firme travaille donc depuis 2008 avec des clients de toutes les tailles en adaptant son expertise dans les infrastructures (*serving*), dans les technologies d’information (*tracking* et *targeting*) et en veille informationnelle (*analytics*) aux pratiques plus traditionnelles du monde de la publicité (Rieder et Sire, 2013).

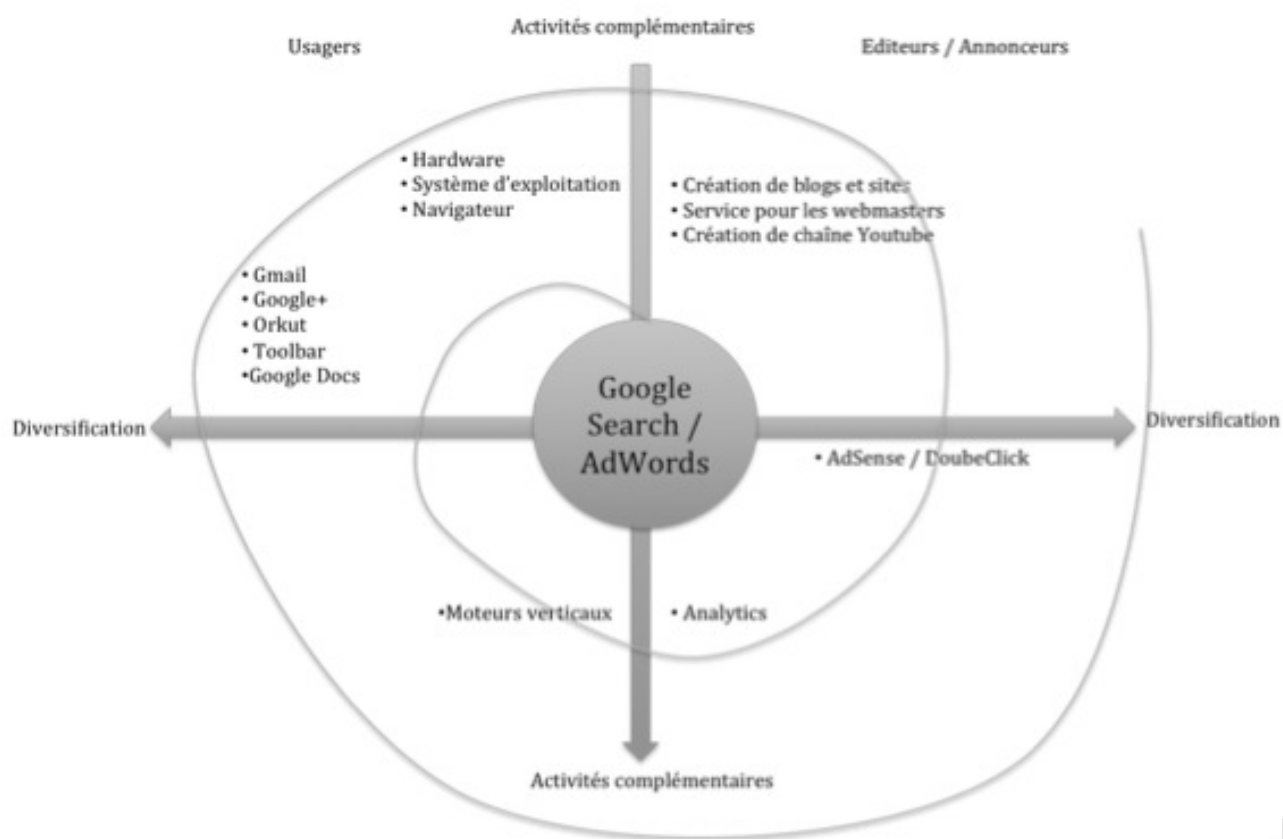
5.2.2 Diversification concentrique

En plus d’être en monopole sur le marché des moteurs et en situation de position dominante sur tous les étages du marché de la publicité, Google œuvre sur de nombreux autres marchés : les navigateurs (Chrome), les messageries en ligne (Gmail), les réseaux sociaux (Google+ et Orkut), le stockage de fichiers en ligne (Google Drive), les barres d’aide à la navigation (Google Toolbar), les outils de mesure et de qualification d’audience (Google Analytics), les outils de création de site internet (Blogger et Google Sites), les logiciels (Google SketchUp), la gestion de portefeuilles boursiers (Google Finance), le traitement de texte, de chiffres et d’images (Google Docs), les systèmes d’application (Android and Chrome OS) et le *hardware* depuis son acquisition du constructeur Motorola en 2011 pour 12,5 milliards de dollars¹⁵¹.

Chaque nouveau service développé ne se substitue pas au précédent mais l’enrichit et le complète, tout en étant autonome. Si un service était arrêté, les autres pourraient continuer de fonctionner. Cela est vrai pour chaque activité sauf pour le moteur de recherche, qui constitue le corps du « couteau-suisse » (Girard, 2008), sur lequel viennent ensuite se greffer les autres services. L’activité permet ainsi d’articuler différents marchés et de développer leurs synergies, grâce aux effets de réseaux que nous présenterons dans la sous-partie suivante. Il s’agit d’une stratégie de *diversification concentrique*.

¹⁵¹ Cette énumération n’est pas exhaustive, même si les principaux services y sont listés.

Figure 15. Schéma : Diversification concentrique (1)



La multiplication des rôles joués par Google est possible notamment grâce aux liquidités considérables que la firme tient prêtes à être mobilisées. En 2011, Google procéda à l'acquisition de *ITA Software* pour 676 millions de dollars et annonça l'achat de *Motorola* pour 12,5 milliards de dollars, tout en réalisant 78 autres acquisitions pour un montant total de 1,3 milliard de dollars (Google, rapport annuel 2011, p. 71). Cela permet à Google de bénéficier de techniques qu'elle n'a pas développées elle-même et de jouir de l'exploitation des brevets afférents. En outre, Google investit à l'interne en Recherche et Développement. Les dépenses dans ce domaine furent de 2,8 milliards, 3,8 milliards et 5,2 milliards de dollars pour les années 2009, 2010 et 2011, soit plus de 11 milliards en trois ans (Google, rapport annuel 2011). A cela, il convient d'ajouter les 20% de temps que les ingénieurs de Google ont le droit de consacrer à un projet personnel susceptible de devenir un nouveau produit de Google ou d'améliorer un service existant (dont provient Google Actualités, cf. Introduction).

Lorsque Google lance un nouveau service, la firme peut décider d'utiliser ce service comme support d'affichage publicitaire, de faire payer les utilisateurs du service, ou bien de proposer le service gratuitement sans y ajouter de module publicitaire. En général, lorsqu'un nouveau service est mis en ligne, il n'y a pas de publicité, et puis, quelques mois ou quelques années plus tard, lorsqu'une masse critique d'utilisateurs a été atteinte, un module AdWords ou AdSense est implémenté sur la page en question. Ce fut notamment le cas de Google Actualités, qui ne générait aucun revenu en tant que tel lors de son lancement en 2004, mais en génère depuis 2009 aux Etats-Unis. Certains services, qui ne font pas forcément, ou pas encore, l'objet d'affichage publicitaire, sont quant à eux directement rémunérateurs. Lorsque cela arrive, c'est aux éditeurs ou aux annonceurs que le service en question est adressé. Le cas de Google Shopping est particulièrement intéressant à mentionner. Google Shopping (anciennement Froogle et Google Product Search) est un comparateur de prix référençant et hiérarchisant des produits vendus en ligne. Le référencement était au départ gratuit pour les e-commerçants. Il s'agissait d'un moteur vertical représentant pour eux ce que Google Actualités est aux éditeurs de presse en ligne. Mais le 31 mai 2012, Google annonça que l'indexation des produits serait désormais payante, sous forme d'enchères, pour les e-commerçants. Très critiquée, cette mesure a montré que la firme, même si ses porte-parole avaient plusieurs fois assuré y être opposés, pouvait choisir de faire payer le référencement aux éditeurs et dégager ainsi de nouveaux bénéfices. Nous pouvons tout à fait imaginer que ce qui s'est passé pour Google Shopping se passe pour Google Actualités : les éditeurs de presse seraient alors obligés de payer pour y être référencés.

Certains services enfin ne sont pas payants et ne font pas l'objet d'affichage publicitaire. L'exemple de Google Actualités, au départ gratuit puis support d'affichage publicitaire (aux Etats-Unis seulement), et de Google Shopping, nous montrent toutefois que ces services peuvent très facilement devenir de nouvelles sources de revenus pour la firme. Par ailleurs, ces services ont deux fonctions. D'une part, ils permettent de développer des synergies grâce auxquelles les internautes finiront par utiliser un service qui, lui, sera rémunérateur : c'est le cas par exemple du navigateur Chrome dont la barre d'adresse URL est par défaut liée au moteur Google Search. D'autre part, tous les services de Google, rémunérateurs ou non, permettent de récolter des données à propos des

internautes, des éditeurs et des annonceurs, et donc de vendre un meilleur ciblage aux annonceurs.

5.2.3 Les données : clef de voûte stratégique

Selon Siva Vaidhyanathan, le cœur de métier de Google n'est pas, ou plus, la recherche d'information, mais le ciblage publicitaire (Vaidhyanathan, 2011, p. 9). Google enregistre des données à propos des sites visités, des chemins de navigation, des taux de conversion, ce qui lui permet de raffiner à la fois la qualité de son moteur et la précision de son ciblage. Le score de qualité peut ainsi prendre en compte l'historique de l'annonceur aussi bien en taux de clic qu'en taux de conversion, mais également considérer auprès de quel type d'internaute cet annonceur a le plus de chance de succès, où, quand, quelle tranche d'âge, quelle fréquence d'utilisation et d'achat on-line, etc.

L'ouverture d'une session Gmail, Google+, iGoogle, Youtube ou sur tout autre service exigeant que l'internaute crée une session personnelle, permet d'identifier l'utilisateur même si celui-ci utilise des adresses IP différentes. Ce n'est plus une adresse IP qui est ciblée mais un individu directement, et ce quelle que soit la machine qu'il utilise ou le lieu depuis lequel il se connecte. Dès lors qu'il se sera identifié, des publicités ciblées pourront continuer à lui être adressés et ce même s'il change de lieu et/ou d'ordinateur.

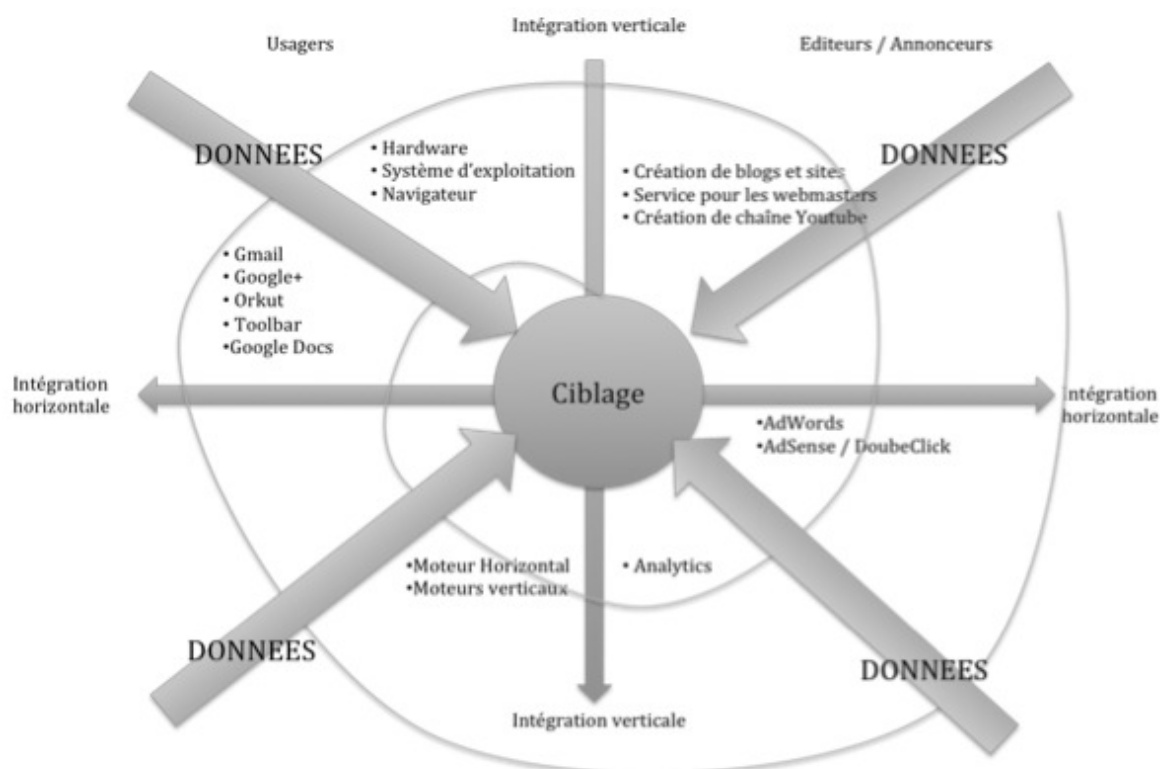
Le réseau social Google+ permet de savoir qu'un individu a apprécié tel ou tel contenu sur tel ou tel sujet et que, par conséquent, il y a une chance pour que les contacts de ce dernier apprécient ces contenus, ces produits, etc. Une des ambitions du réseau social en termes de données est ainsi de relier les chemins de navigation qui jusque là étaient considérés individuellement : utiliser le « graph social » comme signal d'entrée pour l'algorithme d'appariement.

Le navigateur Chrome, quant à lui, est arrivé dans un marché où Mozilla et Internet Explorer jouissaient de positions dominantes depuis déjà un moment. Google a fait d'importantes campagnes de publicité pluri-médias pour pousser son navigateur, et a fini par atteindre 18,9 % des parts du marché européen en mars 2012,

derrière Internet Explorer et Mozilla¹⁵². Le navigateur permet d'enregistrer les informations à propos du comportement d'un internaute et ce même si celui-ci n'utilise aucun autre service de Google. Cela est d'autant plus intéressant que, rappelons-le, ce n'est pas parce qu'un internaute n'utilise jamais les services de Google qu'il ne visitera pas des sites sur lesquels les publicités proviennent des plates-formes AdSense et DoubleClick. Google Toolbar est également un moyen d'enregistrer des données à propos des chemins de navigation. Grâce à Chrome et Toolbar, Google peut ainsi suivre les déplacements des internautes dans l'arbre de navigation.

Si nous reprenons le graphique précédent (Fig. 16), le ciblage comportemental donne une cohérence d'ensemble à la stratégie de diversification concentrique menée par Google :

Figure 16. Schéma : Diversification concentrique (2)



¹⁵² <http://www.atinternet.fr/documents/google-chrome-et-safari-renforcent-encore-leurs-positions-en-europe/> (dernière visite le 21 août 2013)

Le point commun de l'ensemble des activités de Google est de générer des données qui ensuite sont indexées par la firme, traitées, rapprochées, stockées, afin de cibler le mieux possible le profil d'un individu ou d'une adresse IP, d'un éditeur ou d'un annonceur. Ce sont ces informations que Google vend ensuite aux publicitaires et aux éditeurs des réseaux de syndication AdSense et DoubleClick. Il s'agit du phénomène que Yann Moulier-Boutang nomme « pollinisation » : les internautes produisent des données en se déplaçant d'une page à l'autre, d'un service à l'autre, et ce sont ces données et ces déplacements qui permettent à l'ensemble de porter ses fruits, tout comme des abeilles permettent en se déplaçant de transporter les pollens et les spores grâce auxquels un écosystème se maintient et produit (Moulier-Boutang, 2009, 2010). L'auteur y voit un changement de paradigme fondamental. D'une économie de l'échange et de la production, nous serions passés à une économie de la pollinisation et de la contribution, centrée sur l'information, et propre à ce qu'il nomme le « capitalisme cognitif » (Moulier-Boutang, 2007)¹⁵³.

L'immense majorité des informations récoltées par Google ne sont pas rendues publiques, et l'on ne sait pas lesquelles Google enregistre et lesquelles sont utilisées, quand, pourquoi. En France, la Commission nationale de l'informatique et des libertés (CNIL), pour le compte de l'organe européen G29¹⁵⁴, a récemment mené une enquête à ce sujet et a jugé que les réponses de Google étaient trop floues, trop évasives, recommandant à la CE de mettre en œuvre certaines sanctions si la firme ne clarifiait pas sa position.

5.2.4 Marché multiversant

Notre description des activités de Google nous a permis de montrer à quel point la réussite de Google ne reposait pas sur la seule efficacité technique, mais également sur le mécanisme économique à l'œuvre, dont l'ensemble doit être

¹⁵³ Il est intéressant de rapprocher le postulat de Yann Moulier-Boutang selon lequel nous aurions changé de capitalisme des conclusions récentes du rapport rendu au gouvernement français le 18 janvier 2013 par le membre du Conseil d'Etat Pierre Collin et l'inspecteur des finances Nicolas Colin. Ce rapport préconise en effet qu'une nouvelle fiscalité soit progressivement mise en place afin de s'adapter aux modèles d'affaires d'entreprises comme Google dont le modèle économique est essentiellement basée sur la collecte et le traitement de données relatives au comportement des internautes. Ainsi, si nous exprimons les conclusions de Collin et Colin dans le langage de Moulier-Boutang, nous dirions que l'environnement fiscal, hérité de l'ancien paradigme capitaliste, n'est pas adapté à l'économie de la pollinisation.

¹⁵⁴ Le G29, autrement appelé « Groupe de travail Article 29 sur la protection des données », est un organe consultatif européen indépendant travaillant pour conseiller la CE sur la réforme de la législation relative à la protection des données.

considéré dès lors que l'on veut mettre à plat les associations propres au dispositif et comprendre ce qu'elles impliquent. Nous proposons à présent de passer par l'analyse microéconomique — c'est-à-dire l'étude, sur des marchés spécifiques, de la manière dont les ressources sont allouées étant donné les décisions et les interactions des participants (Varian, 2010, p.1) — afin de comprendre à la fois la manière dont la firme inter-relie les cours d'action, et comment ces liaisons se répondent les unes aux autres selon des mécanismes appelés « effets de réseaux ». Cela nous permettra de voir dans quels rouages économiques la liste hypertexte est enserrée et de localiser certains phénomènes d'intéressement et d'enrôlement qui sous-tendent les rapports de forces de Google et des éditeurs.

Nous allons voir dans le cas de Google que les phénomènes observés peuvent être décrits grâce au modèle des marchés multiversants¹⁵⁵, largement étudié depuis les années 2000 et fréquemment mobilisé dans l'étude des dispositifs de communication. Sur un marché multiversant, une plate-forme réunit plusieurs groupes d'agents distincts sur un versant qui leur est propre, chacun de ces groupes ayant la particularité d'avoir potentiellement intérêt à interagir avec les agents du groupe d'au moins un autre versant. Ainsi, la présence d'agents sur le versant n°2 rend plus attractif le bien vendu sur le versant n°1, l'inverse étant également possible (Armstrong, 2006). Il s'agit de ce que les économistes nomment « effets de réseaux croisés ». Pour en bénéficier, le propriétaire d'une plate-forme doit s'assurer que les différents consommateurs concernés par les marchés sur lesquels il opère embarquent à bord de la plate-forme, car si un groupe manque à l'appel sur un versant, la demande tend à disparaître sur l'autre versant (Evans, 2003, p. 195). Pour régler ce problème de circularité, appelé « de la poule et de l'œuf » (Caillaud et Jullien, 2003), David Evans explique qu'il existe deux types de solutions : la subvention d'une catégorie d'agent ou l'investissement sur un des versants (Evans, 2003). On peut par exemple offrir gratuitement un bien à une catégorie d'agents ou bien investir sur un versant en particulier en espérant que les effets sur l'autre versant suffiront à rendre l'investissement profitable (Gabszewicz *et al.*, 1999). Si ces subventions et/ou ces investissements sont importants, des effets de réseaux croisés se mettent en place,

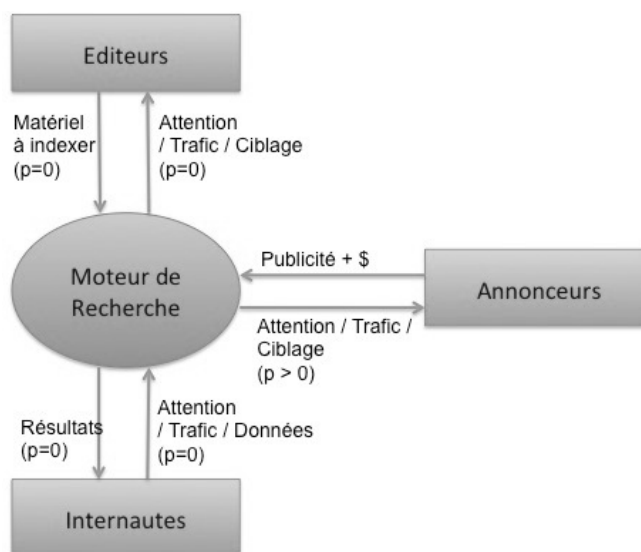
¹⁵⁵ Parmi les microéconomistes, il y a un débat sur la pertinence de l'utilisation de ce modèle dans le cas de Google [(Pollock, 2010) ; (Luchetta, 2012)]. Cependant, ce débat ne nous concerne pas, car nous utilisons la modélisation simplement comme méthode de figuration et de compréhension du réseau, et non pour étudier les conditions de maximisation des surplus. Chez nous, le modèle est une façon de voir et de saisir l'objet. C'est un outil descriptif, pas normatif.

« par lesquels l'attraction d'un seul côté du marché suffit, par effet de débordement, à capturer l'ensemble du marché » (Wauthy, 2008, p. 49).

Nous retrouvons ici de nombreuses expressions qui peuvent être rapprochées de celles employées par l'ANT. Un marché multiversant est en effet un *réseau* au sens économique du terme (Curien, 2005) mais également au sens de l'ANT : un ensemble d'attachements entre des acteurs agissant dans le but de satisfaire leurs intérêts et d'actualiser leurs programmes d'action. Ce que la notion d'effet de réseaux croisé nous apprend, c'est que la mobilisation d'un acteur peut avoir un impact positif sur l'intéressement/enrôlement d'autres acteurs.

Partons du modèle de marché à trois versants simple, articulé autour de la seule activité « moteur de recherche » de Google, inspiré du schéma déjà effectué par Rufus Pollock (2010, Fig.2) :

Figure 17. Schéma : Modèle économique du dispositif d'infomédiation (1)



« Comme on le voit ici, explique Rufus Pollock, le modèle a l'aspect classique d'un marché à trois versants, dans lequel le moteur de recherche agit en tant que plate-forme générant une intermédiation entre les "éditeurs" (qui cherchent à atteindre des "internauts"), les internautes/chercheurs (qui cherchent du contenu) et les publicitaires (qui cherchent à accéder à des "usagers") » (Pollock, 2010, p.7). Les éditeurs fournissent au moteur du matériel à indexer gratuitement, en échange de quoi ils reçoivent

gratuitement du trafic. Les internautes fournissent de l'attention, créent du trafic et divulguent des données à propos de ce qui les intéresse et de leur comportement, en échange de quoi ils reçoivent, gratuitement, des listes de résultats. Les annonceurs, enfin, fournissent des publicités et reçoivent le trafic des internautes qui ont cliqué sur les liens – pour lequel ils payent Google. Les effets de réseau croisés sont les suivants :

- Effet n°1 : S'il n'y avait pas d'éditeur à bord de la plate-forme, les internautes n'auraient aucun intérêt à effectuer une requête. Plus les éditeurs sont nombreux en revanche, et plus un internaute a de chances de trouver l'information qui l'intéresse et par conséquent a intérêt à utiliser Google. Google doit donc s'assurer que les éditeurs n'utiliseront pas massivement le « robots.txt ».
- Effet n°2 : Plus les internautes sont nombreux à utiliser Google, et plus les éditeurs ont intérêt à ce que leurs documents soient indexés, car ils pourront recevoir et monétiser leur trafic. Cet effet permet de les intéresser suffisamment pour qu'ils ne désindexent pas leurs contenus, et éventuellement pour qu'ils se mobilisent davantage en se conformant au programme d'action des concepteurs.
- Effet n°3 : Plus les internautes sont nombreux, plus les annonceurs sont intéressés : ces derniers ont plus de chances d'atteindre leur cible, en plus grand nombre (White, 2009). C'est particulièrement intéressant pour les sites de e-commerce dont le taux de conversion est faible (Moe et Fader, 2004).
- Effet n°4 : De la même manière que dans le cas des éditeurs, si les annonceurs sont nombreux, les internautes auront davantage de chances de trouver ce qu'ils cherchent, car ils auront davantage de réponses proposées, lesquelles seront parfois plus pertinentes que les réponses données dans la liste de résultats. En effet, pour certains types de requêtes, notamment lorsque l'internaute compte acheter une marchandise, les résultats donnés par les liens sponsorisés peuvent être plus satisfaisants que les résultats du référencement naturel (Jansen, 2007).

Ainsi, pour qu'il y ait des internautes, il faut qu'il y ait des éditeurs. Et pour qu'il y ait des annonceurs, il faut qu'il y ait des internautes. Google ne peut résoudre

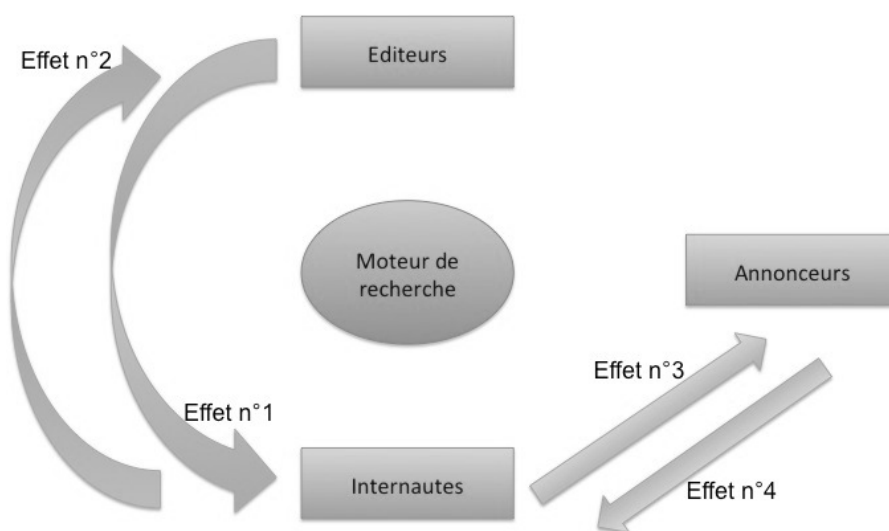
« le problème de la poule et de l'œuf » qu'à condition d'atteindre une masse critique d'éditeurs et d'internautes. C'est pourquoi la firme subventionne intégralement ces deux versants ($p=0$), qui exercent une externalité sur le versant des annonceurs. Les associations se renforcent mutuellement : plus l'association des éditeurs au réseau est forte et plus celle des internautes l'est également, entraînant à son tour une association plus forte de la part des annonceurs.

Nous retrouvons ici une dynamique très proche de celle que l'économiste Lars Furhoff désignait dès 1973 comme une « spirale de diffusion ». Selon lui, le financement par la publicité d'un titre de presse qui dispose d'un avantage par rapport à ses concurrents en termes de diffusion, et donc de plus grands rendements d'échelle, permet au titre en question de proposer des tarifs plus bas aux annonceurs, tout en étant plus attractif puisqu'il touche davantage d'individus. Les revenus qu'il tire du marché de la publicité sont donc potentiellement supérieurs à ceux de ses concurrents, et permettent au titre d'accroître ses rendements d'échelle tout en investissant dans la qualité des articles produits, afin d'attirer de nouveaux lecteurs et d'être encore plus compétitif sur le marché de la diffusion. Cela augmente encore plus sa compétitivité sur le marché de la publicité, et entraîne alors une spirale, nourrie tantôt par la compétitivité sur le marché de la publicité, tantôt par la compétitivité en termes de diffusion, jusqu'à ce que le titre soit en position dominante à la fois sur le marché de la diffusion et sur le marché de la publicité. De son côté, le titre concurrent va subir la spirale inverse. Ne générant pas suffisamment de recettes publicitaires pour recouvrer ses coûts, il révisera à la baisse l'étendue de sa diffusion et la qualité de sa production, ce qui rendra encore moins attractif son titre pour les publicitaires, etc. (Furhoff, 1973 ; Le Floch, 1997 ; Gabszewicz *et al.*, 2007).

Si nous appliquons ce modèle de spirale à Google, il apparaît que la firme, dès lors que son moteur a un grand nombre d'utilisateurs, attire d'autant plus les annonceurs, grâce au paiement desquels elle peut investir dans la qualité de son moteur et dans sa stratégie de diversification, et attirer davantage d'internautes, ce qui attirera d'autant plus les annonceurs et lui permettra de générer davantage de profit, grâce auquel elle continuera d'investir, etc. Non seulement les autres moteurs peineront à demeurer compétitifs, mais, de surcroît, toutes les entreprises (éditeurs de presse y compris) dont le modèle d'affaires est basé sur un coût nul pour les internautes et un financement par la publicité facturée aux

annonceurs seront susceptibles d'être affectées par la spirale ascendante de leur concurrent sur le marché de la publicité.

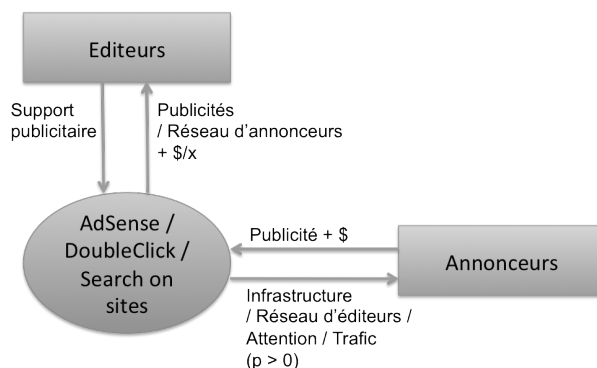
Figure 18. Schéma : Effets de réseau croisés du dispositif d'infomédiation (1)



Ce modèle constitue un excellent point de départ. Cependant, il néglige les autres activités de Google. Par exemple, les régies publicitaires AdSense et DoubleClick en sont totalement absentes. En effet, nous n'avons considéré ici que ce qui apparaissait à l'écran d'un internaute lorsque la liste de liens naturels et sponsorisés était générée par le moteur, négligeant les autres activités de Google. En conséquence, nous proposons d'enrichir cette représentation de manière à représenter autant d'associations qu'il nous sera possible d'identifier.

Si nous établissons un autre schéma, cette fois centré sur les services publicitaires proposés aux éditeurs, nous obtenons :

Figure 19. Schéma : Modèle économique du dispositif d'infomédiation (2)

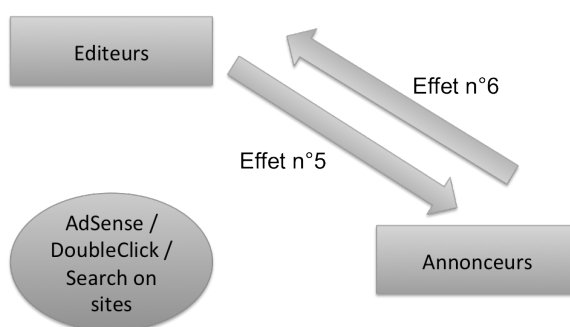


Cette fois, les éditeurs implémentent gratuitement sur leurs pages des modules AdSense, les bannières DoubleClick ou le moteur « Search on sites », en échange de quoi ils bénéficient du réseau d'annonceurs de Google et/ou d'une infrastructure technique. La firme joue ainsi un rôle d'« intermédiaire entre les espaces publicitaires d'un grand nombre de sites, [qu'elle] « [prend] en régie », et les annonceurs qui peuvent acheter un très grand nombre d'impressions à coût réduit en passant par un seul interlocuteur (*one-stop shopping*) » (Ouakrat *et al.*, 2010, p.155). Les annonceurs s'engagent à payer à chaque fois qu'un lien est cliqué (CPC), ou qu'une page est affichée (CPM), en échange de quoi leurs publicités apparaîtront sur le réseau d'éditeurs partenaires. Une partie des revenus revient à Google, l'autre à l'éditeur (c'est pourquoi nous avons écrit sur le schéma « \$ divisé par x »). Nous identifions grâce à notre représentation deux effets de réseaux croisés supplémentaires :

- Effet n°5 : Plus les éditeurs sont nombreux à être clients d'AdSense, DoubleClick ou « Search on Sites » et plus les annonceurs sont intéressés, car leurs campagnes peuvent avoir une plus grande portée, auprès de sites diversifiés et potentiellement plus pertinents, permettant à leurs publicités d'être vues et cliquées par un nombre d'internautes plus important.

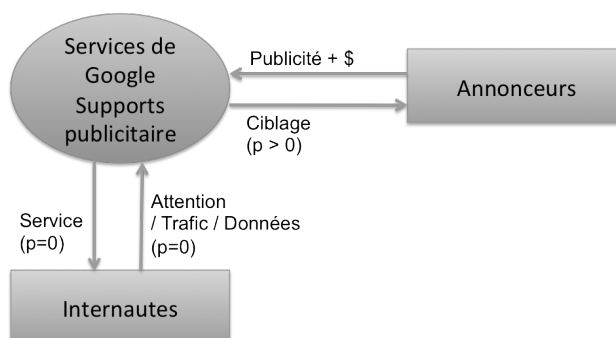
- Effet n°6 : Plus les annonceurs sont nombreux, et plus les éditeurs sont intéressés, car les prix des espaces publicitaires et des clics sur leurs pages sont soumis à une pression à la hausse étant donné l'augmentation de la demande pour ces espaces.

Figure 20. Schéma : Effets de réseau croisés du dispositif d'infomédiation (2)



Si nous représentons à présent les autres services de Google, supports d'affichage publicitaire, comme par exemple le service de messagerie Gmail, nous obtenons :

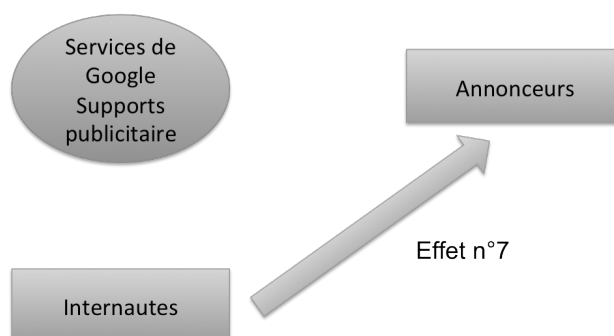
Figure 21. Schéma : Modèle économique du dispositif d'infomédiation (3)



Les internautes utilisent gratuitement les services de Google, en échange de quoi ils divulguent des données à propos de ce qui les intéresse et de leur comportement en ligne. Google affiche les publicités des annonceurs sur ses services, en échange de quoi les annonceurs le payent au CPC le plus souvent, au CPM parfois. Plus les internautes sont nombreux à utiliser ces services, et plus les pages où figureront des publicités seront nombreuses, meilleur sera le ciblage, et plus grandes les opportunités de profit pour Google et les annonceurs.

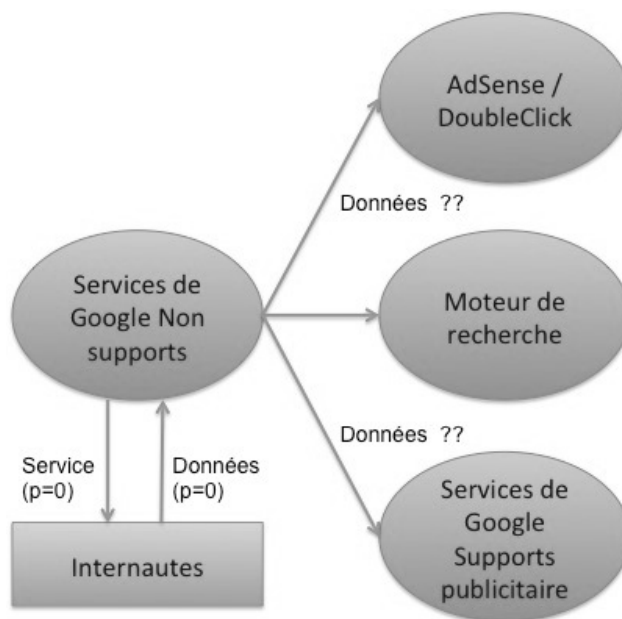
- Effet n°7 : Plus les internautes utilisent les services de Google supports publicitaires, et plus les annonceurs ont intérêt à s'engager.

Figure 22. Schéma : Effets de réseau croisés du dispositif d'infomédiation (3)



Nous représentons enfin les services de Google qui ne sont pas supports d'affichage publicitaire, comme c'est le cas pour la plupart des moteurs verticaux ou pour le navigateur Google Chrome. Cette fois-ci, l'enjeu pour Google concerne les données. Les internautes, même s'ils ne sont pas directement valorisés auprès d'annonceurs, donnent des informations à propos de leurs centres d'intérêt et de leur comportement navigationnel : ils « pollinisent » (Moulier-Boutang, 2009, 2010). Ces données peuvent ensuite servir à Google à établir un meilleur ciblage, lequel pourra être valorisé auprès des annonceurs aussitôt que l'internaute se rendra sur un des services utilisés comme supports publicitaires ou sur le site d'un éditeur appartenant au réseau de partenaires. Toutefois, nous ne savons pas à quel point les données récoltées sur un certain service sont utilisées pour le ciblage sur les autres services.

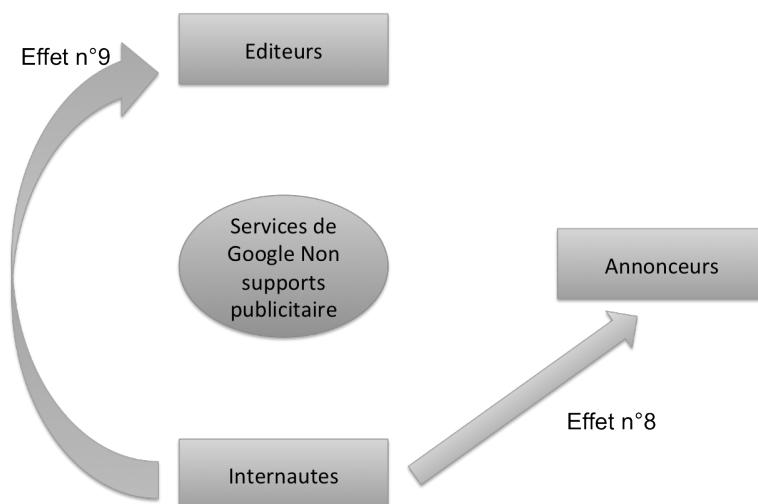
Figure 23. Schéma : Modèle économique du dispositif d'infomédiation (4)



Les services qui ne sont pas directement rémunérateurs pour Google génèrent des données et des effets de réseaux, permettant à Google de consolider sa position sur le marché de la publicité — marché rémunérateur et concurrentiel. A ce sujet, l'explication de Lawrence Lessig, interviewé par Ken Auletta, est lumineuse : « [Les dirigeants de Google] ont produit une machine fascinante pour récolter des données, et ces données ont leurs propres effets de réseau. [...] Chaque fois que vous effectuez une recherche, vous ajoutez de la valeur à la base de données de Google. Cette base de données devient tellement riche que le modèle de financement par la publicité qu'elle soutient s'avère bien plus rentable que tous les autres modèles, simplement parce qu'il a de meilleures données » (Auletta, 2010, p. 138).

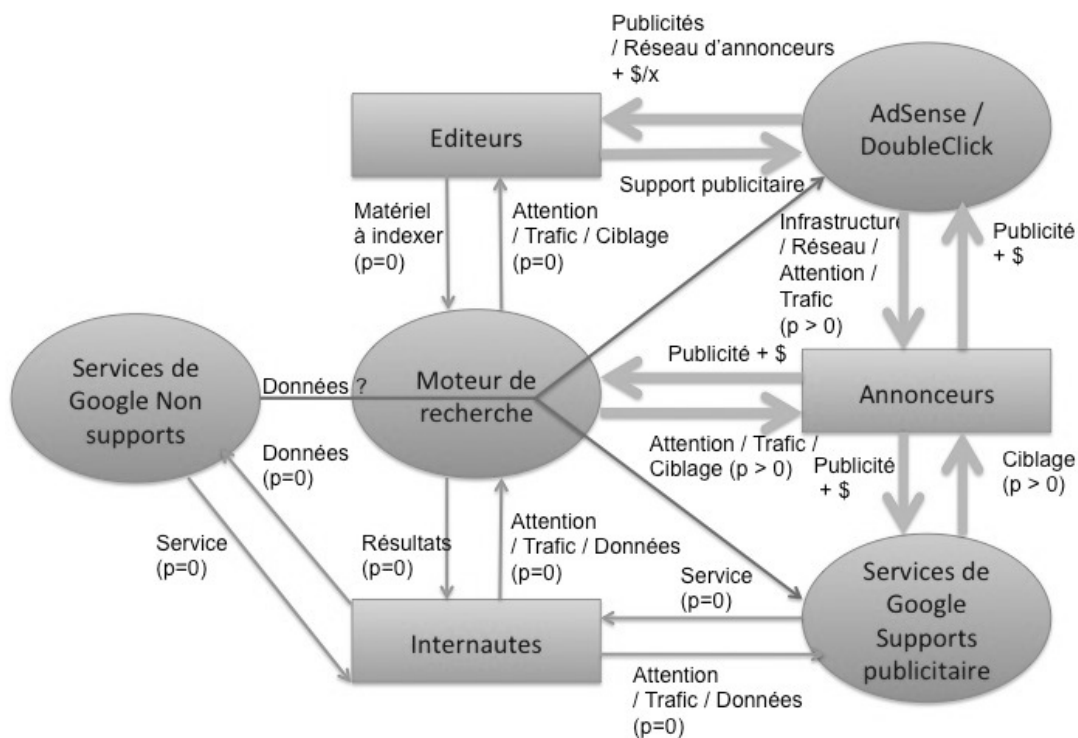
- Effet n°8 : Plus le ciblage comportemental des internautes est efficace, et plus les annonceurs ont intérêt à embarquer à bord de la plate-forme. Leur utilité perçue sera en effet d'autant plus grande que le taux de conversion estimé sera grand.
- Effet n°9 : Plus le ciblage comportemental est efficace, et plus les éditeurs ont intérêt à devenir des partenaires de Google *via* AdSense ou DoubleClick, car Google saura quelles publicités afficher pour optimiser le nombre de clics payants, maximisant ainsi à la fois ses propres recettes et celles des espaces publicitaires des éditeurs.

Figure 24. Schéma : Effets de réseau croisés du dispositif d'infomédiation (4)



Si nous rassemblons nos schémas en une seule représentation du modèle (Figures 18, 20, 22 et 24), nous visualisons le réseau d'échanges dans son ensemble :

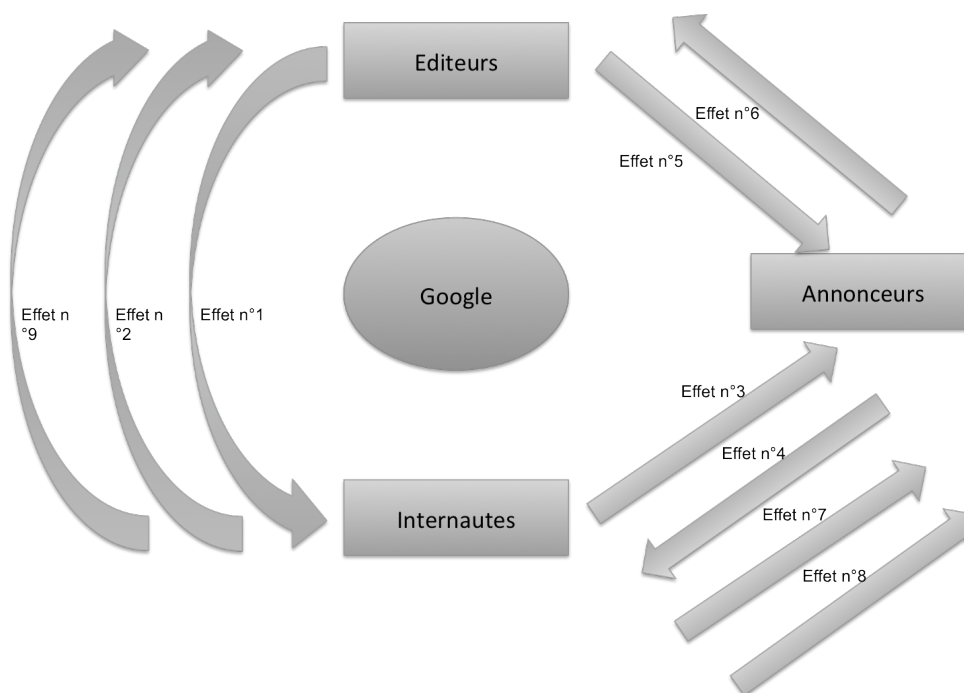
Figure 25. Schéma : Modèle économique du dispositif d'infomédiation (Complet)



Chaque flèche représente le lien économique rendu possible par les outils de *crawling*, d'indexation, d'appariement et d'affichage dont nous avons présenté les modalités dans le chapitre précédent. Nous avons épaissi les flèches désignant un échange de valeur, par rapport à celles qui désignaient un échange basé sur la gratuité. Ainsi, notre analyse descriptive du dispositif et des rapports de pouvoir est enrichie par une visualisation des intérêts économiques et de la galaxie des échanges dont la liste hypertexte est le centre de gravité.

Enfin, si nous représentons l'ensemble des effets de réseaux identifiés (Figures 19 , 21, 23 et 25), nous voyons la manière dont les liaisons propres à notre réseau peuvent s'influencer les unes les autres. Pour cela, nous rassemblons chacune des bulles représentant un type d'activité en une seule bulle représentant l'acteur-réseau Google dans son ensemble :

Figure 26. Schéma : Effets de réseau croisés du dispositif d'infomédiation (Complet)



Dans la troisième sous-partie de ce chapitre, nous positionnerons les éditeurs de presse en ligne dans ce modèle, en analysant ce qu'impliquent pour eux les effets de réseaux 1, 2, 5, 6 et 9 (concernant les éditeurs). Mais avant cela, il nous faut revenir à la liste hypertexte et à notre arbre de navigation dans le but de comprendre par quelles incitations, étant donné les intérêts dévoilés, ceux-ci sont travaillés.

5.2.5 Arbre des opportunités

Nous avons récemment proposé avec notre collègue Bernhard Rieder une typologie des liens hypertextes générés par le dispositif d'infomédiation dans le but de décrire les incitations économiques attachées à chacun d'eux (Rieder et Sire, 2013). Cette typologie n'a pas pour but d'être exhaustive mais de révéler, à partir de notre arbre de navigation, que les liens naturels peuvent eux aussi être générateurs de profit et que, selon les chemins envisagés, les opportunités ne sont pas les mêmes pour Google. La voici brièvement résumée :

1. Liens propriétaires (*Ad garden links*) : pointent vers un site de Google où des publicités peuvent être affichées et des informations à propos du comportement de l'internaute collectées. (ex : Google Actualités ou Gmail)
2. Liens concurrents (*Competitor links*) : pointent vers des sites fournissant des services concurrents des services de Google (produits substitués). (ex : Bing).
3. Liens affiliés (*Ad affiliated links*) : pointent vers des sites partenaires du programme AdWords, directement depuis un des sites de Google.
4. Liens réseau (*Ad network links*) : pointent vers des sites partenaires des réseaux de syndication AdSense et/ou de DoubleClick, sur lesquels l'impression ou le clic sur une publicité entraîne un gain par Google.
5. Liens potentiels (*Ad potential links*) : pointent vers des sites qui ont les ressources pour devenir partenaires du programme AdWords et/ou du programme AdSense mais ne le sont pas.
6. Liens compétiteurs (*Ad competitor links*) : pointent vers des sites partenaires d'un service concurrent d'AdSense et/ou de DoubleClick.
7. Liens persistants (*Ad resistant links*) : pointent vers des sites qui ne font pas de publicités sur leurs pages et dont il est extrêmement peu probable qu'ils en fassent un jour (ex : Wikipédia).

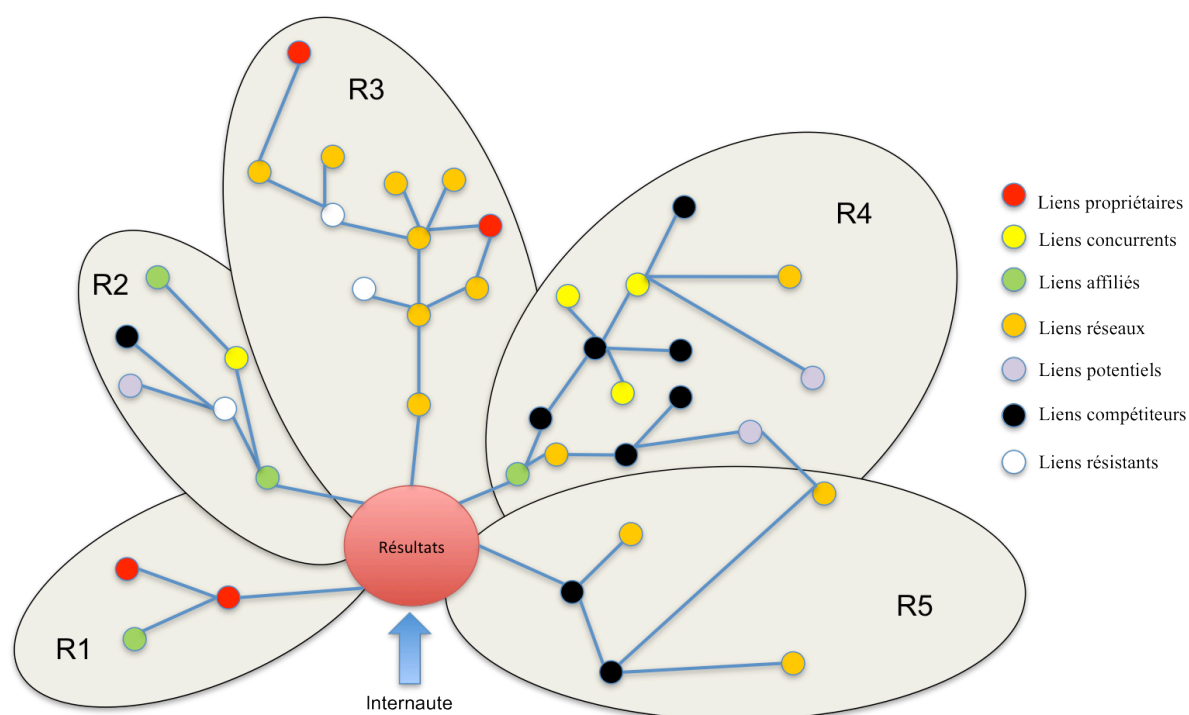
Bien entendu, notre typologie peut être complétée, discutée, et certains cas de figure sont inclassables dans la mesure où ils correspondent à plusieurs des catégories décrites. Néanmoins, nous pensons qu'une telle classification permet de dévoiler des aspects systémiques en incorporant à la notion d'arbre de décision celles d'incitation, d'opportunité et de profit.

L'observateur Dan Sullivan prétend qu'il existe un conflit d'intérêts fondamental dans la situation de Google et se demande dans quelle mesure une entreprise qui délivre un important trafic aux sites devrait également leur fournir les moyens de monnayer ce trafic (Sullivan, 2009b). Etant donné sa double casquette, Google pourrait en effet être tentée de privilégier les sites partenaires de ses réseaux AdSense et DoubleClick dans les résultats naturels. Par ailleurs, le concept d'arbre de navigation et notre typologie permettent de révéler que, dans certains cas, il pourra être économiquement plus intéressant pour la firme que l'internaute clique sur un lien naturel plutôt que sur un lien sponsorisé, car ce lien mènera vers une page où se trouvent une bannière DoubleClick et un module AdSense, elle-même pointant vers des pages partenaires de Google. Une page qui n'est partenaire ni d'AdSense ni de DoubleClick peut elle-même être intéressante dans la mesure où elle renvoie vers de nombreuses pages rémunératrices.

Les opportunités de profit ourdissent autour du processus communicationnel une toile d'incitations. Considérons par exemple l'arbre de navigation schématisé ci-dessous (Figure 28). Chaque bulle représente une page, et chaque trait un lien. R1, R1, R3, R4 et R5 représentent des ensembles navigationnels différents, dont chacun est porteur d'opportunités de profit. Nous voyons sur cette représentation simplifiée à quel point les incitations se trouvant derrière la liste peuvent être complexes. Par exemple, le réseau R4, auquel l'internaute accède via un lien Adwords rémunérateur pour Google, est truffé de liens pointant vers des sites partenaires d'un réseau concurrent de Google. Dès lors, R4 peut potentiellement générer moins de profits que le réseau R3, truffé de sites partenaires de AdSense et DoubleClick. R5, quant à lui, auquel l'internaute accède via un lien compétiteur, comprend des « liens réseaux » et donc de bonnes opportunités de profit. Nous comprenons désormais qu'un arbre de navigation, du point de vue de Google, est aussi un arbre d'opportunités de profit et

donc, inévitablement, constitue un ensemble d'incitations qui justifieraient de biaiser les résultats. Nous ne prétendons pas que ces incitations influencent les résultats, mais plutôt qu'elles font partie du dispositif et qu'elles ne peuvent pas en être exclues *a priori*. Les dirigeants de Google doivent choisir. Mais s'ils décident de ne pas établir de biais, comme c'est apparemment le cas, les incitations ne disparaissent pas, et ils pourront à tout moment revenir sur leur décision. C'est pourquoi nous avons tenu à dévoiler ces incitations qui, même invisibles, font partie intégrante du dispositif et du processus sur lequel nous disons qu'elles « pèsent ». En fait, l'arbre de navigation est, tout autant que la liste, le résultat de l'infomédiation à l'œuvre. En publiant une liste hypertexte, le dispositif ouvre des chemins possibles et crée des opportunités de profits, le résultat de son action n'étant réductible ni à la seule liste, ni aux seuls chemins ou aux seules opportunités, mais à ces éléments conjointement.

Figure 27. Schéma : Arbre des opportunités



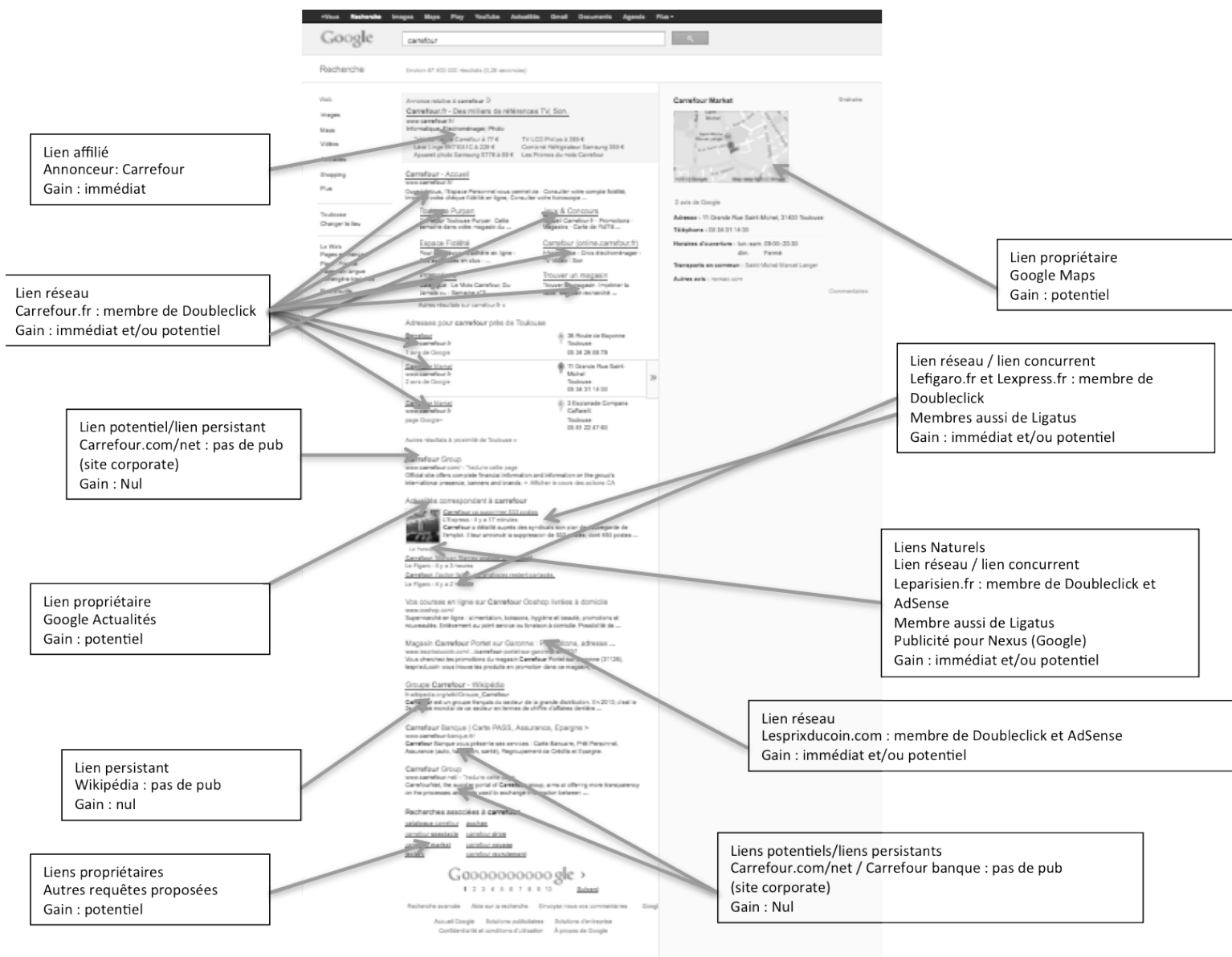
Google peut avoir intérêt à diminuer la présence dans les résultats naturels des « liens potentiels », pour inciter les éditeurs qui pourraient devenir des clients d'AdWords à le devenir effectivement, parce qu'ils auront moins de visiteurs. Par

ailleurs, Google peut avoir intérêt à mettre des « liens résistants » dans les résultats naturels, en pointant par exemple vers Wikipédia, pour réduire le trafic non-payant envoyé vers des concurrents ou vers des clients potentiels de la plate-forme AdWords. A cette aune, la donation versée par Google de 2 millions de dollars à la Fondation Wikimedia¹⁵⁶ en 2010 pourrait être considérée autrement que comme un geste philanthropique (Rieder et Sire, 2013). La liste ne peut en effet pas simplement être considérée étant donné ce qu'elle inclue, mais également étant donné ce qu'elle exclut : lorsque certains sites n'apparaissent pas dans les résultats naturels, leurs propriétaires peuvent avoir intérêt à paramétrer des enchères sur AdWords, et donc à payer dans le but de figurer malgré tout sur la première page de résultats.

Considérons maintenant un exemple tiré d'une observation en ligne (Figure 29). Le vendredi 12 octobre 2012, nous avons fait la requête « Carrefour » sur Google Search, alors qu'il y avait une actualité concernant la chaîne de magasins du même nom, dont les dirigeants avaient annoncé la suppression de 533 postes. Nous avons obtenu une liste de résultats puis nous avons visité chaque page, demeurant à un degré de navigation « +1 ».

¹⁵⁶ La Fondation Wikimedia est propriétaire de l'encyclopédie participative Wikipédia.

Figure 28. Capture d'écran : Requête « Carrefour » (12/10/2012)



Nous voyons dans cette liste qu'il n'y a que quatre liens sur lesquels un clic n'a aucune chance d'être rentable pour Google (on ne considère ici l'arbre de navigation qu'à un niveau « n+1 »). Pour les autres, la potentialité est plus ou moins élevée. Ce simple exemple déconstruit la séparation entre résultats naturels et liens sponsorisés comme elle est présentée par les porte-parole de Google quand ils prétendent que la firme ne génère pas de profits *via* les résultats naturels. Nous voyons apparaître ici le tissu d'incitations auquel Sergey Brin et Larry Page ont eux-mêmes fait référence lorsqu'ils ont expliqué dans le célèbre article donnant naissance

à Google que la publicité créait des « incitations entremêlées » (Brin et Page, 1998). Des incitations s'entremêlent en effet, et sont d'autant plus prégnantes que les crawleurs enregistrent la structure hypertexte et donc que le dispositif a accès aux informations lui permettant de tracer les ramifications de l'arbre de navigation et de calculer les opportunités attachées aux différents ensembles navigationnels.

En guise de conclusion pour cette sous-partie, nous voudrions insister sur le fait que nous ne prétendons pas que l'arbre de navigation est considéré par Google comme un signal d'entrée dans l'algorithme, mais simplement : 1) que cet arbre existe ; 2) que les outils dont Google dispose lui permettent d'en avoir connaissance ; 3) que cet arbre est indissociable de la liste qui apparaît à l'écran ; 4) qu'il génère des incitations qui, même ignorées, continuent d'exister et pourront, à tout moment, ne plus être ignorées ; 5) que les éditeurs peuvent prendre des décisions étant donné ces opportunités et ces incitations.

5.3 Position des éditeurs

A présent que nous avons déployé le dispositif, sous forme de marché multiversant, et enrichi la compréhension de la liste que l'internaute voit apparaître à l'écran grâce aux notions d'interaction, d'expérience et d'arbre de navigation, il nous faut y positionner les éditeurs de presse.

5.3.1 Environnement coopératif

La valeur attribuée au positionnement dans la liste hypertexte peut être plus grande que la valeur attribuée au contenu lui-même. C'est en ce sens qu'il faut comprendre les propos de Pierre-Jean Benghozi, selon lequel « le développement des marchés en ligne s'est accompagné de la création de nouvelles formes d'intermédiation, où le référencement des œuvres [...] a une valeur parfois supérieure à l'œuvre proprement dite » (Benghozi, 2011, p. 112).

D'autre part, Google est un concurrent des éditeurs de presse en ligne sur le marché de la publicité, où certains annonceurs vont potentiellement tourner le dos aux éditeurs pour effectuer tout ou une partie de leurs campagnes via AdWords. Nous avons vu notamment, étant donné une dynamique de « spirale de diffusion »,

comment le succès de Google sur le marché des moteurs pouvait nourrir son succès sur le marché de la publicité et nuire à l'activité des éditeurs sur ce même marché, ainsi qu'à celle des autres moteurs et, de manière général, à l'activité de tous ceux qui comptent sur la publicité pour recouvrer leurs coûts.

Les contenus repris par Google sous forme de liens peuvent se substituer aux contenus du site de l'éditeur. En effet, selon une étude réalisée en 2010 par la compagnie Outsell, 44% des utilisateurs américains de Google Actualités lisent les titres et les accroches sans cliquer sur les liens (Wauters, 2010). Le rôle ainsi joué par les liens générés par le moteur est ambigu dans la mesure où le lien n'est pas seulement une voie d'accès vers la production journalistique, mais également un *substitut*. Dès lors, certains éditeurs accusent Google de saborder le marché de la publicité et de vampiriser leurs contenus. Google leur répond qu'ils sont tout à fait libres d'utiliser le « robots.txt » pour exclure leurs documents et que, par conséquent, s'ils ne le font pas, c'est qu'ils estiment que la valeur du service rendu par le moteur est supérieure à la perte causée par l'action de Google sur le marché publicitaire. La firme propose également aux éditeurs des solutions pour monnayer leurs contenus et partager les profits générés avec elle (AdSense, DoubleClick). Dans ce cas, la firme Google propose aux éditeurs de les aider à remédier à leur manque de compétitivité sur le marché de la publicité, et cela à condition qu'ils partagent avec elle les revenus générés, alors même que ce manque de compétitivité peut en partie être attribuée à la spirale dont la firme bénéficie.

Comme l'ont très bien montré les économistes Dellarocas, Katona et Rand, les effets empirique de l'action d'une entreprise comme Google sont à la fois positifs et négatifs pour les éditeurs dont le modèle économique ne repose que sur l'affichage publicitaire (Dellarocas *et al.*, 2010). D'une part, en rendant plus facile l'utilisation du web et l'accès aux contenus, Google donne de la valeur à l'écosystème de la presse en ligne et lui permet de rivaliser en termes de visibilité avec la presse imprimée. Mais d'autre part, Google génère sur le marché de la publicité des revenus qui auraient potentiellement pu être générés par les éditeurs. Nous retrouvons ici les effets ambigus de la spirale de diffusion : les éditeurs gagnent en matière de visibilité grâce à l'action d'une entreprise qui les concurrence d'autant plus sur le marché de la

publicité qu'elle leur permet de gagner en visibilité. Par ailleurs, les économistes ont montré qu'en hiérarchisant les contenus des éditeurs, et en les plaçant auprès des contenus produits par des amateurs, Google augmentait les pressions concurrentielles sur leur marché. Dès lors, « l'action des agrégateurs profite aux éditeurs de contenu à la seule condition que le référencement permette à ces derniers d'attirer suffisamment de visiteurs additionnels dans leur écosystème pour compenser la perte de revenus publicitaires générée par la présence sur ce marché des agrégateurs » (Dellarocas *et al.*, 2010, p.5). Intéressés par la même question, les économistes Jeon et Esfahani ont d'ailleurs confirmé, que les effets de l'action d'un agrégateur pouvaient être ambigus quant aux profits de la presse en ligne, et qu'il existait une forte probabilité pour que cette action réduise leurs profits étant donné un contexte de congestion publicitaire (Jeon et Esfahani, 2013)¹⁵⁷.

5.3.2 Les éditeurs et les effets de réseaux

Reprenons les schémas réalisés en 5.2.4 .

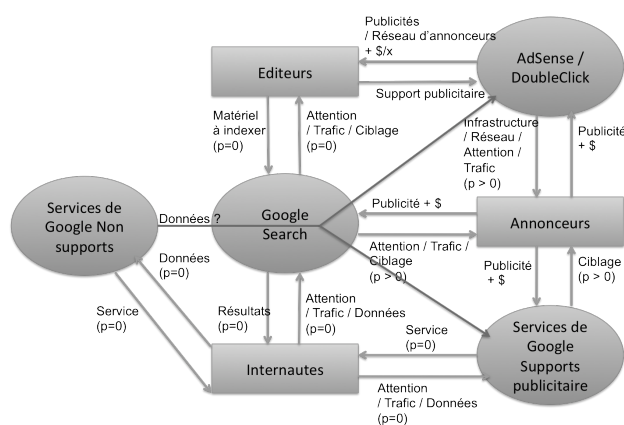


Figure 26

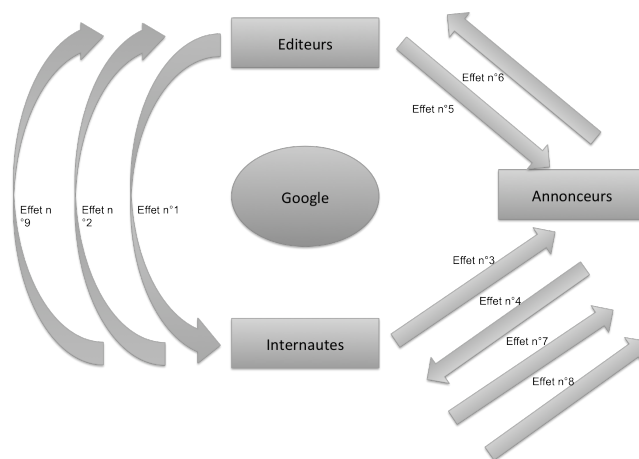


Figure 27

Les producteurs de contenus journalistiques se retrouvent parmi « les éditeurs », auprès des sites amateurs, des sites de relation publique, des fermes de contenus, des e-commerçants. Si l'on considère les effets de réseaux 1, 2, 5, 6 et 9, et qu'on les applique en particulier aux éditeurs de presse, notre modèle permet de

¹⁵⁷ Ce contexte de congestion a très bien été décrit par Anderson *et al.* (2012).

détailler l'ensemble des relations possibles entre les éditeurs et la firme ainsi que les mécanismes qui peuvent contribuer à fortifier ces relations.

- Effet n°1 : Plus les éditeurs de presse en ligne seront nombreux à laisser Google indexer leurs contenus et plus les internautes auront, théoriquement en tout cas¹⁵⁸, accès à des traitements diversifiés de l'actualité, aussi bien en matière d'angles que de choix des sujets. Si les éditeurs décidaient de ne plus laisser la firme indexer leurs contenus, le moteur fournirait aux internautes un service dont la qualité serait moindre et les internautes pourraient choisir d'utiliser un moteur concurrent. Le désengagement des éditeurs pourrait ainsi remettre en cause l'ensemble de la chaîne d'associations, car « un contenu professionnel de qualité demeure indispensable afin d'attirer le grand public » (Smyrnaio et Rebillard, 2011, p.8). Cependant, le poids d'un éditeur, s'il est seul à décider de ne plus indexer ses contenus, est insuffisant pour rompre la chaîne.
- Effet n°2 : Plus les internautes sont nombreux à utiliser le moteur et plus les éditeurs ont intérêt à ce que leurs contenus soient indexés s'ils veulent atteindre une audience la plus large possible — audience qu'ensuite ils monétisent auprès de leurs annonceurs. L'enjeu ici n'est pas seulement économique. En effet, plus les internautes sont nombreux à utiliser le moteur, et plus celui-ci est le porte-parole à qui on a délégué la tâche de hiérarchiser l'espace public. Dès lors, les éditeurs *doivent* y être référencés s'ils veulent être des leaders d'opinion au même titre que leurs homologues de la presse écrite, de la télévision ou la radio.
- Effet n°5 : Plus les éditeurs seront nombreux à être clients des réseaux de syndication AdSense et DoubleClick et plus les annonceurs seront intéressés, car leurs publicités seront juxtaposées à un contenu de qualité. La firme Google pourra donc être plus compétitive sur le marché publicitaire si, dans son réseau, figurent des éditeurs de presse en ligne nombreux et valorisés par les annonceurs.
- Effet n°6 : Plus les annonceurs sont nombreux, et plus les éditeurs de presse en ligne ont intérêt à adhérer aux plate-forme AdSense et DoubleClick. En revanche,

¹⁵⁸ En effet, il existe en fait une forte redondance des contenus journalistiques sur le web (Smyrnaio *et al.*, 2010 ; Marty *et al.*, 2012).

le fait d'implémenter le module AdSense peut être perçu comme un risque de dévalorisation de la marque¹⁵⁹ et donc ne pas être bénéfique pour les éditeurs de presse, ce qui peut fragiliser ou rompre l'association.

- Effet n°9 : Plus le ciblage comportemental est efficace, et plus les éditeurs ont intérêt à devenir des partenaires de Google *via* AdSense ou DoubleClick, car Google saura quelles publicités afficher pour optimiser le nombre de clics payants effectués par les internautes, maximisant ainsi à la fois ses propres recettes et celles des éditeurs. Cela peut contribuer à pallier la dévalorisation mentionnée ci-avant : les éditeurs peuvent optimiser leurs chances que leurs lecteurs soient intéressés (et non plus gênés) par les publicités, et en faire des « publiphiles »¹⁶⁰. En revanche, le fait qu'ils ne puissent pas savoir à l'avance quelles publicités figureront sur leurs sites peut gêner certains éditeurs (Google ne communique pas les noms des entreprises de son réseau d'annonceurs), dont certains redoutent que s'affichent sur leurs pages des publicités venant d'entreprises porteuses de valeurs avec lesquelles ils ne sont pas d'accord.

Maintenant que nous avons situé les éditeurs dans le modèle économique de Google, nous allons nous intéresser à deux tentatives de partenariats économiques effectuées conjointement par la firme et les éditeurs de presse. Cela nous permettra de dévoiler certaines tensions susceptibles d'influencer les stratégies des éditeurs et, donc, la nature et le résultat du processus d'infomédiation.

¹⁵⁹ « Les *ad-networks* proposent en effet aux régies de la presse de commercialiser leurs espaces invendus. La plupart des grands sites, médias et portails, se trouvent alors face au même dilemme. D'un côté, il est tentant de confier cet inventaire à des *ad-networks*, qui pourront l'intégrer à leurs offres à bas prix ou à la performance, et faire en sorte qu'« il rapporte quelque chose plutôt que rien ». De l'autre, accepter de vendre une partie de leurs inventaires à la performance revient à accepter une opération de déqualification, le transfert de leur produit d'un dispositif de qualification à un autre, qui contredit l'ensemble de leurs efforts pour défendre la qualité de leurs espaces. La présence d'espaces publicitaires « de marque » au sein des *ad-networks* crée un risque de dévalorisation de la marque : auprès des acheteurs, qui sont susceptibles de s'en rendre compte et d'être plus réticents à payer 5 € les mille affichages l'espace que d'autres ont acheté à 1 € le clic ; et auprès des annonceurs, qui peuvent estimer que le contrat de lecture de la page web d'un site de qualité est rompu lorsqu'y apparaissent des publicités émanant d'acteurs très peu prestigieux. » (Ouakrat *et al.*, 2010, p.156)

¹⁶⁰ « ... si les lecteurs sont publiphiles, ils seront d'autant plus satisfaits de la consommation de leur titre si la quantité de publicité y est importante (la publicité est considérée alors comme informative) ou encore, l'externalité sera considérée comme positive, si le prix d'accès à l'information est nul » (Sonnac, 2009, p.30).

5.3.3 Deux tentatives échouées

Devant composer avec des entreprises de presse dont le modèle économique est en crise, les dirigeants de Google ont tenté de mettre au point différents dispositifs censés permettre aux éditeurs de monnayer leurs contenus. Fin 2008, alors que Google venait d'acquérir DoubleClick, l'employé en charge de Google Actualités Josh Cohen fut accueilli en France aux *Etats Généraux de la Presse* par des critiques extrêmement virulentes. Il annonça alors que ses équipes travaillaient sur certaines pistes dont l'objectif était de reconsidérer la relation de Google et des éditeurs. Mais quelques mois plus tard, les deux modalités de collaboration auquel ces pistes aboutirent, *Fast Flip* et *One pass*, furent montrées du doigt par les éditeurs comme des opérations de communication visant à illustrer la bonne foi de la firme plutôt que réellement à trouver des solutions. Si nous y revenons ici, c'est parce que ces échecs sont révélateurs de la ligne de front qui oppose Google et les éditeurs.

A. *Fast Flip*

En septembre 2009, Google lança « Fast Flip », agrégateur visuel permettant de naviguer parmi les pages d'une quarantaine de sites web anglophones sélectionnés pour la phase de test. Cette expérience visait, selon les propos tenus sur le blog de Google, à encourager les lecteurs à consommer davantage de contenus : « L'industrie de l'édition affronte de nombreux défis aujourd'hui, pour lesquels il n'y a pas de solution miracle. Mais nous croyons qu'encourager les lecteurs à lire davantage de nouvelles est une composante nécessaire de la solution. Nous pensons que Fast Flip pourrait être un moyen d'aider les éditeurs, et nous chercherons à développer d'autres moyens de les aider à l'avenir »¹⁶¹. Les dirigeants de Google avaient annoncé que les revenus générés par la publicité présente sur « Fast Flip » seraient partagés avec les éditeurs¹⁶². Une question se posa : pourquoi partager les revenus de Fast Flip et pas de Google Actualités ? Interrogé à ce sujet dans le cadre du *Nieman Journalism Lab*, le porte-parole de Google Michael Kirkland éluda : « L'intérêt de Google ici est

¹⁶¹ « The publishing industry faces many challenges today, and there is no magic bullet. However, we believe that encouraging readers to read more news is a necessary part of the solution. We think Fast Flip could be one way to help, and we're looking to find other ways to help as well in the near future. » (Notre traduction)

¹⁶² Un tel partage de revenus existait déjà pour la vente de contenus d'archives, et également pour Google Books et sur Youtube.

d'essayer d'être un bon partenaire de l'industrie des *news* et des fournisseurs de contenus de qualité et d'essayer de trouver des moyens d'aider les éditeurs à gagner davantage. C'est la première étape d'une série de plusieurs types d'expériences ; plusieurs solutions vont être essayées »¹⁶³.

Quelques semaines après le lancement de Fast Flip, Eric Schmidt publia une chronique dans le *Wall Street Journal*, à propos de la relation entre Google et les éditeurs de presse¹⁶⁴ : « de la même manière qu'il n'y a pas une seule cause responsable des problèmes actuels de l'industrie, il n'y aura pas une seule solution. Nous voulons travailler avec les éditeurs afin de les aider à augmenter leur audience, à accroître l'engagement de leurs lecteurs, et à gagner plus d'argent. Surmonter ces défis nécessitera d'utiliser une technologie permettant de développer de nouvelles façons d'accéder aux lecteurs et de les conserver plus longtemps, ainsi que de développer de nouvelles façons d'augmenter les revenus en combinant modèle gratuit et modèle payant. » Le Président de la firme demanda également aux éditeurs de bien vouloir « changer de ton » dans les débats les opposant à Google.

Du côté des éditeurs, « Fast Flip » n'a pas été reçu comme un réel moyen de remédier à la crise qu'ils traversaient, et le discours d'Eric Schmidt ne les a certainement pas rassurés. Par exemple, selon Emmanuel Parody, responsable de *CBS Interactive*, « Fast Flip n'est qu'un jouet, un gadget, une diversion par rapport au débat sur la publicité. Déjà on attend rien de Google Actualités en termes de revenus, mais de Fast Flip, ça tient de la plaisanterie ! »¹⁶⁵. Finalement, Google a arrêté Fast Flip en septembre 2011, sans donner d'explication¹⁶⁶.

¹⁶³ « Google's interest here is in trying to be a good partner to the news industry and to quality providers of news and try to frankly find ways to help publishers get more out of the web. It's the first of several different types of experiments, different things we'll be trying ». (Notre traduction) <http://www.niemanlab.org/2009/09/google-sharing-revenue-with-publishers-for-first-time/> (dernière visite le 21 août 2013)

¹⁶⁴ « just as there is no single cause of the industry's current problems, there is no single solution. We want to work with publishers to help them build bigger audiences, better engage readers, and make more money. Meeting that challenge will mean using technology to develop new ways to reach readers and keep them engaged for longer, as well as new ways to raise revenue combining free and paid access. » (Notre traduction) <http://online.wsj.com/article/SB10001424052748704107104574569570797550520.html> (dernière visite le 21 août 2013)

¹⁶⁵ <http://www.geste.fr/spip.php?article203> (dernière visite le 21 août 2013)

¹⁶⁶ <http://paidcontent.co.uk/article/419-google-kills-its-fast-flip-news-reading-experiment/> (dernière visite le 21 août 2013)

B. Google One Pass

Le 16 février 2011, Google a lancé « One Pass » (ou « Pass Media »), dispositif adressé cette fois directement aux éditeurs. « One Pass » avait pour objet de permettre de simplifier les tâches d'authentification des visiteurs et de gestion des contenus payants. Ainsi les éditeurs pourraient, selon les porte-parole de Google, « se concentrer sur la qualité des contenus qu'ils proposent à leurs lecteurs [...] Basé sur la technologie Google Checkout, le Pass Média est facile à mettre en place et à gérer. En effet, la procédure de configuration se résume à quelques étapes et la gestion du contenu s'effectue via une interface Web simple d'utilisation. Basé sur la technologie Google Checkout, le Pass Média évite d'avoir à intégrer un système de paiement tiers sur le site »¹⁶⁷.

Le « One Pass » prévoyait également une application sur les mobiles et tablettes numériques, et permettait à l'éditeur de fixer lui-même, sans que Google ait son mot à dire, le prix des articles et des abonnements ainsi que les conditions de vente et d'accès au contenu (durée et étendue de l'accès). Il s'agissait d'un « pouvoir de faire » paramétré par Google à destination des éditeurs. La première raison invoquée par la firme était la volonté d'aider les éditeurs à faire payer leurs contenus : « Nous faisons cela parce que nous voulons aider. Je sais que ça paraît naïf, mais c'est la vérité »¹⁶⁸ a expliqué Madhav Chinnappa, en charge des partenariats avec les éditeurs européens. Il semble toutefois que les choix qui ont présidé au lancement de « One Pass » n'étaient pas si « naïfs », dans la mesure où l'on considère que l'objectif était de concurrencer Apple. Google devait percevoir 10 % du montant de chaque transaction effectuée par ce biais, contre les 30 % perçus par Apple via les abonnement « *in-app* » lancés le 15 février 2011, la veille du lancement de One Pass. « C'est une pure attaque vis-à-vis de la stratégie d'Apple et ce à un moment critique, a expliqué l'analyste James McQuivey au *New York Times*¹⁶⁹. C'est exactement ce que l'industrie veut en ce moment : savoir qu'il existe une alternative à Apple et qu'un partenaire proposera des taux plus raisonnables. »

¹⁶⁷ Présentation de « One Pass » par Google.

¹⁶⁸ « We're doing this because we want to help. I know it sounds kind of cheesy, but it's true. » (Notre traduction)

¹⁶⁹ « This is purely a shot across Apple's bow at a critical point in time That's what the industry wants right now, to know there is an alternative to Apple and someone willing to talk about a more reasonable rate. » (Notre traduction)
http://www.nytimes.com/2011/02/17/business/media/17google.html?_r=2&ref=media (dernière visite le 21 août 2013)

Ainsi, si Google présentait « One Pass » comme une solution destinée à aider les éditeurs, cette solution était également destinée à contrer la stratégie d'Apple.

L'initiative a été bien perçue par les éditeurs, ravis de trouver une alternative aux prix pratiqués par Apple, jugés trop importants. En France, le groupement d'intérêts économiques ePresse a signé un accord le 9 novembre 2011 avec Google selon les termes duquel « One Pass » devenait « le principal système de paiement du kiosque numérique ePresse.fr, à la fois pour la vente au numéro et pour les abonnements, sur tous les supports : smartphones, tablettes et sur le web »¹⁷⁰. A propos de cet accord, Frédéric Filloux, alors directeur général du consortium ePresse, déclarait : « le GIE se réjouit de contribuer au développement d'une relation durable et constructive entre Google et la presse française. Nous sommes convaincus que les compétences de Google en termes d'ingénierie, associées à sa vision des médias numériques du futur, peuvent considérablement améliorer les conditions d'accès à l'information. Nous sommes également satisfaits de voir que des médias français s'impliquent dans le développement et le test de nouveaux services qui valoriseront les contenus et contribueront à préserver une information de qualité. »

Mais finalement, alors que le GIE ePresse devait annoncer le lancement officiel de son système de paiement le 1er février 2012, celui-ci fut annulé par Google. « Nous avons écouté les retours de nos utilisateurs et de nos partenaires et nous avons décidé de modifier One Pass, qui était en version bêta, [...] a expliqué à *L'Express* Madhav Chinnappa. Mais je ne peux pas vous donner les détails du nouveau projet, ni quand cela surviendra. Il arrive que nous ne tenions pas nos délais. Nous communiquerons en temps utile. »¹⁷¹ Deux mois plus tard, en avril 2012, à l'occasion de ce que la firme nomme « son ménage de printemps », Google annonça qu'elle mettait un terme définitif à « One Pass », sans donner d'explication et assurant que la firme continuerait « à travailler avec les éditeurs dans le but de construire de nouveaux outils. »¹⁷²

¹⁷⁰ Communiqué de presse : http://www.google.fr/press/pressrel/20111109_pressrelease.html (dernière visite le 21 août 2013)

¹⁷¹ <http://blogs.lexpress.fr/tic-et-net/2012/01/31/one-pass-google-fait-faux-bond-a-la-presse-francaise/> (dernière visite le 21 août 2013)

¹⁷² <http://googleblog.blogspot.co.uk/2012/04/spring-cleaning-in-spring.html> (dernière visite le 21 août 2013)

« Fast Flip » et « One Pass » ont fait l'objet d'annonces largement diffusées par Google puis de retraits rapides, sans explication. Ils doivent être considérés à la lumière de ce que nous avons vu dans le chapitre précédent : Google incite les éditeurs à offrir leurs contenus d'actualité gratuitement (cf. 4.5.1.C). En effet, la solution du « Premier Clic Gratuit », à raison d'un minimum de cinq article par jour et par internaute, n'a, elle, pas été retirée. La firme facilite ainsi l'action des éditeurs qui acceptent de mettre en ligne des contenus gratuits et se financent par la publicité — marché sur lequel elle leur propose des solutions de monétisation.

5.3.4 Dilemme de l'éditeur

Finalement, l'éditeur de presse en ligne se retrouve dans un jeu, au sens économique du terme, à propos duquel il lui manque un certain nombre d'informations. Notamment, il ne sait pas à l'avance quel sera le trafic provenant de Google ni quels seront les résultats de son investissement s'il a décidé de déployer une stratégie de SEO. Par ailleurs, il ne sait pas si le fait d'appartenir aux réseaux AdSense et DoubleClick peut permettre à ses contenus d'être mieux référencés. L'éditeur doit jouer étant donné certaines incertitudes concomitantes aux qualités intrinsèques du service produit et consommé dans le cadre du dispositif. Il peut choisir de désindexer ses contenus, d'actionner les leviers d'actions décrits dans le chapitre 4, de faire payer pour son contenu ou bien de le proposer gratuitement, de traiter directement avec les annonceurs ou bien d'adhérer à une régie publicitaire, qui peut très bien être Google.

L'éditeur est libre d'effectuer différentes actions étant donné ce qu'il sait des dynamiques économiques qui sous-tendent le dispositif. Il peut par exemple se positionner dans l'arbre de navigation de telle sorte que Google aura tout intérêt à ce que ses contenus remontent en haut des classements. Il peut également tenter de négocier directement un partenariat avec Google. Enfin, il peut attaquer la firme en justice dans le but de la contraindre à partager la valeur générée à partir de la liste. L'éditeur doit se poser les questions quant à ses actions futures étant donné ce qu'il sait et ce qu'il ne sait pas à propos des résultats de ces actions et des actions des autres éditeurs, des internautes, des annonceurs et, bien sûr, des concepteurs du

moteur. Et comme dans le célèbre « dilemme du prisonnier », il est possible que le jeu aboutisse à une solution qui ne sera optimale ni pour Google ni pour l'éditeur.

Nous avons vu tout au long de ce chapitre qu'il y avait à la fois un « déplacement de la valeur » (Farchy, 2006, p. 147) et un caractère distribué et évanescent de la valorisation de la liste de positions hypertextes. Chacun de ceux qui ont contribué à sa production pourrait potentiellement revendiquer le droit de toucher une partie des profits qu'elle génère. Pourtant, les annonceurs paramètrent des enchères et payent Google, seul destinataire du paiement, tandis que les éditeurs ne reçoivent aucun revenu. Une telle situation soulève plusieurs questions : Qui est propriétaire de la liste hypertexte ? Si les éditeurs n'utilisent pas le robots.txt ou les balises méta-tags pour désindexer leurs documents, cela signifie-t-il pour autant qu'ils sont d'accord pour que ceux-ci soient utilisés par Google sans être rémunérés ? De quelle forme de *contractualisation* la liste est-elle la partie visible ? De quelles négociations est-elle le résultat ? Comment les modalités de reprise par le moteur du contenu des éditeurs sous forme de liens hypertextes sont-elles déterminées ? Par qui ? Pour quel rapports de forces ? Quelle géographie des responsabilités ? Quelles juridictions responsables ? C'est à ces questions que nous nous intéresserons dans le chapitre suivant.

CHAPITRE 6 : LE MOTEUR COMME TERRAIN D'ENTENTE

« [U]n processus de dispute est toujours une épreuve, c'est-à-dire une situation dans laquelle les individus déplacent et refondent l'ordre social qui les lie. »

(Lemieux, 2007, p. 193).

Dans le chapitre 4, nous avons déployé le dispositif en tant que champ de manœuvres et présenté la liste comme le résultat d'actions possibles et/ou rendues possibles, rassemblées et interprétées en un même lieu, un même instant, situés entre plusieurs cours d'action : une *inter*-actions. Après quoi nous avons montré dans le chapitre 5 comment le dispositif pouvait être travaillé par des modèles économiques complémentaires et/ou concurrents, avant de déployer l'arbre de navigation qu'il irrigue. A présent que nous savons de quels acteurs et de quels intérêts la liste est le catalyseur, l'objectif sera de sonder l'entente qui en sous-tend la production. Nous reviendrons pour cela sur plusieurs exemples nous permettant à la fois de jauger le flou juridique dont le moteur fait l'objet et de décrypter les stratégies des acteurs et leurs rapports de forces. Puis nous nous intéresserons à « l'entente cordiale » conclue récemment en France entre les éditeurs de presse et Google. Enfin, nous analyserons les négociations dont le protocole d'exclusion « robots.txt » a lui-même fait l'objet, en nous attachant à décrypter ce qu'elles révèlent quant à la nature de l'entente nécessaire à l'élaboration de la liste.

6.1 Quelque part entre la loi et l'exception

Des différents travaux académiques et décisions de justice concernant Google ressort une confusion liée à l'hétérogénéité des métiers de la firme en même temps qu'à sa dimension internationale et au manque de transparence dont son action fait l'objet. Aussi bien aux Etats-Unis qu'en Europe, les institutions se heurtent à des cas de figure que les jurisprudences ne permettent pas de trancher avec certitude. Pour le Président de Google Eric Schmidt, « les lois [couvrant le secteur de l'information]

sont incohérentes »¹⁷³. Ces lois doivent selon lui s'assouplir si les pays souhaitent que la recherche d'information puisse être opérée. A partir du même constat, le juriste Alain Strowel ne plaident pas pour un assouplissement mais pour une mutation, qui empêcherait de laisser faire à la firme ce qu'elle veut (Strowel, 2011, p. 227).

Nous essaierons de montrer ci-après la manière dont Google contribue à fabriquer le droit en même temps que le droit contribue à fabriquer Google. La loi et la justice obligent en effet la firme à revenir sur certaines de ses décisions, tandis que dans d'autres cas la firme réussit à orienter l'interprétation du droit dans un sens qui lui convient, sinon à passer outre. Dès lors, plutôt que de nous livrer à l'argument simpliste selon lequel la firme régirait le web, nous tâcherons de montrer comment émerge une jurisprudence co-fabriquée par les concepteurs, les éditeurs, le pouvoir législatif, le pouvoir judiciaire, les usagers du moteur. L'objet de cette sous-partie sera ainsi de présenter les limites de l'action de Google et la manière dont ces limites sont négociées juridiquement et techniquement. Pour cela, nous indiquerons en premier lieu à quelles conditions, aux Etats-Unis et en Europe, le dispositif conçu par Google peut générer des liens hypertextes à partir de documents qui n'appartiennent pas à la firme. Puis nous nous poserons la question du droit dont dépend l'activité du moteur, et nous décrirons en détails certains cas européens, qui nous permettront de comprendre à la fois comment la firme est contrainte par les lois en vigueur et comment, et à quel point, elle peut contourner les lois et les décisions de justice.

6.1.1 *Fair use*

La société mère de Google se trouvant en Californie, la firme se revendique en général du droit étatsunien. Or, dans le *Copyright Act* américain, le degré de créativité suffisant pour qu'un contenu puisse être protégé est très bas, et les titres et les accroches produits par les éditeurs, même extrêmement factuels, peuvent faire l'objet d'une protection. Dès lors, Google doit se prévaloir du « *fair use* » pour avoir le droit de les reprendre sous forme de lien hypertexte. En cas de litige, la justice se posera plusieurs questions sans que la réponse à l'une d'entre elles ne suffise à elle seule à décréter le recours au « *fair use* » (Isbell, 2010) :

¹⁷³ Interview d'Eric Schmidt dans le *Financial Times* du 4 juin 2010, cit. in. (Strowel, 2011, p. 227)

1) *L'usage fait par Google des contenus est-il le même que l'usage qu'en font les sites des éditeurs ?*

Si l'usage est transformatif et qu'il bénéficie à l'utilisateur final, il y aura plus de chances que le *fair use* puisse s'appliquer. Ça a été le cas avec Google Images à l'occasion du procès opposant Google à *Perfect 10 Inc*, où les juges de la Cour d'Appel du *Ninth Circuit* ont estimé que les photos étaient incorporées à un service complémentaire et original.

2) *Le contenu est-il factuel et publié ?*

Les dirigeants de Google ont intérêt à ce que ne remontent sur Google Search et Google Actualités que des contenus factuels et déjà publiés, ce qui risque à la fois d'entraver l'ambition d'exposer des points de vue différents (cf. Introduction) et qui, d'autre part, peut être problématique dans le cas où Google mettrait « en cache » un contenu qu'un éditeur a publié dans un premier temps, puis modifié ou dépublié ensuite.

3) *Quelle est la quantité et la substantialité du contenu repris ?*

La partie du contenu reprise ne doit pas être trop « importante » en termes de quantité et de substance . La substantialité pose problème dans notre cas en particulier, car étant donné la logique de pyramide inversée propre à l'écriture journalistique (Mouriquand, 2011), le cœur du message (sa *substance*) se trouve en général dans les premiers mots.

4) *Quels peuvent être les effets de la reprise de contenu sur le marché où l'éditeur de ce contenu opère et sur les marchés potentiels où il pourrait opérer ?*

Dans la situation de coopération qui caractérise les relations de Google et de la presse et la ligne, il faudrait pouvoir dire à quel point la reprise des contenus fait perdre ou gagner son lectorat au site d'un éditeur. Dans nos entretiens, il est clairement apparu que lorsqu'un contenu est bien référencé sur Google Search ou Google Actualités, l'audience est plus importante. Une telle situation plaide donc en faveur de l'application du *fair use*. Cependant, nous avons également vu avec la « spirale de

diffusion » (cf. 5.2.4) que l'activité de Google pouvait nuire aux éditeurs sur le marché de la publicité, ce qui pourrait plaider contre l'application du *fair use*.

Nous voyons ici comment la firme est incitée par la loi à se positionner le plus possible comme un complément, et non comme un substitut aux services fournis par les éditeurs. C'est bien dans cette direction que s'oriente le discours des porte-parole de la firme, selon lesquels Google n'a aucune ambition éditoriale (Battelle, 2005, p. 3 ; Strowel, 2011, p. 62). Mais les éditeurs n'en accusent pas moins Google de détourner une partie de leur lectorat¹⁷⁴ et de casser les coûts sur le marché de la publicité, c'est-à-dire d'être un concurrent sur les deux versants de leur activité.

6.1.2 *Hot news misappropriation*

Une autre doctrine que le *fair use* existe en droit américain, la « *hot news misappropriation* », susceptible de donner raison aux éditeurs dont le contenu est repris par Google, et dont l'application obligerait la firme à les rémunérer¹⁷⁵. Cette doctrine concerne les informations dont la valeur décroît rapidement dans le temps, et donc spécifiquement les informations qui ont trait à l'actualité. Selon cette doctrine, même si des faits ou des idées ne peuvent théoriquement pas faire l'objet d'une protection en tant que tels (dont seul un contenu en particulier peut faire l'objet, c'est-à-dire un angle de traitement, un texte, une ponctuation, des images), celui qui a récolté des informations, et a supporté des coûts pour cela, peut prétendre être le seul à les utiliser tant qu'elles auront de la valeur (McDonnell, 2012). C'est pourquoi cette

¹⁷⁴ Par exemple, les porte-parole du site *lefigaro.fr* écrivent : « il s'avère que de plus en plus d'internautes ne cliquent pas sur les liens des éditeurs et se contentent des quelques lignes référencées par Google Actualités » (Strowel, *op. cit.*, p. 63).

¹⁷⁵ La doctrine américaine « *hot news misappropriation* » date d'une décision de la Cour Suprême prise en 1918, lors du procès opposant les agences de presse *International News Service* (INS) et *Associated Press* (AP). Ces agences fonctionnaient sur un modèle classique : elles produisaient du contenu qu'elles fournissaient aux titres de presse abonnés à leurs « fils de dépêches », qui pouvaient ensuite utiliser ces dépêches pour écrire leurs articles, ou bien les publier en l'état. Pendant la Première Guerre Mondiale, le propriétaire d'INS, William Randolph Hearst, s'était montré très critique vis-à-vis de la Grande Bretagne et de l'entrée en guerre des Etats-Unis, et sympathisant de la cause allemande. Aussi, la Grande Bretagne décida d'interdire les reporters d'INS de produire des dépêches à propos de la guerre. Cependant, pour conserver ses abonnés, INS devait renseigner ses abonnés à propos de la guerre. Ainsi, INS fut accusé d'utiliser les contenus des journaux de la côte est des Etats-Unis (que l'agence achetait), clients d'AP, pour produire et envoyer ses propres dépêches. Etant donné la rapidité de la diffusion d'INS, certains de ses abonnés pouvaient ainsi publier certaines informations à propos de la guerre avant les abonnés d'AP. A l'issue de la guerre, AP attaqua INS en justice, et la Cour Suprême énonça à l'issue du procès la doctrine « *hot news* », selon laquelle il est interdit de reprendre ou d'utiliser une « histoire », une « information », une « nouvelle », qui a coûté cher à produire et dont la valeur décroît extrêmement rapidement dans le temps, car alors celui qui a supporté ses coûts de collecte et de production voit diminuer ses ventes au moment même où l'information a le plus de valeur.

doctrine nous intéresse en particulier. La juriste Kimberley Isbell (2010) explique que pour que cette doctrine puisse être utilisée, il faut que :

1) *l'information ait un coût significatif.*

Dans notre cas, il y a bien un coût significatif, dès lors que les éditeurs de presse en ligne ont payé des journalistes qui ont récolté l'information.

2) *l'information soit sensible au temps.*

Cela est toujours vrai pour de l'information d'actualité et devrait, par conséquent, contribuer à protéger le contenu des éditeurs. Google est incitée à reprendre des messages qui ne sont pas les plus récents, ce à quoi incite également le *fair use*. Cela va contre le traitement en temps réel des informations d'actualité, dont la valeur décroît à mesure qu'on s'éloigne de l'événement concerné.

3) *l'usage de l'information par un tiers soit un « free riding » (i.e. n'y n'ajoute rien).*

Cela ne concerne pas Google dans la mesure où son dispositif accole des contenus à d'autres contenus provenant d'autres sites, et produit ainsi un nouvel énoncé, ou méta-énoncé.

4) *le tiers soit un concurrent direct de l'éditeur.*

Comme nous l'avons expliqué dans le chapitre 5, Google est un partenaire des éditeurs sur le marché de l'accès à l'information et un concurrent sur le marché de la publicité, aussi ce point, dans notre cas, est-il extrêmement difficile à déterminer pour un juge.

5) *l'activité du tiers nuise commercialement à l'action des éditeurs.*

Nous avons plusieurs fois expliqué les ambiguïtés liées à l'effet de l'action de Google sur l'activité des éditeurs, autant sur le versant de la diffusion des contenus que sur le versant publicitaire de leur activité (cf. 5.2.4 et 5.3.1). Les économistes ne sont pas clairs sur ce point (Dellarocas *et al.*, 2010 ; Jeon et Esfahani, 2013), ce qui complique là encore la tâche du juge. Le fait que 44 %

des internautes consultent les liens de Google Actualités sans y cliquer (Wauters, 2010) plaide cependant clairement selon le juriste John McDonnell pour une application de la *hot news misappropriation* en cas de procès entre les éditeurs de presse en ligne et Google (McDonnell, 2012, p. 268).

6.1.3 Une jurisprudence favorable

Finalement, les moteurs de recherche et les agrégateurs, et donc Google, bénéficient d'une relative sécurité juridique aux Etats-Unis, même si la *hot news misappropriation* peut être invoquée, ce qui semble de plus en plus difficile étant donné les conclusions du *Second Circuit* (Calman et Balin, 2011 ; McDonnell, 2012). L'influence du système juridique américain « autorise à dire que la même observation peut être formulée pour les territoires canadien, australien, néo-zélandais, israélien ou indien » (Manara, 2011, p.152). Cependant, aucune conclusion définitive ne peut être avancée avant qu'une juridiction en particulier n'ait tranché un cas particulier. La jurisprudence peine à s'établir et ce qui est aujourd'hui favorable à Google aux Etats-Unis peut très bien ne plus l'être dans quelques mois. Le vent aurait notamment pu tourner lorsque le service Google Actualités a commencé à afficher de la publicité en 2009 sur ses pages de résultats et donc à générer des revenus grâce à leur contenu¹⁷⁶. Plusieurs éditeurs ont décrété que cela devait changer les choses mais, malgré leurs protestations, l'utilisation commerciale n'a pas suffi à rendre obsolète le *fair use* (Bayard, 2009). La justice américaine a en effet clairement statué que le *fair use* pouvait être appliqué dans le cas d'un usage rémunérateur. Par ailleurs, à l'occasion de l'affaire opposant *Campbell* à *Acuff Rose Music Inc.*, la Cour Suprême a déclaré que « plus l'usage est transformatif, moins les autres facteurs, comme l'usage commercial, pèseront en matière d'application du *fair use* »¹⁷⁷. Selon Sam Bayard (*op. cit.*), étant donné cette jurisprudence, le *fair use* pourra être appliqué à Google Actualités même si le service affiche des publicités, car le service, lorsqu'il est utilisé comme un moteur, et non comme un agrégateur, est une version spécialisée de

¹⁷⁶ Les publicités apparaissent aux Etats-Unis sur les pages de résultats lorsqu'une requête est effectuée sur Google Actualités, mais n'apparaissent pas sur la homepage du service. En France, aucune publicité n'apparaît jamais sur Google Actualités, mais de nombreux observateurs disent que cela pourrait changer prochainement (Strowel, 2011).

¹⁷⁷ « the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use. » (Notre traduction)

Google Search. Ainsi, si le *fair use* est appliqué à ce dernier, qui affiche de la publicité, il doit être appliqué à l'autre, qui peut alors afficher lui-aussi de la publicité. En revanche, pour le service d'agrégation, un tel raisonnement plaide contre l'application du *fair use* (*ibid.*). C'est pour cette raison que Google n'affiche pas de publicité sur Google Actualités si une requête n'a pas été effectuée.

6.1.4 Triple test et droit d'auteur

En Union européenne, il existe depuis mai 2001 une directive¹⁷⁸ concernant le copyright (*EU Copyright directive*), dont l'article 5.5 introduit ce qu'on appelle le « Triple test »¹⁷⁹. Ce dernier stipule que des exceptions au *copyright* peuvent être faites : 1) dans certains cas spéciaux 2) à condition de ne pas être en conflit avec l'exploitation normale du travail concerné 3) à condition de ne pas porter de préjudice injustifié aux intérêts légitimes de l'auteur et du propriétaire du contenu. Nous ne rentrerons pas ici dans une analyse détaillée du Triple test, mais il est intéressant de noter que son interprétation est débattue par les juristes (Gervais, 2005). Il est également intéressant de constater que nous y retrouvons certaines des idées du droit étatsunien. En effet ce test, ramené à notre objet, plaide en faveur d'une utilisation qui complètera la production des éditeurs sans nuire à leurs intérêts. Une fois de plus, Google se retrouve devant une situation où il faudrait prouver à la justice que ses outils poursuivent un but complémentaire, et non similaire, aux buts des éditeurs.

Comme le souligne Barry Sookman (2011), la mise en « mémoire cache » et l'absence de toute compensation versée aux éditeurs posent particulièrement problème dans le cas du triple test : « Ces exceptions [celles du triple test] ne peuvent pas être rencontrées dans la mesure où le cache de Google interfère avec l'exploitation normale du travail, cela car les internautes peuvent accéder aux articles qui ont été dépubliés par l'éditeur ou bien qui sont accessibles seulement après un paiement effectué par l'internaute. En outre, le service de Google Actualités empêche

¹⁷⁸ Rappelons ici que contrairement au règlement, la directive européenne n'est pas applicable en tant que tel mais doit être transposée par les pays membres, à qui elle fixe pour cela des objectifs et des délais.

¹⁷⁹ En plus de la directive européenne, le Triple test était déjà présent dans la Convention de Berne après sa révision à Stockholm en 1967, et apparaît également dans plusieurs autres traités et accords internationaux, dont les principaux sont le Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) (Article 13), le WIPO Copyright Treaty (WCT) (Article 10) et le WIPO Performances and Phonograms Treaty (WPPT) (Article 16).

les porteurs de droit de recevoir une compensation substantielle correspondant à une nouvelle publication de leur travail »¹⁸⁰ (*op. cit.*). C'est pourquoi Google propose aux éditeurs d'implémenter la balise « noarchive » (cf. 4.1.4), permettant d'indiquer aux crawleurs qu'aucune version « en cache » ne doit être proposée.

Le flou qui entoure l'application du Triple Test a permis aux différents Etats signataires de transposer la directive selon qu'ils étaient des pays de droit d'auteur ou des pays de *copyright* (Geiger, 2007). La principale différence entre le droit d'auteur européen et le *copyright* américain réside dans le fait que le premier protège une personne physique (auteur, héritier) alors que le second peut attribuer des droits à l'éditeur ou au producteur dont il protège les investissements économiques. Dans la réalité, ces deux droits ont tendance à se rapprocher, mais le droit d'auteur à l'européenne, par rapport au droit étatsunien, demeure attaché à l'existence d'un droit moral reposant sur des bases plus éthiques qu'économiques (Farchy, 2001). Ainsi, dans les pays anglo-saxons, « la balance des intérêts est biaisée en faveur des intérêts économiques en jeu. Le système français, par contre, est un exemple de système où le juge protège autant les droits patrimoniaux que les droits moraux » (Van Der Wielen, 2010). C'est pourquoi il est plus intéressant pour Google d'être soumis au droit américain plutôt qu'au droit européen. Le juriste Jonathan Band explique que la directive européenne et ses transpositions nationales, ainsi que les directives concernant le e-commerce et les bases de données, aboutissent à des juridictions beaucoup moins souples en matière d'exception au droit à la propriété intellectuelle que ne l'est la loi américaine (Band, 2008, p.27¹⁸¹). Cédric Manara va également dans ce sens, arguant que la liste limitative d'exceptions du droit européen ne prévoit pas

¹⁸⁰ « These exceptions could not be met as the reproductions in Google's cache interfered with the normal exploitation of the works because internet users were able to gain access to articles that had either been removed from the publishers' sites or were only available from the sites for a fee. Further, the Google News service deprived rights holders of reasonable compensation for new publications of their works » (Notre traduction).

¹⁸¹ La conclusion de Brand résume très bien son propos : "As Justice Ginsburg noted in *Eldred*, the fair use doctrine acts as one of copyright law's built-in accommodations to the First Amendment. Hence, it is particularly fitting that two of Google's endeavors that advance First Amendment values-its search engine and Library Project-depend on the fair use doctrine for their lawfulness. In contrast, European nations with less developed free speech traditions lack a fair use analog. These jurisdictions present a hostile copyright environment to Google's search engine and library project" (Band, 2008, p. 28).

de dérogation¹⁸² et que l'activité des moteurs de recherche « entre mécaniquement en conflit avec le droit d'auteur » (Manara, 2011, p.147).

6.1.5 Particularités françaises

La directive européenne de 2001 a été transposée en France par la loi du 1^{er} août 2006. Cette transposition étant relativement tardive, le texte était selon les experts décalé par rapport à ce que le web était devenu (Lucas et Sirinelli, 2006). La transposition donne une place centrale au juge qui doit trancher au cas par cas en appliquant le Triple Test. Ce dernier a par conséquent un pouvoir d'interprétation très large et peut refuser la protection d'une exception prévue par la loi dans un cas où il considère qu'il y a atteinte au titulaire du droit en question (Van Der Wielen, 2010). Ce pouvoir est contesté par certains auteurs qui se demandent si le Triple Test ne donnerait pas au juge un « pouvoir discrétionnaire de remettre en cause toutes les exceptions sur le fondement de son sentiment de l'équité » (Gaudrat, 2007). Cela aboutit à un flou jurisprudentiel (Geiger, 2007). Nous retrouvons le manque de prévisibilité dont nous avons parlé au sujet du droit étatsunien, mais cette fois dans un pays attaché au droit d'auteur, ce qui va clairement à l'encontre de Google en cas de litige mais n'en fait pas non plus nécessairement un perdant.

Par ailleurs, en France, la reproduction, selon l'article L. 122-3 du Code de la Propriété Intellectuelle, « consiste dans la fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public d'une manière indirecte ». Il s'agit bien de ce que la firme Google fait lorsqu'elle enregistre la totalité d'une page web sur ses propres serveurs et la rend disponible « en cache » (Peguera, 2009). La firme pourrait par conséquent être accusée de reproduction illicite.

A propos du système français, Françoise Benhamou et Joëlle Farchy soulignent : « Les différents éléments du droit moral tels que l'envisage le système juridique français ne sont pas tous affectés de la même manière. Le droit de divulgation, qui donne à l'auteur le pouvoir de décider le moment et le mode de

¹⁸² « La directive relative au droit d'auteur dans la société de l'information n'a pas non plus prévu d'exception qui bénéficierait spécifiquement aux moteurs de recherche pour l'exercice de leur activité. Si certains actes de reproduction transitoires ou accessoires qui constituent une partie intégrante et essentielle d'un procédé technique sont exemptés du droit de reproduction, c'est quand ils ont pour unique finalité de permettre une transmission dans un réseau entre tiers par un intermédiaire, ou une utilisation licite » (Manara, 2011, p. 153).

communication des œuvres au public, n'est pas touché ; sur le réseau comme ailleurs, il maîtrise la décision de diffuser son œuvre. De même, rien n'interdit à une œuvre numérisée d'être signée : le droit au nom et à la paternité peut donc être respecté. En revanche, le droit de repentir est plus délicat à appliquer. La densité des connexions rend irréversible le passage dans la sphère publique que constitue le réseau. C'est surtout le droit à l'intégrité qui se trouve affecté. La dématérialisation des supports constitue pour l'auteur une forme de désacralisation des œuvres puisque tout utilisateur peut la manipuler comme il l'entend. La numérisation ainsi que l'interactivité sur les réseaux rendent les œuvres aisément transformables » (Benhamou et Farchy, 2009, p. 59).

Ce sont ces deux derniers droits qui posent problème dans la relation de Google et des éditeurs de presse. Puisque le contenu reste disponible pour une durée indéterminée « en cache », le droit au repentir est remis en cause. Et puisque Google Search et Google Actualités placent un article à côté d'autres articles, d'autres liens vers d'autres sites, au sein d'une hiérarchie et d'une interface qui dépendent en partie des choix des ingénieurs, le droit à l'intégrité est lui aussi concerné.

Lors de différents procès, les porte-parole de Google ont présenté leurs services comme relevant du régime exceptionnel accordé en Europe aux revues de presse. Cependant, pour bénéficier d'une telle exception, la revue doit être un texte original comportant des critiques ou des réflexions. Dès lors, il semble que cela plaide en défaveur de Google, qui copie une partie du texte intacte tout en traitant de plusieurs thèmes regroupés en grappes. C'est pourquoi il ne peut s'agir ici de revue de presse, mais à la rigueur de ce qu'on appelle dans le droit français un « panorama de presse ». Dans le cas de la presse imprimée, un tel panorama ne peut être effectué sans qu'une redevance ne soit versée au *Centre Français d'exploitation du droit de Copie* (CFC). Le CFC n'a en revanche aucun mandat légal pour les panoramas de presse en ligne, à moins que ceux-là ne soient constitués de versions électroniques de contenus imprimés. Il semble donc y avoir une carence en ce qui concerne la presse en ligne : non seulement les contenus sont moins coûteux à copier intégralement ou partiellement, mais en plus ils ne bénéficient pas d'une protection semblable à celle des contenus papier.

Interrogés en juillet 2012 par la Haute Autorité pour la Diffusion des Œuvres et la Protection des droits sur Internet (Hadopi) dans le cadre d'une étude concernant les exceptions au droit d'auteur, les porte-parole de Google ont expliqué que la jurisprudence en matière d'exception pour revue de presse (L. 122-5 3° b) leur paraissait « limitée » et « devrait pouvoir couvrir toute activité d'information qu'elle qu'en soit la source » (Hadopi, 2012). Il a également été souligné par Google lors de cet entretien que : « Toute personne devrait pouvoir reprendre une ou plusieurs œuvres d'autrui à des fins d'information, dès lors que la reprise est limitée, et que cela n'a pas pour effet de concurrencer la jouissance économique des droits qu'ont des tiers sur ces œuvres » (*ibid.*).

6.1.6 Quel droit appliquer ?

Nous venons de voir qu'un flou juridique nimbait les deux côtés de l'Atlantique. Mais il se trouve que ce flou en Europe est plus défavorable pour la firme Google qu'il ne l'est aux Etats-Unis (Manara, 2011). Par ailleurs, l'Europe est aussi caractérisée par des transpositions différentes de la directive de 2001 dans les droits nationaux, ce qui n'est pas sans ajouter du brouillard au brouillard. En effet, les tribunaux en Espagne, en France, en Italie et en Belgique sont arrivés à des conclusions différentes quant à ce qui constituait ou non une infraction au droit d'auteur. Il semblerait par conséquent que Google ait intérêt à ce que l'action du dispositif relève du droit américain. Cela apparaît d'ailleurs clairement dans les réponses données par Google à l'Hadopi, où la firme milite pour que l'Union Européenne et la France adoptent le *fair use* :

« La doctrine du *fair use* est née de la reconnaissance de ce que la créativité et l'innovation ne peuvent être prescrites par des règles figées. L'objectif du *fair use* est d'encourager la créativité et l'innovation. [...]Le *fair use* permettrait d'éviter l'absence de flexibilité d'une liste arrêtée des exceptions. Il peut également être créateur de valeur, il a d'ailleurs été évalué que son existence a permis de générer 4,7 milliards de dollars de revenus aux Etats-Unis. L'activité de nombre d'entreprises de la société d'information, dont certains services offerts par Google, repose sur ce principe. Outre qu'une telle réforme simplifierait la compréhension des exceptions au droit d'auteur, elle serait aussi source d'opportunité économique (en plus de supprimer un désavantage concurrentiel pour les entreprises françaises face à leurs homologues américaines) » (Hadopi, 2012, p. 32-34).

Selon l'article 5.2 de la Convention de Berne, la loi qui doit être appliquée lors de la reprise d'un contenu est celle du pays où a lieu la reproduction finale de ce contenu. Dans le cas de Google, ce serait donc le droit du pays où se trouve l'internaute qui devrait être appliqué, dès lors que c'est l'action de celui-ci qui permet que la liste apparaisse sur son écran. Cependant, les infrastructures qui permettent la reproduction se trouvent en Californie, aussi un juge pourrait-il décréter que le droit américain prévaut (Sookman, 2011). Google a donc intérêt à enregistrer les contenus outre-Atlantique. Nous sommes là en présence d'un problème juridique lié au caractère distribué du fonctionnement du moteur : le contenu en ligne sur les serveurs de l'éditeur est crawlé et indexé sur les serveurs de Google, puis la liste hypertexte est générée par l'ordinateur, la connexion et le navigateur de l'internaute. Ainsi, le contenu qui apparaît à l'écran se trouve en quelque sorte dans plusieurs endroits à la fois, dont chacun peut dépendre d'une juridiction différente.

Dans la décision prise par la justice française en 2011 dans le cadre du procès opposant la Société Des Auteurs des Arts Visuels et de L'image Fixe Visual Auteurs (SAIF) à Google France et Google Inc., le juge s'en est référé au postulat de « traitement national » de la Convention de Berne, pour soumettre les contenus produits majoritairement par des Français à l'article 107 du *Copyright Act* étatsunien. Le juge a déclaré que l'usage des images par la firme relevait du *fair use*. Et il a décrété que Google France n'était pas concernée par cette affaire, parce que ses employés n'avaient aucun pouvoir d'administration sur le moteur. Pour se prévaloir du droit américain, Google a donc intérêt à poursuivre sa stratégie de structuration « en silos » (Smyrnaio, 2005), c'est-à-dire de conserver la société mère en Californie et d'implanter de simples satellites partout ailleurs, tandis que les infrastructures, où sont copiés et indexés les documents, demeurent aux Etats-Unis. Ainsi, comme le souligne Jonathan Band : « Ce n'est pas un accident si les plus importants moteurs de recherche du monde sont tous basés aux Etats-Unis ; le *fair use* fournit à leurs innovations un environnement légal bien plus fertile que ne le font les régimes prévoyant une palette d'exceptions spécifiques. Cependant, au fur et à mesure que Google et ses concurrents américains étendent leurs activités à l'échelle du globe, ils s'exposent à la possibilité de commettre des effractions répressibles ailleurs qu'aux

Etats-Unis. Et tandis que la doctrine *fair use* a permis la croissance explosive des moteurs de Google, Yahoo, Ask et MSN, l'environnement légal de l'Union Européenne s'est montré bien plus hostile vis-à-vis des moteurs de recherche »¹⁸³ (Band, 2008, p.27).

Nous voyons ici à la fois comment les aspects techniques du fonctionnement du moteur, et notamment le caractère coproduit et distribué de la liste, peuvent dérouter la jurisprudence et laisser libre cours à l'appréciation du juge, tandis que les lois en vigueur peuvent orienter les stratégies de Google, aussi bien sur le plan technique (infrastructures aux Etats-Unis) que sur le plan organisationnel (structuration en silos). Nous allons désormais nous intéresser à certains litiges ayant eu lieu en Union Européenne entre les éditeurs de presse et Google. Cela nous permettra de voir comment la firme peut conduire les éditeurs à coopérer malgré les pressions exercées sur son activité par les pouvoirs judiciaire et législatif.

6.1.7 Négociations en Union européenne

La relation entre Google et les éditeurs européens, en particulier les éditeurs de presse, est significative de la tension qu'il existe « entre la recherche de la diffusion la plus large des œuvres de l'esprit et la volonté de protéger, respecter et rémunérer la création par des dispositifs juridiques qui limitent de fait la diffusion » (Benhamou et Farchy, 2009, p. 4). Nous l'avons dit à plusieurs reprises, le référencement a une portée symbolique dans la mesure où les journalistes souhaitent être à l'avant-scène de l'espace public. Cependant, leurs employeurs sont tiraillés entre d'un côté la course à l'audience propre à leur modèle d'affaires (gratuit/publicité) et de l'autre le recouvrement de coûts fixes qu'ils sont seuls à supporter, tandis que leurs contenus sont utilisés par Google qui, en plus de siphonner le marché de la publicité, y tire les prix à la baisse. Dès lors qu'ils n'implémentent pas le robots.txt, les éditeurs ne font rien cependant pour empêcher que leurs

¹⁸³ « It is no accident that the world's leading search engines are all based in the United States; fair use provides a far more fertile legal environment for innovation than regimes with a handful of specific exceptions." However, as Google and its U.S. search engine competitors expand their operations globally, they will increasingly expose themselves to infringement liability overseas. While the fair use doctrine in the United States has allowed for the explosive growth of the Google, Yahoo, Ask, and MSN search engines, the legal environment in the European Union has been much more hostile to search engines. » (Notre traduction)

contenus soient indexés, et ce même si le moteur « porte [...] atteinte tant au droit de reproduction qu'au droit de communication au public des titulaires d'un droit d'auteur » (Strowel, 2011, p. 69). Toutefois, le tiraillement ne disparaît pas et nombre d'éditeurs cherchent des moyens d'obliger Google à supporter une partie de leurs coûts fixes. A travers plusieurs cas exemplaires, nous sonderons les négociations dont cette question a fait (et continue de faire) l'objet.

A. Copiepresse

Un procès emblématique en Europe a opposé la firme Google à 17 éditeurs de presse belges en 2006. Ces derniers, regroupés au sein de l'association *Copiepresse*, accusaient Google de reprendre une partie de leurs contenus pour générer des liens sur ses services sans les rémunérer. La firme s'est défendue en se prévalant de la décision de justice déjà prise pour un cas similaire aux Etats-Unis dans l'affaire l'opposant à *Blake A. Field*. Finalement, Google a été condamnée en septembre 2006 par le Tribunal belge à retirer de ses serveurs tous les articles ou morceaux d'articles, photographies et représentations graphiques appartenant aux éditeurs de *Copiepresse*, sous astreinte d'un million d'euros par jour de retard. Le 22 septembre 2006, la demande de Google de suspendre cette décision a été refusée par le Juge des référés. La firme, parce qu'elle copiait les pages en cache, a été explicitement considérée comme un « portail d'information » et non comme un simple moteur de recherche. Ses porte-parole ont argué en vain que le dispositif ne faisait que recopier le code des pages web et que, par conséquent, seul l'internaute qui affichait *effectivement* la liste hypertexte sur un terminal pouvait être considéré comme étant l'auteur de la reproduction (Van Asbroeck et Cock, 2007, p. 464).

Google a également tenté de se prévaloir de l'exception de revue de presse prévue par le droit belge, selon laquelle une exception au droit d'auteur peut être faite dans le cas où les contenus repris aux éditeurs sont courts et qu'ils servent à la critique, la polémique, l'éducation ou au travail scientifique. Mais la justice a rejeté cet argument en expliquant qu'aucune opinion n'était exprimée par l'algorithme, ce qui, nous l'avons vu, est faux, dans la mesure où l'algorithme d'appariement est un ensemble de critères et de pondérations définis selon ce que son concepteur pense à propos de ce qui rend un document pertinent, et donc de son opinion.

Google a tenté ensuite de se prévaloir de l'article 10 de la Convention Européenne des Droits de l'Homme qui garantit la liberté d'expression. Mais la décision de justice relève que « Copiepresse s'interroge, de façon pertinente, sur la question de savoir si on peut, en l'espèce parler d'exercice de son droit d'expression dans le cas de Google dans la mesure où le système mis en place par Google News n'a rien d'humain, que Google n'emploie aucun rédacteur en chef en vue de la sélection des articles [...] »¹⁸⁴ Nous retrouvons ici le problème posé par l'action manuelle d'un côté, issue d'une tradition chère aux médias traditionnels, et de l'autre le traitement algorithmique qui n'aurait, selon la justice belge, rien d'humain, et ce quand bien même des êtres humains auraient pensé et conçu le dispositif qui l'induit.

Enfin, Google a tenté de se prévaloir de « l'accord tacite » en mettant en avant le fait qu'il soit possible à tout moment pour l'éditeur de désindexer ses contenus grâce au protocole d'exclusion robots.txt. Mais la justice belge a statué que le droit d'auteur n'était pas en Europe un droit de ne pas être repris par un tiers, mais un droit qui interdisait à un tiers de reprendre un contenu sans l'autorisation *a priori* du propriétaire et de l'auteur du contenu en question (Van Asbroeck et Cock, 2007, p. 466). Nous retrouvons ici une remise en cause de l'accord tacite que constitue selon Google la non implémentation du protocole d'exclusion ou de balises méta-tags. Ce n'est pas, selon la justice, parce qu'il est possible pour l'éditeur d'exclure ses documents du champ d'indexation que celui-ci est d'accord, dans le cas où il ne ferait rien, pour que ses documents y soient inclus¹⁸⁵.

Finalement, la justice belge a décrété que les documents conservés par Google en mémoire cache permettaient aux internautes d'y accéder gratuitement, alors que ces mêmes contenus faisaient parfois partie des archives payantes des éditeurs, lesquelles génèrent des revenus que l'action de Google pouvait contribuer à amoindrir. Par ailleurs, le juge a estimé qu'il y avait violation du droit moral, puisque seulement une partie du contenu était reprise sans que l'avis de l'auteur ne soit demandé et que l'association de sa production à d'autres productions ne correspondait pas forcément aux opinions éditoriaux et moraux de l'auteur. Enfin, il a été

¹⁸⁴ http://cedriclaurant.files.wordpress.com/2010/12/copiepresse_google-070213.pdf (dernière visite le 21 août 2013)

¹⁸⁵ Nous reviendrons sur ce point en 6.3.

mentionné que la paternité des auteurs était bafouée puisque leurs noms n'apparaissaient pas sur Google Actualités.

L'Association Mondiale des Journaux (AMJ) a salué la décision belge en estimant par voie de communiqué que « les fournisseurs de contenu devraient avoir leur mot à dire sur la façon dont sont utilisées leurs productions »¹⁸⁶. Une *task force* a été mise en place à la suite du gain de cause de *Copiepresse* en première instance, regroupant plusieurs acteurs voyant là l'opportunité d'être entendus par les dirigeants de Google¹⁸⁷. Interrogé par l'AFP, le directeur du World Editors Forum (WEF, regroupant les rédacteurs en chef au sein de l'AMJ) Bertrand Pecquerie a expliqué : « Tous les journaux reconnaissent qu'il y a bien un service rendu, l'objectif n'est pas de condamner Google mais d'établir un partenariat [...] C'est une prise de conscience. On est dans un combat Europe/Etats-Unis, cela pousse Google à négocier »¹⁸⁸. Quant à la Fédération nationale de la Presse Française (FNPF), elle est allée dans le même sens en estimant qu'il était « nécessaire que [ce jugement] fasse jurisprudence dans l'ensemble de l'Union européenne »¹⁸⁹.

Après cinq années, la décision du tribunal de première instance fut confirmée en appel par la Neuvième Chambre belge le 5 mai 2011. Celle-ci a en effet estimé que même si l'affaire concernait des actes commis dans plusieurs pays, c'était le droit belge qui devait s'appliquer, cela car la Belgique était « la plus concernée ». La décision rejoint le cas opposant Google à la SAIF : même si le moteur indexe sur des serveurs en Californie et transmet les contenus depuis les Etats-Unis, ce serait, selon l'article 5.2 de la Convention de Berne, la loi du pays où les contenus sont consultés qui s'appliquerait. La Cour d'Appel a également confirmé qu'une partie, même infime, du contenu ne pouvait être reprise sans enfreindre le droit d'auteur. L'utilisation par Google des contenus empêchait selon elle les éditeurs d'en faire un usage normal, notamment lorsque des documents payants (archives) étaient disponibles gratuitement via le « cache » de Google. Google a eu beau se défendre en

¹⁸⁶ <http://technaute.cyberpresse.ca/nouvelles/internet/200609/19/01-13640-google-news-les-editeurs-de-presse-rassures.php> (dernière visite le 21 août 2013)

¹⁸⁷ <http://www.linformaticien.com/actualites/id/592/les-editeurs-de-presse-reclament-la-remuneration-du-contenu-de-leurs-articles-diffuses-sur-les-moteurs-de-recherche.aspx> (dernière visite le 21 août 2013)

¹⁸⁸ <http://technaute.cyberpresse.ca/nouvelles/internet/200609/19/01-13640-google-news-les-editeurs-de-presse-rassures.php> (dernière visite le 21 août 2013)

¹⁸⁹ Ibid.

disant que les internautes étaient les véritables auteurs de ces reproductions, la Cour d'Appel a conservé la même ligne que le tribunal de première instance. Enfin, la Cour a souligné que les droits moraux de l'auteur en matière d'intégrité et de paternité étaient effectivement violés : « Le service Google Actualités enfreint le droit de paternité parce que le nom des auteurs ne figure pas dans les résultats. Le droit d'intégrité a quant à lui été enfreint dans la mesure où seuls des extraits sont reproduits et le travail est modifié. La Cour a statué que même dans la mesure où un internaute sait, ou devrait savoir, que ce qu'il ou elle voit est seulement un extrait et qu'il est possible de le lire dans son intégralité en cliquant sur l'hyperlien, cela n'implique pas que l'auteur ait consenti à ce qu'un extrait de son travail puisse être publié »¹⁹⁰ (Sookman, 2011). Ce point est particulièrement intéressant, car nous pouvons voir dans le dispositif de « signature numérique » (cf. 4.3.6) une initiative de la part de Google visant à rétablir l'intégrité et la paternité des auteurs.

Le cas belge est emblématique car il illustre les points communs et les différences entre les transpositions de la directive européenne et le droit américain, entre le droit d'auteur et le *copyright*. Mais ce qui a suivi la décision de la Cour d'Appel s'avère plus intéressant encore. Google a procédé à l'éviction des sites de *Copiepresse* de l'index des moteurs Google.be, Google.com, Google Actualités Belgique, mais aussi d'autres sites comme Google.fr et Google.it. La firme a ainsi procédé à leur endroit à un biais sélectif et systématique. La désindexation fut perçue par *Copiepresse* comme une « application mesquine » de la décision de justice. « Ils l'ont fait pour nous punir », a déclaré la Secrétaire Générale de Copiepresse Margaret Boribon¹⁹¹. Selon les éditeurs, la mesure concernait essentiellement Google Actualités et n'aurait pas dû être étendue à l'ensemble de la galaxie des moteurs de Google. A cette critique, Google a répondu que le dispositif Google Actualités était intégré au dispositif Google Search depuis avril 2011, et que par conséquent l'indexation sur

¹⁹⁰ « The Google News service infringed the right of paternity because the authors' names were not mentioned in the search results. The right of integrity was infringed because only extracts of articles were reproduced and the works had been modified. The Court ruled that even though an Internet user knew, or should know, that what he or she was viewing was only an extract from an article and that it was possible to read it in its entirety by clicking on the hyperlink, that did not imply that the author had given his or her consent that only excerpts from the work could be published. » (Notre traduction)

¹⁹¹ « They have done it to punish us. They have a bad attitude. » (Notre traduction) <http://www.out-law.com/page-7427> (dernière visite le 21 août 2013)

l'un des deux dispositifs était fonction de l'indexation sur l'autre¹⁹². Ainsi, selon les porte-parole de la firme, appliquer la décision de justice revenait à désindexer les sites de *Copiepresse* de l'ensemble du dispositif¹⁹³. Google a explicitement demandé à ce que les éditeurs de *Copiepresse*, s'ils voulaient que les contenus soient réindexés, renoncent à réclamer l'astreinte de 25 000 dollars par effraction constatée et lui assurent que de nouvelles poursuites n'auraient plus lieu. C'est ce qu'ont effectivement fait les éditeurs, sans commenter leur décision. Les porte-parole de Google ont, quant à eux, commenté cette affaire en disant :

« Nous sommes ravis que Copiepresse nous ait donné aujourd'hui l'assurance de pouvoir inclure à nouveau leurs sites dans notre index de recherche Google sans devoir appliquer les pénalités ordonnées par la justice. Nous allons donc procéder à la ré-indexation de ceux-ci le plus rapidement possible. Nous n'avons jamais eu la volonté d'exclure les sites de notre index mais nous étions tenus de respecter le jugement du tribunal, à moins que Copiepresse n'intervienne. Nous demeurons ouverts pour travailler en collaboration avec les membres de Copiepresse à l'avenir »¹⁹⁴.

Nous constatons ici que les problèmes juridiques liés au droit d'auteur sont inscrits dans un tissu de rapports de forces socio-économiques qui dépasse largement le domaine de compétence d'une juridiction nationale. Même si Google a perdu son procès, la firme a réussi à annuler les effets concrets d'une décision de justice, étant donné les intérêts des éditeurs à être indexés sur ses moteurs. Certes, *Copiepresse* et la justice belge ont pu exercer des pressions sur Google à un moment donné, mais, finalement, la firme a repris le dessus étant donné le mécanisme d'intéressement/enrôlement/mobilisation qui lie l'action des éditeurs à la sienne.

B. AFP

Un autre cas emblématique de conflit entre la presse traditionnelle et la firme Google est celui des agences de presse internationales, et notamment de l'Agence France Presse (AFP). Ces agences n'ont en général pas de site ouvert au public. Elles divulguent leurs dépêches à leurs abonnés, qui payent une licence et peuvent ensuite

¹⁹² Ibid.

¹⁹³ Nous voyons très bien ici comment la controverse nous apprend de nouveaux éléments quant au fonctionnement technique du moteur, du moins à ce que veulent nous en dire ses concepteurs.

¹⁹⁴ <http://www.journaldunet.com/solutions/moteur-referencement/google-desindexe-copiepresse-le-soir-et-la-libre-belgique-0711.shtml> (dernière visite le 21 août 2013)

reprendre les documents en l'état ou en partie sur leurs propres sites. En mars 2005, l'AFP a accusé Google de reprendre sur son service Google Actualités le titre, l'accroche et parfois les images de dépêches pour lesquelles aucune licence n'avait été payée ni aucun accord donné. Le montant des dommages et intérêts demandés s'élevait à 17,5 millions de dollars. Le procès a eu lieu devant la Cour fédérale de Washington D.C., où les avocats de l'AFP ont mis en avant la doctrine « *hot news misappropriation* » (cf. 6.1.2). Les avocats de Google se sont défendus en mettant en avant l'incapacité de l'AFP à identifier quels étaient les contenus concernés et, par ailleurs, en arguant qu'étant donné le caractère extrêmement factuel des titres repris, ceux-là ne pouvaient faire l'objet d'un *copyright*.

Ce cas a ceci de particulier que puisque l'AFP fournit ses dépêches à un grand nombre de sites qui, eux, payent une licence sous forme d'abonnement, la demande d'un retrait des contenus appartenant à l'AFP aurait signifié un retrait des pages des sites qui utilisaient ces contenus. Autrement dit, tous les éditeurs utilisant les dépêches AFP auraient pu être pénalisés. Ainsi, l'observateur John Burke expliquait en 2006 : « Je ne crois pas que le quotidien français *Le Figaro* serait content si les lecteurs ne pouvaient pas accéder à un article trouvé sur Google Actualités à propos du décès d'un comédien sous prétexte que les photos apparaissant sur la page viennent de l'AFP. Ou que *Le Guardian* apprécierait si ses lecteurs ne pouvaient pas cliquer sur un article à propos du coup de tête de Zidane sous prétexte que la photo accompagnant l'article est accréditée AFP/Getty. »¹⁹⁵ En accusant Google, l'AFP mettait donc en péril sa relation avec ses clients historiques, lesquels risquaient, par peur de ne plus être indexés sur Google, de mettre un terme à leurs abonnements.

Finalement, en 2007, un accord fut trouvé et l'AFP abandonna ses poursuites. Selon cet accord, dont les termes exacts n'ont pas été communiqués, la firme paye un abonnement à l'AFP comme le font de nombreuses entreprises de presse. Un accord similaire a été trouvé avec l'agence américaine Associated Press (AP), l'agence britannique Press Association (PA) et l'agence Canadian Press (CP). Le PDG de

¹⁹⁵ « I don't think the French daily *Le Figaro* would be too happy if readers couldn't access its article found through Google News about a deceased French comedian because of the photos in the slideshow it offers come from AFP. Or that the *Guardian* would appreciate it if readers couldn't click through to an article about the Zidane headbutt because a picture next to the story is accredited AFP/Getty. », disponible ici : http://www.editorsweblog.org/analysis//2006/07/is_it_possible_to_take_afp_content_off_o.php#more (dernière visite le 21 août 2013)

l'AFP Pierre Louette a expliqué par voie de communiqué : « Cet accord avec plusieurs agences de presse exprime la reconnaissance par Google de la qualité de leur travail comme de l'originalité de chacun de leurs regards sur le monde »¹⁹⁶. Sur le blog de Google, on trouve :

« Parce que Associated Press, Agence France-Presse, UK Press Association et Canadian Press n'ont pas de sites où ils publient leurs contenus, ils n'ont pas bénéficié du trafic que Google Actualités renvoie vers les autres éditeurs »¹⁹⁷.

Dans le même communiqué, Google prévenait les éditeurs qui voudraient nouer un partenariat similaire que cela n'arriverait pas dans la mesure où ils ont un site web adressé au grand public et qu'ils bénéficient du trafic que leur envoie le dispositif d'infomédiation. Les dirigeants de la firme considèrent en effet le trafic comme un service rendu gratuitement en l'échange d'une permission d'indexation. Puisque la firme ne peut pas « payer » les agences avec du trafic, elle les paye avec de l'argent. Par ailleurs, deux autres arguments existent en faveur d'un tel accord. D'une part, cela permet à Google et à l'AFP de se prémunir : si Google doit désindexer tous les sites qui reprennent des dépêches de ces agences, cela aboutira à une perte de qualité pour le moteur, privé d'un grand nombre de contenus dont il a besoin et des effets de réseaux croisés afférents. D'autre part, grâce à cet accord, le dispositif peut mieux réagir en temps réel, parce qu'il prend connaissance en même temps que les journalistes des actualités traitées par l'AFP¹⁹⁸. Cela permet d'optimiser la constitution des grappes et l'identification des requêtes potentiellement liées à une actualité, ainsi que d'éviter de proposer à l'internaute trop de pages de sites différents menant vers un message qui est le même¹⁹⁹.

Enfin, notons que cet accord encourage la firme à faire remonter des documents issus de dépêches AFP. Car en effet, puisqu'elle paye son abonnement, la firme a le droit de reprendre ces contenus, alors que son action est parfois contestée lorsqu'elle reprend des contenus originaux. Cette incitation va dans le sens de ce que

¹⁹⁶ <http://www.generation-nt.com/google-actualite-agence-presse-afp-actualite-44628.html> (dernière visite le 21 août 2013)

¹⁹⁷ <http://googlenewsblog.blogspot.com/2007/08/original-stories-from-source.html> (dernière visite le 21 août 2013)

¹⁹⁸ Interrogé par Ken Auletta, un représentant de Google explique : « The agreement with the wire services means we'll be able to display a better variety of sources with less duplication. Instead of 20 "different" articles (which actually use the same content), we'll show the definitive original copy and give credit to the original journalist ». (Auletta, 2010, p. 165)

¹⁹⁹ Sur le blog de Google, toutes les améliorations permises par cet accord avec les agences sont expliquées : <http://googlenewsblog.blogspot.com/2007/08/original-stories-from-source.html> (dernière visite le 21 août 2013)

nous ont dit certains des acteurs interrogés dans le cadre de notre étude de terrain, selon lesquels Google privilégierait les reprises de dépêches AFP plutôt que les contenus originaux.

C. Editeurs italiens

La Fédération italienne des éditeurs de journaux (Fieg) a porté plainte en août 2009 auprès de l'Autorité italienne de la concurrence (AGCM) contre Google pour abus de position dominante (Macchiati, 2010). L'argument avancé par les éditeurs était lié cette fois à la déconstruction opérée par Google Actualités : dès lors que le dispositif pointait directement vers une page, il permettait de contourner, *via* un « lien profond », la page d'accueil du site, ce qui, selon la presse italienne, nuisait à leur activité dans la mesure où cette page devenait moins attractive pour les annonceurs. Par ailleurs, la Fieg reprochait à Google de les contraindre à accepter d'intégrer le champ de *crawling* de Google Actualités. En effet, selon la Fieg, lorsqu'un éditeur de presse reconnu indiquait à Google qu'il ne souhaitait pas intégrer le champ du moteur vertical, la firme lui signifiait qu'il ne pourrait pas intégrer dans ce cas le champ du moteur généraliste (pouvoir sur). Cela est étonnant car, comme nous l'avons expliqué dans le chapitre 4, ce n'est normalement pas de cette façon que Google procède : l'éditeur désireux d'intégrer Google Actualités doit lui-même en formuler la demande et répondre à certaines conditions, tandis que n'importe quel éditeur qui n'implémente pas le protocole « robots.txt » est éligible à l'indexation sur Google Search. Google se serait ainsi servi de sa situation dominante sur le marché des moteurs pour contraindre certains éditeurs aux contenus de qualité à se *mobiliser* davantage. La firme aurait alors traité différemment les éditeurs de presse en leur refusant un service fourni à d'autres éditeurs qui, eux, n'intègrent pas Google Actualités. Enfin, dans la même plainte, la Fieg reprochait à Google de ne pas divulguer les modalités de partages de revenus générés par les modules AdSense.

Sans attendre la décision des autorités italiennes, Google a tout de suite indiqué que les éditeurs qui choisissaient d'exclure leurs contenus du champ de Google Actualités le pouvaient, et que cela ne nuirait pas à leur référencement sur Google Search. La firme a expressément demandé à ceux-là de prendre contact avec elle. Par ailleurs, elle a révélé les modalités d'AdSense (cf. 5.2.1.B, Fig.15). En 2010,

L'AGCM a déclaré, à la suite de son enquête, qu'il serait désormais obligatoire pour Google de laisser les éditeurs choisir s'ils veulent ou non intégrer Google Actualités, et que cela ne pourrait pas avoir d'influence sur leur classement sur Google Search. Cela permettrait d'empêcher la firme d'agir comme elle l'avait fait en désindexant la presse belge de l'ensemble de ses services, prétextant de l'indissociabilité des champs. Google a pris cet engagement en Italie pour une durée de trois ans. L'AGCM a également recommandé au gouvernement italien de faire en sorte d'adapter le cadre législatif, jugé trop flou, de manière à ce que ce genre de problème ne puisse plus avoir lieu à l'avenir. « Une enquête antitrust ne peut pas régler le problème de la rémunération adéquate de l'activité des entreprises qui produisent des contenus éditoriaux – online – en cas d'utilisation des œuvres par d'autres sujets » a estimé l'autorité dans son communiqué²⁰⁰.

D. Editeurs allemands

En août 2012, le gouvernement allemand a rendu public le projet de loi surnommé « Lex Google » prévoyant que les moteurs de recherche et les agrégateurs devraient désormais rémunérer les éditeurs des contenus qu'ils citaient. Tandis que les grands groupes de presse allemands, comme *Axel Springer* et *Bertelsmann*, se sont félicités de ce projet, le syndicat professionnel allemand des technologies de l'information et de la communication (*Bitkom*) jugea que le gouvernement envoyait un mauvais signal à l'ensemble du secteur du numérique, tout en pointant des imprécisions quant aux services concernés par la loi²⁰¹.

Dans l'espoir que le Bundestag invaliderait le projet lors du vote qui aurait lieu fin novembre 2012, Google mit en ligne sur la page d'accueil de Google Search en Allemagne, un lien visant à faire adhérer les internautes à sa cause. Sur ce lien, il était écrit : « *Toi aussi tu veux trouver à l'avenir ce que tu cherches ? Défends ton réseau !* » L'internaute accédait, après avoir cliqué, à une page²⁰² sur laquelle le projet de loi était présenté brièvement, après quoi Google avançait les arguments suivants :

²⁰⁰ Traduction : AFP.

²⁰¹ http://www.bitkom.org/de/themen/55269_73263.aspx (dernière visite le 21 août 2013)

²⁰² <http://www.arretsimages.net/vite.php?id=14785> (dernière visite le 21 août 2013)

- Il n'y pas de publicité sur Google Actualités,
- Chaque éditeur décide s'il veut être référencé par Google Actualités ou non,
- En 2003, la cour constitutionnelle [allemande] a permis aux moteurs de recherche d'afficher des extraits d'articles.

Par ailleurs, les porte-parole de Google se sont largement prononcés en Allemagne pour alerter l'opinion publique. Le patron de Google Allemagne Stefan Tweraser a ainsi déclaré :

« Une telle loi affecterait chaque utilisateur d'Internet en Allemagne en privant d'information les consommateurs et en entraînant des coûts supplémentaires pour les entreprises. La plupart des citoyens n'ont encore jamais entendu parler de cette proposition de loi (...) alors que celle-ci pourrait concerner tous les internautes en Allemagne »²⁰³.

Nous trouvons ici une montée en généralité par laquelle le porte-parole de Google tente de se positionner comme le porte-parole de « tous les internautes en Allemagne ».

Google demanda aux internautes d'agir pour se rallier à sa cause en faisant pression sur les députés dans le but de faire avorter le projet de loi. Ainsi, une carte interactive de l'Allemagne fut mise en ligne, sur laquelle les internautes pouvaient trouver les prénom, nom, coordonnées (adresse postale, mail, fax, téléphone) de chaque député, pour chaque circonscription. Google demandait ainsi aux internautes de contacter les députés pour les inciter à voter non au projet de loi.

Figure 29. Carte interactive du lobbying contre la loi allemande *Lex Google*



²⁰³ <http://www.zdnet.fr/actualites/lex-google-google-tente-de-rallier-les-internautes-allemands-39784880.htm> (dernière visite le 21 août 2013)

Google a voulu utiliser l'association des internautes à son cours d'action, en les mobilisant dans le but de peser sur les éditeurs (selon le cabinet AT Internet, les parts de Google sur le marché des moteurs en Allemagne étaient de 92% en décembre 2012). Il s'agit d'une traduction du *script* de Google, qui crée de nouvelles page, une carte, un discours — certaines inscriptions qui ont modifié, alourdi et transformé son script initial (Latour *et al.*, 1991) — afin de peser dans le processus de négociation et de pousser les autres acteurs du réseau à agir dans le sens qui lui convient. La firme tentait ainsi de *mobiliser* chaque utilisateur de son moteur de manière à ce qu'il défende son programme d'action contre les « anti-programmes » du gouvernement et des éditeurs de presse en ligne.

Les grands groupes de presse, membres de la fédération des éditeurs de journaux et de l'association des éditeurs de magazines, ont aussitôt accusé Google d'abus de position dominante et de « propagande haineuse »²⁰⁴. L'initiative a également été très mal perçue par les députés. Par exemple, pour Günther Krings et Ansgar Heveling, la campagne de Google constituait une tentative visant à « mobiliser les utilisateurs pour faire du lobbying pour ses propres intérêts, au nom de la liberté sur Internet [...] »²⁰⁵.

Dénonçant l'impérialisme de Google, une partie de la presse allemande a vivement critiqué la firme. Ainsi, par exemple, selon le quotidien conservateur *Frankfurter Allgemeine Zeitung* : « Google nous a enfin dit ce qu'il est. Google c'est nous. Nous avons les questions, Google livre les réponses. Nous fournissons les données que Google transforme en argent. Google appelle cela la liberté, mais pour nous, c'est le capitalisme de la Silicon Valley, dans lequel une entreprise mondiale fait passer son intérêt commercial pour un droit fondamental de l'utilisateur. »²⁰⁶ Le quotidien renverse ici la montée en généralité : plutôt que de dire « Google nous représente », il prétend que « Google, c'est nous », et rappelle aux internautes que Google est riche grâce à eux, ce qui ne signifie pas que les internautes doivent faire une confiance aveugle à Google mais au contraire que la firme doit

²⁰⁴ <http://www.arretsurlimages.net/vite.php?id=14785> (dernière visite le 21 août 2013)

²⁰⁵ <http://www.lefigaro.fr/medias/2012/11/29/20004-20121129ARTFIG00703-presse-l-allemande-se-defend-face-a-l-offensive-de-google.php> (dernière visite le 21 août 2013)

²⁰⁶ <http://www.lefigaro.fr/medias/2012/11/29/20004-20121129ARTFIG00703-presse-l-allemande-se-defend-face-a-l-offensive-de-google.php> (dernière visite le 21 août 2013)

rendre des comptes aux internautes et respecter la procédure législative du pays dans lequel elle se trouve. En disant aux citoyens « Google devient riche grâce à vos données », le *Frankfurter Allgemeine Zeitung* se sert de son statut pour essayer de ranger ses lecteurs de son côté. Il s'agit ici d'un exemple où c'est parce que le fonctionnement du dispositif agrège les cours d'action de plusieurs acteurs que la possibilité de parler au nom de l'ensemble des membres enrôlés dans le dispositif est contestée aux porte-parole de Google. En disant « Google, c'est nous », les députés et les journalistes s'érigent, eux, comme porte-parole du dispositif.

Finalement, la loi est passée au Bundestag et devrait entrer en vigueur à l'automne 2013. Il est intéressant de constater qu'en plus d'avoir provoqué de très nombreuses réactions contre elle de la part des députés et des médias, étant donné le dispositif de lobbying mis en place, Google n'a pas reçu de la part des internautes le soutien escompté. Il semble en effet que les internautes aient été plutôt indifférents à ses appels²⁰⁷. Ainsi, nous voyons que le pouvoir d'enrôlement/mobilisation de Google est limité : ce n'est pas parce que les internautes utilisent son moteur et ses services qu'ils sont prêts à agir comme la firme le leur demande. Ils ont beau être *intéressés*, ils n'acceptent pas le rôle qui leur est attribué par Google et la définition qui est donnée par la firme de leur identité : le phénomène d'*enrôlement/mobilisation* est limitée. Le niveau d'implication des internautes dans le dispositif ne signifie donc pas qu'ils se reconnaissent dans les valeurs que la firme prône et qu'ils sont prêts à défendre ses intérêts. C'est pourquoi la montée en généralité a échoué, empêchant Google d'aboutir au « faire faire » escompté. Il reste à voir ce qu'il se passera à l'automne, lorsque la loi entrera en vigueur. A la lumière de ce que nous avons expliqué à propos de *Copiepresse*, nous pouvons imaginer que Google menacera de désindexer les contenus des éditeurs qui lui demandent de payer, et ce même si la loi les autorise désormais à le faire. Certains éditeurs, par peur d'être exclus du champ d'indexation, n'oseraient alors pas demander à Google des sommes qui pourtant leurs seraient dues.

²⁰⁷ Par exemple : <http://www.lefigaro.fr/medias/2012/11/29/20004-20121129ARTFIG00703-presse-l-allemande-se-defend-face-a-l-offensive-de-google.php> (dernière visite le 21 août 2013)

6.2 France : entente cordiale

En France, les éditeurs de presse en ligne ont observé attentivement les cas belges, italiens et allemands, et mené leurs propres négociations avec Google. L'objet de cette sous-partie est de revenir sur ces négociations et sur l'issue récemment (et provisoirement) trouvée.

6.2.1 Le meilleur ennemi

En 2003, Philippe Jannet, directeur du site des éditions électroniques des Echos et président du Geste (Groupement des éditeurs de services en ligne), envoya une lettre à Google France, priant la firme de bien vouloir entamer des discussions avec les éditeurs français afin qu'un accord puisse être trouvé quant aux modalités de reprise des contenus. Mais comme l'a expliqué Nikos Smyrnaiois dans sa thèse, « à cette époque l'équipe de Google France était constituée uniquement des commerciaux qui étaient chargés de vendre les offres publicitaires de la société. Les responsables de Google France n'avaient aucune prise sur le service d'actualité et ne connaissaient pas les modalités de son fonctionnement, puisque celui-ci était géré directement par les équipes américaines. De plus, dans le fonctionnement « matriciel » de la société, chaque responsable local répond directement à son supérieur hiérarchique aux Etats-Unis et non pas au directeur de la division du pays en question. Par conséquent il n'y avait pas en France d'interlocuteur qualifié pour répondre aux interrogations des éditeurs au nom de Google » (Smyrnaiois, 2005, p. 299). Même si, à la suite de cette lettre, les représentants du Geste ont rencontré le directeur de Google France, le responsable des contenus au niveau mondial et un avocat, aucun accord n'a été trouvé et les éditeurs ont clairement eu le sentiment de ne pas avoir été entendus.

Le rapport de forces était alors en faveur de Google, dont les dirigeants savaient que l'audience apportée aux sites de presse était trop cruciale pour que ces derniers envisagent d'être désindexés. C'est l'équivalent en économie d'une *demande inélastique* : lorsque le producteur sait que les consommateurs sont prêts à payer n'importe quel prix ou presque, il exerce un pouvoir très fort en cas de négociation. Ainsi, par exemple, les dirigeants du site *Lemonde.fr*, qui avaient demandé en 2004

que leurs pages soient retirées de l'index, ont fait volte-face en demandant à Google que leurs contenus soient réintégrés.

Quelques années plus tard, la tension n'avait toujours pas disparu. Le responsable de Google Actualités Josh Cohen fut très mal accueilli lors de son intervention aux Etats généraux de la presse de décembre 2008. « Nous vous considérons comme un danger pour notre existence, pour nos entreprises, lui a expliqué Pierre Conte, directeur général adjoint du groupe *Figaro*. En l'espace de quelques mois, l'ami Google est devenu le danger Google dans le domaine publicitaire. Vous êtes devenu notre pire ennemi. Nous ne demandons pas la pitié, mais un meilleur partage des revenus. C'est comme cela que nous comprenons notre partenariat »²⁰⁸. Et Josh Cohen de répliquer : « Nous ne pouvons pas faire en sorte que les gens achètent davantage vos journaux. Votre modèle est en train de changer, il va falloir vous adapter »²⁰⁹. Enfin, Bruno Patino, ex-président du *Monde Interactif* et nouveau directeur général de *France Culture*, de rétorquer : « L'écosystème des *news* est en train de mourir. Et ce que nous entendons aujourd'hui signifie que nous sommes livrés à nous-mêmes, malgré les déclarations d'Eric Schmidt. Vous acceptez donc la fin des *news* comme nous les avons connues »²¹⁰.

L'un des objectifs des éditeurs était explicitement d'être moins dépendants de Google en diluant son pouvoir de *gatekeeping*. C'est pourquoi certains d'entre eux ont créé le kiosque « *e-presse* », rassemblant les acteurs du groupement d'intérêt GIE-ePresse : *Le Parisien-Aujourd'hui en France*, *Le Figaro*, *Libération*, *L'Equipe*, *Les Echos*, *Le Nouvel Observateur*, *L'Express* et *Le Point*. Cette application web avait pour but de proposer une plate-forme sur laquelle les éditeurs pouvaient commercialiser leurs contenus : articles et exemplaires vendus à l'acte et l'abonnement. Les utilisateurs avaient accès aux sommaires et pouvaient acheter un article ou un titre « en un clic ». Parallèlement, les membres du GIE mirent en place un partenariat avec *Orange* et son service d'agrégation : 24/24 Actu, qui rémunérait les propriétaires de contenus. Mais *Orange* a finalement fermé son portail en 2011 en

²⁰⁸ <http://www.geste.fr/spip.php?article203> (dernière visite le 21 août 2013)

²⁰⁹ <http://www.20minutes.fr/article/569955/Economie-Google-en-panne-d-algorithme-politique.php> (dernière visite le 21 août 2013)

²¹⁰ http://blogs.afp.com/mediawatch/post/2008/12/11/Etats-Generaux-%3A-Google-en-position-daccuse_ (dernière visite le 21 août 2013)

raison d'une « audience trop faible pour assurer une viabilité suffisante sur internet »²¹¹.

Fin 2012, la tentative fut relancée par les mêmes acteurs. « Presse.lemoteur.fr » fut créé par *Orange*, l'entreprise espérant cette fois attirer davantage d'internautes. Ce moteur restreint son champ d'indexation aux contenus des huit entreprises du GIE, ce qui permet une fréquence de *crawling* élevée. Les revenus publicitaires générés par le moteur sont quant à eux redistribués équitablement entre les huit éditeurs et *Orange*. De telles mesures semblent prendre le contre-pied de Google Actualités. C'est en tout cas de cette façon que les observateurs la présentent : « un tacle envers Google à peine caché »²¹².

6.2.2 Projet de loi

Au moment où Orange et le GIE lançaient leur moteur, fin 2012, les tensions entre Google et les éditeurs français se sont ranimées. Certains des éditeurs—rassemblés via L'Association de la presse d'information politique et générale (IPG), le Syndicat de la presse quotidienne nationale et le Syndicat des éditeurs de la presse magazine — ont demandé au gouvernement de créer une loi semblable à celle qui existe dans l'industrie de la musique, laquelle conduirait, grâce à une révision du Code de la propriété intellectuelle (L. 218), à la création d'un droit voisin au droit d'auteur qui permettrait d'obliger les infomédiaires, Google en tête, à rémunérer ceux dont le contenu serait repris. L'IPG mentionnait explicitement dans son projet le cas allemand comme un exemple à suivre.

Le projet (auquel nous avons eu accès) indique que les éditeurs ne peuvent pas s'opposer à ce que des liens hypertextes soient faits vers leurs contenus mais que, lorsque c'est le cas, une « rémunération équitable peut être versée », et en particulier « lorsque le lien est utilisé par un prestataire d'un service de référencement ou l'exploitant d'un moteur de recherche ». D'autre part, le projet précise : « il est indifférent que le service soit offert à titre onéreux ou à titre gratuit ». Ainsi, même si Google ne tire pas directement de revenus de la liste de liens naturels et de Google

²¹¹ <http://www.orangeinfo.fr/19997-orange-fermer-2424-actu.html> (dernière visite le 21 août 2013)

²¹² <http://www.pcinpact.com/news/76094-orange-devoile-google-actualites-maison-avec-seulement-huit-journaux.htm> (dernière visite le 21 août 2013)

Actualités, la firme aurait dû, si la loi avait été votée et appliquée en l'état, rétribuer les éditeurs. « De même, précisait le projet, il est également indifférent que ces prestataires ou exploitants aient eu ou non un rôle actif, notamment d'éditeur leur conférant une connaissance et un contrôle des liens. Il suffit que les liens hypertextes visent manifestement le public français. » Ces précisions sont apportées pour se prémunir du caractère automatisé du moteur et de l'absence de contrôle de la part des employés de Google. Enfin, le projet de loi stipulait qu'« afin que la rémunération équitable soit raisonnablement cantonnée, elle n'est due que lorsque les liens sont l'objet unique d'un service ou, en cas d'une offre plurale de services, lorsque le service qui propose des liens occupe une place importante et substantielle ». Nous avons vu dans le chapitre 5 pourquoi et comment le moteur était au centre des activités de la firme. Dès lors, il semble que de telles stratégies, et bien sûr celle de Google en particulier, aient été directement visée par ce projet. Dans le cas où le projet aurait effectivement modifié le Code de la Propriété Intellectuelle selon ses termes initiaux, la rémunération des éditeurs aurait été calculée sous le contrôle d'une société de perception et de répartition agréée par le Ministère de la culture.

« Nous ne sommes pas en capacité de négocier seuls face à ce géant, expliquait Nathalie Collin, présidente de l'IPG, après avoir remis le projet de loi au gouvernement. Dès lors, nous avons préféré trouver une solution juridique en élaborant ce projet de loi et nous avons recherché le soutien des pouvoirs publics. [...] Nous voulons que cette rémunération soit indexée sur la croissance du marché de la publicité sur les moteurs de recherche qui croît d'environ 7 %, alors que dans le même temps la publicité par bannière baisse de 6 %. »²¹³ En échange, les éditeurs s'engageaient à ne pas utiliser le robots.txt pour désindexer leurs contenus.

Mais tous les éditeurs français n'étaient pas d'accord avec le projet. Notamment, les sites nés en ligne, représentés par le *Syndicat de la Presse indépendante d'information en ligne* (SPIIL), pensaient que « que la proposition de loi présentée [...] en France par le Syndicat de la presse quotidienne nationale

²¹³ <http://www.lefigaro.fr/medias/2012/10/28/20004-20121028ARTFIG00166-j-espere-que-google-a-des-propositions-concretes.php> (dernière visite le 21 août 2013)

(taxation de l'indexation des contenus de presse) constitu[ait] des solutions inappropriées »²¹⁴. Ils estimaient notamment que :

1. « Pour être visibles sur la toile, les éditeurs de presse sont largement dépendants du référencement de leurs informations par Google. Cela provoque une uniformisation des contenus, les mêmes mots-clés et les mêmes thèmes étant traités en même temps par les éditeurs qui tentent d'arriver en tête des pages de recherche de Google. Cette course effrénée à l'audience entraîne un affaiblissement de la qualité des informations, et tue l'indispensable diversité de la presse.
2. Si les éditeurs sont un jour rémunérés également par des taxes versées par Google, ils seront soumis à une double dépendance : dépendance pour l'audience et dépendance pour les recettes. »²¹⁵

Johan Hufnagel, rédacteur en chef de *Slate.fr*, publia une tribune dans le quotidien *Libération*, expliquant que selon lui le projet de loi visait « juste à prolonger artificiellement des modèles économiques à bout de souffle et des groupes de médias détenus pour la plupart par des géants du capitalisme, français cette fois, de Dassault à Niel en passant par Arnault, Rothschild, Pinault et Perdriel »²¹⁶. Plus loin, il écrivait : « si Google est en position dominante, à qui la faute ? Combien de formations pour journalistes s'intitulent « écrire pour Google », combien de missions confiées aux experts du SEO [...], et à quel prix, dans le seul but d'être plus gros que le voisin ? Donner accès à son contenu, *via* Google, en respectant certaines des règles du moteur n'est pas honteux. Après tout, si les lecteurs l'utilisent comme un kiosque, il serait stupide de refuser d'aller à leur rencontre. A ce petit jeu, nos confrères de l'Obs, du Huffington Post, du Figaro ne sont pas les plus mauvais. Bien au contraire. Vouloir reprendre un jour ce qu'on a donné la veille s'apparente à de la pure mauvaise foi. »²¹⁷ Johan Hufnagel pointait ainsi du doigt le fait que, par leurs actions, les éditeurs se soient associés librement à l'action de Google, et cela en sachant très bien quels seraient les inconvénients d'un tel partenariat. Selon lui, cette association, et son renforcement par des pratiques de SEO, est justifiée par l'utilisation massive du dispositif par les internautes : à la manière des porte-parole de Google eux-mêmes,

²¹⁴ <http://www.spiil.org/20120913/spiil-contre-taxe-google> (dernière visite le 21 août 2013)

²¹⁵ *Ibid.*

²¹⁶ <http://www.slate.fr/tribune/64021/taxe-google-presse-patrons> (dernière visite le 21 août 2013)

²¹⁷ *Ibid.*

il attribue la légitimité du « pouvoir sur » et du « pouvoir de faire faire » de Google à l'ampleur des effets de réseaux que la firme génère.

Parmi les sites nés en ligne, le site Rue89 s'est retiré du Spiil dont ses fondateurs étaient pourtant avec d'autres à l'origine de la création en 2009. Poussé par un de ses propriétaires — *Le Nouvel Observateur*, dont Nathalie Collin, citée ci-avant, est directrice générale — Rue89 a ainsi rejoint les rangs des sites issus des médias traditionnels dans l'opposition à Google. Une controverse peut ainsi reconfigurer certains réseaux d'acteurs groupés autour d'intérêts communs.

Pour sa part, Philippe Jannet, directeur général du GIE-ePresse et favorable au projet de loi, pensait qu'il fallait porter le débat sur Google Search et non pas seulement sur Google Actualités :

« Le problème ce n'est plus Google Actualités. Google a d'ailleurs été très malin à ce sujet, car on s'est tous focalisé sur ce service. Or son argument, c'était qu'il ne monétisait pas le service, puisqu'il n'y avait pas de publicité. Mais regardez l'audience de Google Actualités : elle ne cesse de baisser depuis trois ans. Tout simplement parce que Google, pour fournir un meilleur service sur la recherche en ligne, a déporté les actualités dans son moteur, monétisé celui-ci. Il a ainsi amélioré son chiffre d'affaires, grâce au travail de la presse. Google a géré la décroissance. Il faut déplacer le débat sur Google Search, et l'obliger à négocier. »²¹⁸

Google a menacé, en envoyant plusieurs courriers aux cabinets ministériels français, de déréférencer l'ensemble des éditeurs qui soutiendraient le projet. La firme a mis ainsi en avant son « pouvoir sur » afin de faire pression sur les éditeurs. Selon elle, le projet de loi mettait « en cause son existence même » et limitait « l'accès à l'information ».

« En soumettant le référencement à rémunération et en punissant le défaut de versement de celle-ci de trois ans d'emprisonnement et de 300 000 euros d'amende, (le texte) ne ferait que multiplier les conflits et ralentir Internet, prévient un porte-parole de Google. Nous pensons qu'une loi telle que celles proposées en France et en Allemagne serait très dommageable pour Internet. Ce n'est pas un secret, cela fait maintenant trois ans que nous le disons publiquement »²¹⁹.

²¹⁸ <http://www.zdnet.fr/actualites/philippe-jannet-la-presse-francaise-doit-obliger-google-a-negocier-39775786.htm> (dernière visite le 21 août 2013)

²¹⁹ <http://www.lefigaro.fr/hightech/2012/10/18/01007-20121018ARTFIG00718-google-menace-de-ne-plus-referencer-la-presse-francaise.php> (dernière visite le 21 août 2013)

Lorsque Google a rendu publique cette menace de désindexation, les associations d'éditeurs se sont déclarées « stupéfaites » par la réaction de Google, « qui n'hésite pas à s'ériger en censeur au mépris de sa mission d'intérêt général de référencement »²²⁰ et ont tenté d'enrayer la montée en généralité en prétendant que Google jouait contre l'intérêt général. Elles ont estimé que la « tentative d'étouffer tout débat par la menace [était] clairement un déni de démocratie »²²¹. « Je suis un peu surprise par le ton de cette correspondance, qui s'apparente à une menace, a réagi auprès de l'AFP la ministre [de la Culture] Aurélie Filippetti²²². Ce n'est pas avec des menaces qu'on traite avec un gouvernement démocratiquement élu ». La Ministre effectuait ainsi sa propre montée en généralité en rappelant à Google qu'elle représentait, elle, un gouvernement élu. Il y avait ainsi concurrence des tentatives de porte-parolat. D'après la Ministre, « ce qui est indispensable, c'est qu'il y ait une forme de rétribution par les sites qui aujourd'hui tirent un profit réel de l'utilisation des contenus riches en information, en savoir faire, en matière grise. Il faut qu'ils participent au financement de l'information »²²³.

Etant donné ce que nous avons dit dans le chapitre 5, il ne faut pas perdre de vue que Google a besoin des contenus des éditeurs si la firme veut optimiser l'ensemble de sa chaîne de création de valeur. En effet, les effets de réseaux font que les contenus produits par les éditeurs à l'origine du projet de loi fortifient, sans qu'on puisse déterminer à quel point, l'ensemble du réseau articulé par Google. Ainsi l'observateur Dan Sullivan, lorsqu'il commenta le projet de loi français, expliqua que si Google était en effet en mesure de mettre ses menaces à exécution, la firme en pâtirait, car elle n'aurait alors « plus de bonne réponse à lier, et pourrait décevoir les internautes »²²⁴.

²²⁰ *Ibid.*

²²¹ *Ibid.*

²²² *Ibid.*

²²³ <http://www.aqui.fr/societes/interview-aurelie-filippetti-laquo-l-avenir-de-la-presse-passe-par-le-numerique-raquo,7276.html> (dernière visite le 21 août 2013)

²²⁴ <http://www.lefigaro.fr/hightech/2012/10/29/01007-20121029ARTFIG00573-google-et-la-presse-les-cles-du-debat.php> (dernière visite le 21 août 2013)

6.2.3 Accord : 1^{er} volet

Finalement, le vendredi 1^{er} février 2013, Google signa un accord avec la presse française sous le patronage de l'Etat, prévoyant le versement d'un fonds de 60 millions d'euros pour aider la presse d'information. L'accord mit un terme au projet de loi, et constituait un « pas de côté » pour les deux camps. Les entreprises concernées par l'accord sont les sites internet d'information générale, les quotidiens nationaux, la presse magazine et la presse quotidienne régionale. La presse de divertissement en est exclue. Chaque entreprise concernée doit déposer un dossier pour justifier de l'aide qu'elle recevra.

Le Président de la République, François Hollande, a parlé dans sa conférence de presse d'un « accord historique » et d'une « première mondiale ». Marc Schwartz, associé au cabinet Mazars, qui avait arbitré la négociation entre Google et l'IPG, indiqua quant à lui que le contenu de l'accord ne serait pas rendu public et expliqua que l'enjeu consistait à « réussir la rencontre entre deux mondes, l'un qui concerne la technologie (Google), l'autre qui concerne les contenus (la presse) et rappeler notamment que la presse apporte au web trois atouts essentiels : la fréquence, la crédibilité, et la qualité de ses contenus. »²²⁵

Au sujet de l'accord, les avis sont partagés. Emmanuel Berretta, chroniqueur au *Point*, regrette qu'un mécanisme reproductible d'année en année n'ait pas été mis en place et s'étonne : « Qu'est-ce que 60 millions d'euros pour une entreprise qui, rien qu'en France, récolte 1,5 milliard d'euros de revenus publicitaires par an ? »²²⁶ Louis Dreyfus, Président du Directoire du journal *Le Monde*, estime pour sa part que même si l'accord est bénéfique, cela ne résoudra certainement pas la crise économique de la presse²²⁷. Certains éditeurs sont quant à eux furieux que la création d'un droit voisin, prévue par le projet de loi, n'ait finalement pas eu lieu. Le site *Numerama* parle quant à lui d'une « arnaque intellectuelle » et critique le caractère

²²⁵ <http://www.challenges.fr/media/20130203.CHA5842/comment-google-et-la-presse-ont-fini-par-s-entendre.html>

²²⁶ http://www.lepoint.fr/chroniqueurs-du-point/emmanuel-berretta/presse-google-s-en-sort-par-un-cheque-de-60-millions-d-euros-01-02-2013-1622962_52.php (dernière visite le 21 août 2013)

²²⁷ http://www.lepoint.fr/chroniqueurs-du-point/emmanuel-berretta/louis-dreyfus-l-accord-avec-google-ne-sauvera-pas-la-presse-02-02-2013-1623139_52.php (dernière visite le 21 août 2013)

« obscur » et « privatisé » de l'accord, qui ne permet pas de savoir à qui et comment, à quelles conditions, le fonds sera redistribué :

« [L]à où une loi votée démocratiquement par le Parlement aurait nécessairement créé un dispositif ouvert à l'examen public, soumis à des règles de principe garantissant indépendance, équité et transparence dans la distribution des fonds, l'accord signé par Google est des plus opaques sur les clés de redistribution. »²²⁸

Le Spiil, qui était contre la création d'un droit voisin, a publié un communiqué pour dire que même si l'accord était scellé entre des partenaires privés, et pouvait à ce titre demeurer confidentiel, il serait préférable, étant donné l'implication du gouvernement, de le rendre public. Dans son communiqué, le Spiil expliquait :

« Il est question aujourd'hui d'une aide privée soutenue par les pouvoirs publics, d'un montant de 60 millions d'euros, soit exactement le montant de l'aide attribuée par l'Etat au développement de toute la presse numérique en trois ans, entre 2009 et 2011 (Fonds SPEL), et six fois l'aide qu'il a accordée en 2012 (10 millions € pour la section 2 du nouveau Fonds stratégique). Au même titre que les aides publiques, le fonctionnement de ce fonds soutenu par les pouvoirs publics doit être transparent : la composition de la commission d'attribution des aides doit être connue rapidement ; les critères d'attribution doivent être divulgués ; les projets financés, leur montant et les bénéficiaires doivent être rendus publics. La question du cumul par une catégorie restreinte de bénéficiaires d'aides publiques et d'aides privées concourant aux mêmes objectifs devra également être posée. Le Spiil sera particulièrement vigilant sur les fortes distorsions de concurrence que de tels mécanismes sont susceptibles de provoquer. »²²⁹

Enfin, en Allemagne, où nous avons vu qu'un débat semblable avait lieu, l'Association des éditeurs de magazines (*Verband Deutscher Zeitschriftenverleger*) et l'Association des éditeurs de presse (*Bundesverband Deutscher Zeitungsverleger*) ont déclaré que l'accord français n'était pas un modèle pour eux, cela notamment car ce n'était pas juste dès lors que d'autres moteurs/agrégateurs se livrant aux mêmes types d'action que Google ne seraient pas soumis au versement d'un fonds.

Même si certains observateurs dressent des listes (selon le consultant Erwann Gaucher²³⁰, 167 éditeurs seront concernés par l'accord), nul ne sait avec

²²⁸ <http://www.numerama.com/magazine/24992-l-accord-google-presse-ou-l-arnaque-intellectuelle-du-gouvernement.html> (dernière visite le 21 août 2013)

²²⁹ <http://www.spiil.org/20130203/spiil-demande-l-accord-google-ipg-soit-rendu-public> (dernière visite le 21 août 2013)

²³⁰ <http://www.erwanngaucher.com/05022013Quels-medias-vont-pouvoir-beneficier-des-60-millions-de-Google-media?a=1066> (dernière visite le 21 août 2013)

certitude quel éditeur touchera une aide et quels seront les montants exacts des aides versées : seront-ils égaux pour tous les éditeurs ? certains éditeurs seront-ils privilégiés ? que devront faire les éditeurs (quels détours) afin d'être éligibles ? Par ailleurs, d'autres questions sont soulevées, à propos notamment du rôle joué par Google dans le choix des éditeurs éligibles. En effet, selon Sandrine Cassini²³¹, Google siègera au conseil de gestion de ce fonds, et donc aura un pouvoir quant à la décision d'accorder ou non une aide à un éditeur en particulier. Cela constituerait un nouveau « pouvoir sur » attribué à Google, et pourrait nourrir d'autant plus son « pouvoir de faire faire ».

6.2.4 Accord : 2^{ème} volet

Dans le cadre de l'accord français, en plus de la création du fonds, les porte-parole de Google ont annoncé que la firme aiderait les éditeurs à monnayer leurs revenus grâce à ses technologies publicitaires. Cela nous amène à la question des conflits d'intérêt qui peuvent avoir lieu. Car en effet, il est légitime de se demander à quel point l'obtention de l'aide sera conditionnée au fait que les éditeurs soient des clients d'une des régies de Google. Cela pourrait constituer une distorsion concurrentielle à l'avantage de la firme, qui étendrait ses réseaux de syndication AdSense et DoubleClick grâce à son monopole sur le marché des moteurs et à sa position de mécène, et fortifierait ainsi la « spirale de diffusion » (cf. 5.2.4).

Dans un article publié sur le site du Geste en mai 2012²³², Philippe Jannet, Président du Geste et président-directeur général *Monde Interactif*, et Emmanuel Parody, éditeur de CBS France, estimaient que l'essentiel de la bagarre entre Google et les éditeurs se situait désormais sur le terrain de la publicité. Selon eux, « le rachat de DoubleClick par Google ne cesse de créer des distorsions de concurrence entre le géant de Mountain View et les éditeurs ». Ils pointent notamment du doigt le fait que les contrats AdSense aient évolué, proposant désormais aux éditeurs partenaires « un petit pourcentage supplémentaire s'ils acceptent le ciblage comportemental de leurs internautes ». Ainsi, Google obtiendrait des données complémentaires à celles

²³¹ <http://www.latribune.fr/technos-medias/medias/20130204trib000746837/google-et-la-presse-les-questions-qui-demeurent.html> (dernière visite le 21 août 2013)

²³² <http://archive-fr.com/page/7059/2012-05-14/http://www.geste.fr/spip.php?article203> (dernière visite le 21 août 2013)

recueillies *via* ses propres services. Et l'éditeur qui rejoindrait le réseau AdSense et donnerait l'autorisation à Google de cibler son lectorat toucherait une rétribution plus grande que s'il ne l'avait pas fait. Une telle démarche, prévenait les porte-parole du Geste, « pourrait permettre immédiatement à Google de proposer au même internaute le même type de liens sur des sites bien moins gourmands, asséchant à court terme les revenus AdSense des premiers ».

Dès lors, il est intéressant de considérer le fait que, lors du partenariat signé avec la presse française, Eric Schmidt a précisé que « Google était également disposé à aider les éditeurs à accroître leurs revenus publicitaires par la mise en place de partenariats qui se traduiraient, au besoin, par la fourniture d'une "aide technique" »²³³. En effet, il semble que Google (qui rappelons-le attire une partie des annonceurs et tire les prix vers le bas sur le marché de la publicité) puisse profiter de la situation pour que les éditeurs se tournent vers ses services d'intermédiaire sur le marché de la publicité. Quand la firme commente l'accord en expliquant que « Google approfondira ses partenariats avec les éditeurs français pour les aider à augmenter leurs revenus en ligne en utilisant ses propres technologies publicitaires »²³⁴, il s'agit d'un discours problématique dans la mesure où la firme a des concurrents sur le marché de la publicité. Marc Schwartz, médiateur des négociations entre Google et les éditeurs français, a d'ailleurs souligné que l'accord comprenait deux volets dont le deuxième consistait en « un partenariat commercial, dont les médias se font peu l'écho, mais qui est très important car il vise à favoriser le développement à long terme de l'audience et des revenus publicitaires de la presse en ligne. C'est un accord par lequel Google va mettre à disposition de la presse ses différentes plates-formes technologiques, AdSense pour la publicité sur PC, AdMob pour la publicité sur mobile, etc... Ce volet est important car, d'une part, il est signé pour une période de trois à cinq ans, et, d'autre part, il est reconductible. »²³⁵ Louis Dreyfus, Président du Directoire du journal *Le Monde*, a quant à lui précisé :

²³³ http://www.lepoint.fr/chroniqueurs-du-point/emmanuel-berretta/presse-google-s-en-sort-par-un-cheque-de-60-millions-d-euros-01-02-2013-1622962_52.php (dernière visite le 21 août 2013)

²³⁴ « Google will deepen our partnership with French publishers to help increase their online revenues using our advertising technology » (Notre traduction) <http://googleblog.blogspot.fr/2013/02/google-creates-60m-digital-publishing.html> (dernière visite le 21 août 2013)

²³⁵ <http://www.challenges.fr/media/20130203.CHA5842/comment-google-et-la-presse-ont-fini-par-s-entendre.html> (dernière visite le 21 août 2013)

« Google a ouvert au passage la possibilité aux éditeurs de profiter de leurs serveurs publicitaires et de leurs outils afin d'optimiser leurs revenus publicitaires. On comprend l'intérêt de Google à faire du business avec les éditeurs de presse, mais nous sommes ici à des années-lumière du point de départ de la négociation. C'est tout le talent de Google ! »²³⁶

Nous voyons comment les différentes marges de manœuvres de Google et des éditeurs sont susceptibles de se télescoper sur le marché de la publicité. Le sujet est d'autant plus délicat que les données récoltées par Google *via* les éditeurs partenaires d'AdSense permettent à la firme d'être plus performante sur le marché publicitaire (cf.5.2.3) et donc d'y concurrencer d'autant plus ces mêmes éditeurs *via* AdWords, dont elle est seule bénéficiaire. Par ailleurs, étant donné l'arbre de navigation décrit dans le chapitre 5, nous comprenons que Google pourrait être incitée à avantager les liens vers les contenus des partenaires de ses réseaux de syndication, c'est-à-dire de ceux qui auront préalablement bénéficié de son aide financière. Le sujet est encore trop brûlant à ce stade pour que nous puissions faire autre chose que de soulever certaines questions fondamentales : Quelles seront les conditions et les modalités du versement ? Quelle sera la nature de la coopération sur le versant publicitaire des activités de Google et des éditeurs ? L'accord aura-t-il vocation à être reconduit ? Le projet de loi est-il définitivement abandonné ?

*

Nous avons vu que les porte-parole Google, lors des différents conflits et négociations décrits, invoquaient systématiquement la possibilité pour les éditeurs d'utiliser le protocole d'exclusion robots.txt pour exclure du champ d'indexation leurs documents. Or ce protocole, au centre du dispositif d'infomédiation, fait lui aussi l'objet de certaines négociations et de certains désaccords sur lesquels nous proposons de revenir dans la dernière sous-partie de ce chapitre. Car de ces négociations et de ces désaccords, ainsi que du champ des possibles sous-tendu par le protocole d'exclusion, dépendent la nature et les limites de l'entente nouée entre Google et les éditeurs et, *in fine*, la nature et le résultat du processus étudié.

²³⁶ http://www.lepoint.fr/chroniqueurs-du-point/emmanuel-berretta/louis-dreyfus-l-accord-avec-google-ne-sauvera-pas-la-presse-02-02-2013-1623139_52.php (dernière visite le 21 août 2013)

6.3 Tacite ou explicite : quel code, pour quel contrat ?

Sur le web, la loi, c'est le code. A partir de ce constat, Lawrence Lessig explique que la question « est de savoir si nous aurons collectivement un rôle dans [le choix du code] - et donc dans la manière dont [nos] valeurs sont garanties - ou si nous laisserons aux codeurs le soin de choisir nos valeurs à notre place. » (Lessig, 1999). Nous avons vu dans le chapitre 4 comment les éditeurs pouvaient communiquer à Google, par code interposé, certaines conditions, spécifications et restrictions quant à la reprise de leurs contenus, mais ont-ils, pour reprendre les mots de Lessig, « un rôle dans le choix de la manière dont leurs valeurs sont garanties » ? Nous avons vu tout au long du chapitre 6 que plusieurs plaintes étaient déposées par des éditeurs trouvant injustes les termes de l'accord qui les lie à Google et, par débordement, insuffisants, ou inappropriés, leurs leviers d'action. Dans presque tous les cas, les porte-parole de Google se réfèrent au même argument : puisque les éditeurs peuvent utiliser le robots.txt et les méta-tags pour extraire leurs site ou certains documents du champ de *crawling*, c'est, dès lors qu'ils ne le font pas, qu'ils sont d'accord pour que leurs contenus soient référencés selon les modalités définies par Google. Il s'agirait donc d'un enrôlement et d'une mobilisation par défaut, sous-tendus par le fait que les éditeurs soient trop intéressés par l'apport de trafic pour utiliser le protocole d'exclusion, ou bien qu'ils n'aient pas les compétences suffisantes pour l'implémenter. Cependant, nous avons vu dans le cas de l'affaire *Copiepresse* que, selon le tribunal belge, le droit d'auteur n'est pas un droit de ne pas être repris par un tiers, mais un droit qui interdit à un tiers de reprendre un contenu sans l'autorisation du propriétaire et de l'auteur. Or il existe justement un langage créé conjointement par les éditeurs de presse en ligne, permettant de donner une autorisation explicite plutôt qu'implicite, nommé *Automated Content Access Protocol* (ACAP). En décrivant ici la syntaxe et le champ d'action de ce langage, ainsi que la controverse dont il a fait l'objet, nous décryptons la nature du liant et des nœuds dont la liste est le catalyseur, et le processus de communication la cause autant que le résultat.

6.3.1 L'initiative ACAP

L'initiative ACAP fut lancée en novembre 2007 par le *European Publishers Council* (EPC), la *World Association of Newspapers* (WAN) et l'*International Publishers Association* (IPA) dans le but que les éditeurs puissent « communiquer leurs politiques d'accès et d'utilisation de leurs contenus dans une forme *machine-readable* aux moteurs de recherche et aux agrégateurs » (ACAP, 2009, p.1). 1250 éditeurs signataires se sont joints à l'initiative, essentiellement des éditeurs de presse en ligne adossés à des médias traditionnels, rassemblés en un même réseau d'influence dans le but d'être entendus par les propriétaires des moteurs. Cela peut sembler peu au regard du nombre d'éditeurs présents sur le web et aux plus de 20 000 sites indexés sur Google Actualités, cependant ACAP compte parmi ses signataires certains éditeurs dont Google et les autres moteurs peuvent avoir intérêt à référencer les contenus pour satisfaire au mieux leurs utilisateurs et bénéficier des effets de réseaux croisés afférents. Par exemple, en France, le Syndicat de la Presse Quotidienne Nationale (SPQN) et le Syndicat de la Presse Quotidienne Régionale (SPQR) sont des membres de la WAN, et donc signataires de ACAP, tandis que le Syndicat National de l'Édition est membre de l'IPA. L'AFP fut quant à elle l'un des pilotes du projet.

Pleinement opérationnel en France depuis juillet 2011²³⁷, ACAP est un protocole non propriétaire destiné à adapter le web aux lois en vigueur en termes de *copyright* et de droit d'auteur. Sur le site dédié à l'initiative (<http://the-acap.org>), le problème auquel ACAP prétend répondre est présenté comme découlant du fait que les conditions d'utilisation d'un site se trouvent en général sur une page dont personne ne prend connaissance et que les moteurs ne peuvent pas lire. Ainsi, ces conditions sont ignorées par les crawlers. ACAP propose par conséquent au propriétaire d'un contenu de *traduire* les conditions dans une forme qui lui permettra à la fois de contrôler la diffusion de ses contenus, mais également de « maximiser les bénéfices de sa relation avec les moteurs et les agrégateurs »²³⁸.

²³⁷ <http://www.gfii.fr/fr/amico/94> (dernière visite le 21 août 2013)

²³⁸ <http://the-acap.org/FAQs.php> (dernière visite le 21 août 2013)

« En général, les éditeurs rendent leur contenu disponible avec l'aide de partenaires et de tiers [...] et sont payés un certain montant ou une part des revenus en retour du service qu'ils fournissent. C'est pourquoi il n'y a aucune raison pour que, dans le monde en ligne, ils ne choisissent pas de faire quelque chose de similaire dès lors qu'ils peuvent négocier commercialement des termes acceptables qui reflètent la contribution à la valeur fournie par l'éditeur et par le tiers. De la même façon, il n'y a aucune raison pour qu'ils ne puissent pas choisir de ne PAS rendre disponible leur contenu de la manière qu'ils veulent. C'est un principe important qui consiste pour l'éditeur à avoir le droit de décider de ce qu'il advient de son matériel et de faire en sorte que sa décision soit respectée. Néanmoins, jusqu'à présent, de nombreux agrégateurs ont choisi d'adopter une interprétation libérale du copyright – « C'est OK tant que personne ne nous dit que ce n'est pas OK » [...] Laissant de côté la question de savoir si une telle attitude est raisonnable, nous avons trouvé une façon pour les éditeurs de faire connaître aux agrégateurs les conditions de reprise de leurs contenus d'une manière qu'ils pourront facilement comprendre. »²³⁹

Les éditeurs signataires d'ACAP effectuent une montée en généralité dans le but de répondre à un obstacle-problème défini par eux comme l'incapacité de traduire en langage informatique les conditions et les modalités de reproduction de leurs contenus. Dès lors, ils tentent de faire de l'ensemble du balisage ACAP un point de passage obligé pour les éditeurs et les moteurs, grâce à la performance duquel le respect du *copyright* sera assuré. Dans la citation ci-après, les éditeurs présentent leurs activités comme garante de l'intérêt commun présent et futur, et le projet ACAP comme le moyen de garantir que ces activités puissent perdurer :

« Les loi du Copyright existe dans le monde entier et donne aux créateurs et aux éditeurs le droit de décider comment le contenu qu'ils ont créé et pour lequel ils ont investi devrait être légalement exploité par les autres. Les industries des médias, qui existent seulement grâce au copyright, contribuent massivement à développer l'économie du monde physique et sont vitales pour le futur économique. La possibilité d'exprimer et de partager des permissions pour l'accès et l'usage [des contenus] de manière standardisée est une part de l'infrastructure nécessaire [à l'activité des industries médiatiques] tant sur le réseau que dans le monde physique. »²⁴⁰

²³⁹ *Ibid.*

²⁴⁰ *Ibid.*

Les défenseurs du projet ACAP tentent également de contourner les lignes de front et les anti-programmes de ceux qui pourraient s'opposer à leur montée en généralité, en prétendant que la défense du copyright et du droit d'auteur ne gênera en rien l'accessibilité des contenus sur le web, au contraire. Ainsi, Dominic Young, directeur des services d'édition de *News International*, explique sur le site d'ACAP :

« Le copyright a conduit non pas à une restriction des contenus, mais à une explosion, un constant et effréné partage des idées ainsi qu'un raz-de-marée de choix pour les consommateurs. Si l'économie du savoir est la clef de voûte de la croissance future, alors le copyright en est la fondation. »²⁴¹

Le protocole d'exclusion robots.txt existait depuis plus de dix ans lorsque l'initiative ACAP fut lancée. Dès lors, il s'agissait pour les signataires de réussir à remplacer le protocole en montrant aux acteurs concernés pourquoi leur proposition était plus appropriée, et tisser pour cela un réseau de partenaires, qui s'enrôleraient soit en implémentant les ressources préconisées, dans le cas des éditeurs, soit en paramétrant les crawlers de façon à prendre en compte ces ressources, dans le cas des propriétaires du moteur.

Dans les faits toutefois, ACAP ne remplace pas le robots.txt. Le nouveau balisage se greffe à l'existant, implémenté directement à l'intérieur du robots.txt, à la racine du site, ou bien dans les balises méta-tags, sur les pages individuelles. En ce sens, il s'agit d'une spécification des protocoles existants visant à exprimer « des permissions plus précises et plus nuancées » (ACAP, 2009, p.2). Sur le site d'ACAP, cette greffe au robots.txt est justifiée étant donnée « l'insistance des moteurs de recherche ». Il s'agit donc d'un détour par rapport au projet initial, dans le but d'intéresser et de mobiliser plus facilement les propriétaires des moteurs dont les crawlers ont été conçus et paramétrés spécifiquement pour la ressource robots.txt. ACAP fut ainsi présenté comme la version 2.0 du robots.txt, ses signataires souhaitant compléter et préciser l'existant plutôt que de le concurrencer. Les éditeurs tentaient ainsi de renégocier le champ de manœuvres décrit dans le chapitre 4, en exerçant à leur tour un « pouvoir de faire faire » vis-à-vis des moteurs en général et de Google en particulier, étant donné les effets de réseaux croisés que leurs contenus génèrent (cf. chapitre 5), et afin que de nouveaux leviers d'action soient mis en place.

²⁴¹ *Ibid.*

6.3.2 Protocole d'inclusion

Nous décrivons ci-après les principales dispositions techniques d'ACAP. Cela nous permettra de connaître ce que les éditeurs signataires du projet souhaiteraient *idéalement* que leur champ de manœuvres HTML ne devînt, i.e. : la manière dont les éditeurs *voudraient pouvoir* performer l'activité des moteurs. Nous saurons ainsi quel est le programme d'action des initiateurs d'ACAP. Par ailleurs, nous comparerons ces propositions à ce que prévoit le robots.txt pour mieux situer et sonder la ligne de front dont le protocole d'exclusion est l'objet.

En premier lieu, les documents ACAP stipulent que le nouveau protocole proposé permet de faire tout ce que le robots.txt permet lui aussi de faire, mais que cela sera signifié désormais (*traduit*) dans les « termes ACAP ». C'est donc essentiellement une différence de syntaxe qui, nous le verrons, dès lors que l'on considère la production de la liste comme le résultat d'un accord contractuel, a toute son importance. Par ailleurs, le guide d'utilisation indique que l'implémentation d'ACAP, en plus de préciser et de nuancer les conditions de reproduction et de diffusion d'un contenu propriétaire, a pour objectif de diminuer le nombre d'opérations nécessaires à l'édition du fichier robots.txt.

A. Annoncer

ACAP permet de mentionner à quel(s) crawler(s) une commande en particulier est destinée. Il est possible de s'adresser à tous les crawlers, à un seul d'entre eux ou bien à un groupe de plusieurs crawlers. Cela était déjà possible avec le robots.txt grâce à la fonction « *User agent* » mais, dans le cas d'ACAP, cela est nécessaire pour dire au crawler dès son arrivée sur le site que le fichier robots.txt contient deux sortes d'informations : l'une destinée aux moteurs qui n'adhèrent pas à ACAP, et l'autre destinée en particulier à ceux qui lisent et respectent les consignes ACAP. Cela permet d'interdire aux crawlers des moteurs qui ne respectent pas ACAP de scanner les contenus tandis que les moteurs qui adhèrent à l'initiative, et donc qui respectent les modalités qu'elle prévoit, y auront, eux, accès. Ainsi, les éditeurs peuvent faire pression sur les propriétaires des moteurs en avantageant ceux qui adhèrent à leurs projets. Dans ce cas, voici ce qu'on trouvera dans le robots.txt :

User-agent: *	(Robots.txt classique)
Disallow: /	
ACAP-ignore-conventional-records	(Syntaxe ACAP)
ACAP-crawler: *	
ACAP-allow-crawl: /	

Les crawleurs sont informés que s'ils ne respectent pas ACAP, ils ne doivent pas crawler le contenu présent sur le site, et que s'il respecte ACAP, ils peuvent ignorer les informations données à ceux qui ne les respectent pas et crawler l'ensemble des documents présents sur le site.

En outre, ACAP permet de s'adresser à un seul crawler ou à un groupe de crawleurs défini dans les pages, via des balises méta-tags spécifiques, ce que ne permettent pas les méta-tags usuelles. Les informations de spécification à propos des crawleurs autorisés sont données une fois pour toutes à la racine du site, pour l'ensemble des documents, via le robots.txt. Avec ACAP, un document en particulier peut être adressé à un crawler ou à un groupe de crawleurs particulier, tandis que le reste du site est adressé indifféremment à tous les crawleurs. La granularité des leviers performatifs est plus fine.

B. Localiser

Une des différences majeures entre ACAP et le robots.txt réside dans le fait que ce dernier permet d'interdire l'accès à l'ensemble du site, à un répertoire de documents ou à un document, tandis qu'ACAP permet d'autoriser explicitement l'accès à un site, un répertoire ou un document. Le robots.txt ne peut utiliser que la balise « Disallow » (interdiction) et ne prévoit pas de balise « Allow » (permission). Ainsi, tandis que le robots.txt est prohibitif et tacite (dès lors que je ne dis rien, c'est que je suis d'accord), le second est permissif et explicite (je dis que je suis d'accord). Ce n'est plus seulement un protocole d'exclusion mais aussi, et surtout, un protocole d'inclusion. Plutôt que d'expulser plusieurs répertoires du champ de *crawling* pour n'en laisser qu'un seul, l'éditeur peut directement préciser au moteur qu'il veut que ce seul répertoire soit indexé. Par exemple, pour le répertoire dont les adresses URL contiennent la mention « /public », cela donnera : ACAP-allow-crawl: /public/

C. Limiter

Il est possible de spécifier l'usage qui ne peut pas être fait du contenu. Le guide d'utilisation donne un exemple où on reconnaît clairement Google : un moteur de recherche généraliste est présenté comme proposant à ses usagers plusieurs services verticaux, notamment des moteurs verticaux d' « images » et de « news ». Ainsi, dans l'exemple, si un éditeur veut être indexé sur le moteur généraliste mais pas sur les moteurs verticaux « images » et « news », il spécifie :

```
ACAP-crawler: *  
ACAP-allow-crawl: /  
ACAP-crawler: SearchBot1  
ACAP-usage-purpose: news.*  
ACAP-usage-purpose: images.*  
ACAP-disallow-crawl:
```

Nous remarquons qu'il s'agit exactement ce que demandait la presse italienne à Google : pouvoir être indexée sur le moteur généraliste Google Search sans être indexée sur le service Google Actualités ; et ce pourquoi elle a porté plainte devant l'autorité italienne de la concurrence (cf. 6.1.7.C).

D. Spécifier

Les commandes principales d'ACAP sont au nombre de quatre :

- « Follow » : Cette commande permet de dire si des liens doivent être suivis ou non. Elle est possible normalement avec les méta-tags spécifiques (cf. 4.1.3), mais n'est pas possible à implémenter dans le robots.txt. Une des particularités d'ACAP est de permettre à l'éditeur de donner cette information au crawler dès la racine du site en ajoutant une commande à la ressource robots.txt. Dans l'exemple ci-dessous, les crawlers ont l'autorisation de visiter le seul contenu du répertoire « /public » (le signe « \$ » annonce une exception) mais ne peuvent pas, dans ce répertoire, suivre les liens du sous-répertoire /comments.

```
ACAP-crawler: *  
ACAP-disallow-crawl: /  
ACAP-allow-crawl: /$  
ACAP-allow-crawl: /public/  
ACAP-disallow-follow: /public/comments/
```

Dans le cas du robots.txt classique, qui ne prévoit pas de commande « allow », cela aurait dû être fait en deux étapes : d'abord communiquer au crawleur qu'il ne peut pas scanner les autres répertoires, mentionnés alors un à un, excepté « /public », puis implémenter une balise méta-tag à la racine du sous-répertoire « /public/comments » (<meta name="robots" content="noindex, nofollow">).

- « Index » : Cette commande permet de spécifier la zone d'indexation, non pas seulement comme le robots.txt et les méta-tags usuels en la diminuant, mais également en la désignant explicitement. Par exemple, dans le cas ci-dessous, où nous reprenons et enrichissons l'exemple donné ci-avant, seules les pages du répertoire « /public » peuvent être crawlées et indexées, tandis qu'aucune autre action que le crawl et l'indexation n'est autorisée : à la dernière ligne, « disallow-other » signifie que toutes les actions qui ne sont pas explicitement permises sont interdites.

```
ACAP-crawler: *  
ACAP-disallow-crawl: /  
ACAP-allow-crawl: /$  
ACAP-allow-crawl: /public/  
ACAP-allow-index: /public/  
ACAP-disallow-other: /public/
```

Il est également possible de spécifier les formats qu'on ne veut pas indexer. Dans ce cas, par exemple pour les images aux formats « .jpg » et « .gif », il suffira d'ajouter :

```
ACAP-disallow-index: /public/*.jpg$  
ACAP-disallow-index: /public/*.gif$
```

Enfin, il est possible de spécifier explicitement le format que le crawleur est autorisé à indexer. Cela devient une permission et un accord explicite plutôt qu'une absence d'interdiction et un accord tacite. Il suffit pour cela de permettre le *crawling* et d'interdire l'indexation des documents tout en permettant explicitement la seule indexation des documents au format, par exemple, HTML, ce qui donne :

```
ACAP-allow-crawl: /public/  
ACAP-disallow-index: /public/  
ACAP-allow-index: /public/*.html
```

- « Preserve » : Cette commande est utilisée pour explicitement permettre au moteur de garder en mémoire le contenu crawlé et indexé. Dans le cas du robots.txt, il n'existe pas une telle commande. En revanche un méta-tag, « noarchive », permet de demander aux moteurs de ne pas garder en mémoire une page en particulier et de ne pas proposer de version « en cache » (cf. 4.1.3). Cette dernière commande est à nouveau une restriction plutôt qu'une permission, et doit être implémentée sur chaque page concernée. A l'inverse, la fonction « Preserve » d'ACAP permet de désigner des répertoires ou l'ensemble du site depuis la racine, ce que ne permet pas la méta-tag « noarchive ». Par ailleurs, si elle n'est pas spécifiée, le moteur doit comprendre qu'il lui est interdit de garder en mémoire sur ses serveurs, même pour une très courte durée, le contenu qu'il aura éventuellement eu le droit de crawler et d'indexer. Si nous reprenons notre exemple, et que nous voulons permettre explicitement de crawler et d'indexer le seul répertoire « /public », et de le mettre en mémoire, excepté le sous-répertoire « /public/comments », cela donnera :

```
ACAP-crawler: *  
ACAP-disallow-crawl: /  
ACAP-allow-crawl: /$  
ACAP-allow-crawl: /public/  
ACAP-allow-index: /public/  
ACAP-allow-preserve: /public/  
ACAP-disallow-preserve: /public/comments
```

- « Present » : Cette commande est utilisée pour permettre au moteur de reprendre et de publier une partie du contenu sous forme de lien hypertexte vers la source originale. On peut s'étonner qu'une telle commande existe, dans la mesure où l'on voit mal pourquoi un éditeur aurait donné la permission de crawler, d'indexer et de garder en mémoire son contenu si c'était pour ensuite interdire au moteur de générer un lien vers lui, mais nous sommes ici dans la continuité de la politique ACAP, selon laquelle chaque action autorisée doit être explicitement mentionnée par l'éditeur, afin qu'il n'y ait jamais aucune ambiguïté quant à l'action du dispositif. En outre, cela permet d'affiner la granularité du contrôle de la publicisation des contenus. Par exemple, cela permet à l'éditeur de dire au moteur qu'il ne veut pas qu'une image soit présentée dans la liste de liens sous forme de « *thumbnail* », alors qu'il accepte qu'un lien et un *snippet* soient générés et

affichés. Le robots.txt et les méta-tags usuels ne permettent pas une telle précision.

Dans notre exemple, cela donne :

```
ACAP-crawler: *  
ACAP-disallow-crawl: /  
ACAP-allow-crawl: /$  
ACAP-allow-crawl: /public/  
ACAP-allow-index: /public/  
ACAP-allow-preserve: /public/  
ACAP-allow-present: /public/  
ACAP-disallow-present-thumbnail: /public/
```

Dans l'exemple ci-dessus, les commandes « crawl », « index » et « preserve » ne sont pas nécessaires dès lors que la commande « present » est mentionnée.

Cependant, le guide ACAP (2009) conseille de les mentionner malgré tout pour que la relation avec le moteur soit la plus explicite possible.

Les autres commandes d'ACAP fonctionnent toutes sur le modèle des quatre commandes décrites et illustrées ci-avant. Il s'agit, à chaque fois, d'affiner la granularité du champ de manœuvres et de rendre l'accord explicite. Il existe des commandes permettant de limiter la date de disponibilité du contenu, de spécifier les attributs des éléments à crawler²⁴², de limiter le nombre de mots du « *snippet* », de prohiber le changement de format parfois effectué par les moteurs (souvent un PDF transformé en HTML), et d'interdire toute modification typographique, toute traduction, toute annotation et toute ré-éditorialisation²⁴³. Une telle granularité permet aux éditeurs de contrôler extrêmement finement les modalités du processus d'infomédiation. Leurs conditions sont ainsi rédigées (*traduites*) dans un langage approprié, et le *crawling* soumis à certaines clauses explicites. Cependant, presque tous les concepteurs et propriétaires de moteurs de recherche, Google en première ligne, ont refusé d'adhérer à ACAP. Ainsi, le projet d'une entente explicite entre Google et les éditeurs n'a jamais vu le jour, et le protocole d'exclusion et les méta-tags restent les seuls moyens pour ces derniers d'établir une communication site/crawler ou document/crawler.

²⁴² Cela est l'objet de plusieurs critiques, car la commande qui permet de spécifier les attributs peut permettre de présenter au moteur un autre contenu que celui que l'internaute verra et peut être utilisée comme une technique appelée « *cloaking* » visant par exemple à publier du contenu pornographique, tout en disant au moteur qu'il s'agit d'une encyclopédie de botanique.

²⁴³ Dans le cas d'une interdiction de ré-éditorialisation, si une version « en cache » est autorisée, le moteur devra publier le texte brut, comme il a été mis en ligne sur ses serveurs, sans aucune remise en forme.

6.3.3 Traduction dans l'oreille d'un sourd

Seul le moteur Exalead, dont les parts de marché sont très faibles, a participé au projet ACAP en qualité de pilote. Google et les autres moteurs, ayant été consultés, ont finalement rejeté en bloc la proposition. Ainsi le pouvoir, en ce qui concerne la fixation de protocoles communs, semble être du côté des propriétaires des moteurs davantage que de celui des éditeurs : à la fois « pouvoir sur » (capacité de rendre caduque la proposition d'un nouveau protocole) et « pouvoir de faire faire » (capacité de faire accepter le protocole de son choix). Par ailleurs, il est intéressant de remarquer que les concurrents de Google, à part Exalead, ont eux aussi refusé d'adhérer à ACAP : c'est-à-dire qu'ils ont préféré être en désaccord avec les éditeurs partenaires de leur dispositif et en accord avec leur concurrent Google. Nous reviendrons ci-après sur le pourquoi d'un tel alignement.

Le déséquilibre dans les rapports de forces entre Google et les éditeurs fut à l'occasion d'ACAP clairement visible : « Les moteurs de recherche sont une sorte de *gatekeepers*, dès lors que ce qu'ils décident de promouvoir devient effectivement une "loi". S'ils ne promeuvent pas une commande d'exclusion en particulier, cela revient à ce que celle-ci n'existe pas. »²⁴⁴ (Sullivan, 2007)

La plupart des observateurs et des experts du domaine se sont montrés très sceptiques vis-à-vis d'ACAP, notamment car, selon eux, le protocole était aussi facile à mettre en place pour les éditeurs qu'il était compliqué à mettre en œuvre pour les moteurs. Les propriétaires de ces derniers auraient en effet dû re-paramétrer les crawlers et les indexeurs dans le but de s'adapter au langage ACAP, ce qui les aurait placés selon eux en face de difficultés techniques indépassables. Ainsi, les éditeurs, soucieux de leurs propres programmes d'action, n'auraient pas assez considéré le programme d'action des concepteurs des moteurs lors de la conception du nouveau protocole, et de ceux de Google en particulier, dans leur tentative de traduction.

²⁴⁴ « The search engines are sort of the gatekeepers, because it's what they decide to support that effectively becomes "law." If they don't support a particular exclusion command, it might as well not exist » (Sullivan, 2007, notre traduction)

« Intégrer ACAP aux sites web est extrêmement simple pour les éditeurs, mais il semble évident que cela place les moteurs de recherche devant de sérieux challenges en matière d'implémentation. Cela est encore pire quand on considère le fait que les restrictions peuvent aussi apparaître à l'intérieur du code HTML des documents. Le Protocole d'Exclusion des Robots et les normes actuelles en matière de méta-tags s'opèrent sur l'ensemble des fichiers, alors qu'ACAP rendrait possible pour les éditeurs d'assigner des permissions et des restrictions sur des sections en particuliers des pages web en mettant des « markup » dans les attributs de chaque élément html. Exiger que les concepteurs des moteurs réussissent à implémenter chaque sorte de metadata prévue par ACAP pour chaque élément individuel d'une page web en particulier fait de l'ensemble un protocole extrêmement difficile à mettre en œuvre. »²⁴⁵

« Le nouveau protocole est entièrement tourné vers les désirs des éditeurs, et en particulier des éditeurs qui ont peur de ce que les internautes feront de leurs contenus s'ils ne gardent pas le contrôle de tous les points de la chaîne de diffusion. »²⁴⁶

Il semble donc que les éditeurs ne soient pas assez revenus sur leur propre programme d'action, ce qui a empêché ACAP d'aboutir, car aucune issue possible n'était assez satisfaisante aux yeux des propriétaires des moteurs. Cela a conduit à ce que nous nommons la « traduction dans l'oreille d'un sourd » : le protocole a bel et bien été créé et mis en service, mais n'a pas reçu l'adhésion des acteurs sans lesquels, pourtant, il n'a pas lieu d'être. En effet, comme le dit le guide d'implémentation lui-même, « pour qu'ACAP fonctionne comme prévu, les opérateurs doivent reprogrammer volontairement leurs crawlers afin d'interpréter les expressions ACAP présentes dans le robots.txt et intégrées au contenu web [i.e. boucler le processus de *traduction*]. La façon dont un crawler interprète ACAP dépendra de la manière dont cette reprogrammation aura eu lieu »²⁴⁷ (ACAP, 2009, p. 4).

²⁴⁵ « Integrating ACAP into existing web sites is trivially easy for publishers, but obviously poses some serious implementation challenges for search engines. Matters are made much worse by the fact that the restrictions can also be described inline in HTML documents. The Robots Exclusion Protocol and the current standard robot header meta tags operates on whole files, but ACAP would make it possible for publishers to assign permissions and restrictions to individual sections of web pages by putting special markup in the class attribute of any html element. Requiring implementors to manage all of the various kinds of ACAP metadata for individual elements of web pages makes the whole thing much harder to implement. » (Notre traduction, disponible ici : <http://arstechnica.com/business/2008/01/skeptical-look-at-acap>, dernière visite le 21 août 2013)

²⁴⁶ « The new protocol focuses entirely on the desires of publishers, and only those publishers who fear what web users will do with the content if they don't retain control over it at every point. » (Notre traduction), disponible ici : http://blogs.telegraph.co.uk/technology/iandouglas/3624601/Acap_a_shot_in_the_foot_for_publishing/ (dernière visite le 21 août 2013)

²⁴⁷ « ACAP is not a technical protection mechanism. For ACAP to function as intended, crawler operators must voluntarily re-program their crawlers to interpret and act upon ACAP forms of expression in robots.txt and embedded in web content.

Dès lors que les moteurs refusent de paramétrer leurs crawlers de façon à prendre connaissance et à respecter les balises du protocole ACAP, la performativité de celui-ci échoue : il ne réussit pas à devenir le discours juridique qu'il aspirait à être, et ne peut pas « faire exister ce qu'il énonce » (Bourdieu, 1982, p. 21). Par ailleurs, les outils ACAP ne fonctionnaient pas ou très mal, selon l'avis de plusieurs éditeurs les ayant expérimentés²⁴⁸. Il ne faut donc pas considérer simplement que Google n'a pas voulu d'ACAP, mais également que le protocole présentait certaines difficultés majeures en termes d'implémentation : l'absence de *performativité* serait ainsi potentiellement liée à une absence de *performance*.

Le porte-parole de Google Rob Jonas, Directeur des partenariats « media et édition » en zone Europe, s'est prononcé en mars 2008 afin d'expliquer pourquoi ACAP ne lui apparaissait pas convenir, et pourquoi le robots.txt suffisait amplement :

« Le point de vue général est que le protocole robots.txt permet à la plupart des éditeurs de faire ce qu'ils doivent faire. Tant que nous ne considérons pas qu'il existe des raisons convaincantes pour l'améliorer, nous pensons que ce protocole permettra à chacun d'agir comme il l'entend »²⁴⁹.

Dans cette citation, nous voyons, d'une part, la montée en généralité que tente d'effectuer le porte-parole : en évoquant le point de vue général et non de celui de Google en particulier, il sous-entend que l'opinion de Google est conforme à celui de la majorité. D'autre part, le porte-parole prétend que le robots.txt fournit les outils que les éditeurs « ont besoin » d'utiliser (*need*), et non pas « peuvent » utiliser. Nous sommes ici dans le registre du « pouvoir de faire » accordé par Google aux éditeurs : le protocole en place est suffisant pour ce que Google pense que les éditeurs doivent pouvoir faire. Or, nous avons vu qu'il ne s'agit pas de ce dont les éditeurs ont besoin en termes purement technique, car la plupart des commandes d'ACAP ont des équivalents avec le robots.txt. En fait, ACAP a pour but de rendre explicite le partenariat entre le propriétaires des contenus et Google, et cela *avant* la publication

How a crawler interprets ACAP will be dependent upon how they are re-programmed to do so. » (ACAP, 2009, p. 4, notre traduction)

²⁴⁸ http://www.currybet.net/cbet_blog/2007/12/acap---flawed-and-broken-from.php (dernière visite le 21 août 2013)

²⁴⁹ « The general view is that the robots.txt protocol provides everything that most publishers need to do. Until we see strong reasons for improving on that, we think it will get every one where they need to be » <http://www.journalism.co.uk/news/google-rejects-adoption-of-acap-standard/s2/a531181/> (dernière visite le 21 août 2013)

des listes hypertextes. Enfin, dans cette citation, nous voyons le « pouvoir sur » de la firme Google qui, tant qu'elle ne verra pas de raisons convaincantes pour changer de protocole, c'est-à-dire tant qu'elle n'aura pas *intérêt* à le faire, fera en sorte de maintenir un protocole en accord avec ses propres intérêts.

Gavin O'Reilly, Président de WAN et fervent défenseur d'ACAP, a répondu à Rob Jonas :

« C'est plutôt étrange de la part de Google de dire aux éditeurs ce qu'ils devraient penser à propos du robots.txt, lorsque les éditeurs du monde entier — de tous les secteurs — ont déjà et clairement dit à Google qu'ils n'étaient absolument pas d'accord. [...] Si la raison pour laquelle Google ne supporte pas (apparemment) ACAP est basée sur ses propres intérêts commerciaux, alors la firme devrait le dire, plutôt que de remettre en cause la confiance qu'on peut nous faire [...] Google devrait réfléchir au fait qu'après douze mois de réflexions intensives et internationales et un développement actif — auquel Google a également pris part — les éditeurs n'ont pas seulement souligné les inadéquations du robots.txt, mais ont construit une solution pratique, ouverte et opérationnelle destinée aux éditeurs et aux agrégateurs. [...] Par conséquent — une fois de plus — nous appelons Google à adhérer à ACAP et à reconnaître le droit qu'ont les propriétaires de contenus de déterminer comment leurs contenus seront utilisés. »²⁵⁰

Nous voyons ici comment Gavin O'Reilly se pose en porte-parole de l'ensemble des fournisseurs de contenus concernés par l'action de Google. Il remet en cause la délégation opérée par les éditeurs vers Google, en arguant que la firme ne peut pas déterminer ce que les éditeurs doivent penser. Il pointe du doigt les opportunités commerciales, sous-entendant que Google défend ses propres intérêts, et non ceux des éditeurs. Il s'agit d'une tentative visant à rompre la chaîne d'associations en remettant en cause la montée en généralité opérée par Google. Enfin, Gavin O'Reilly en vient aux arguments techniques, prétendant que la solution proposée par ACAP fonctionne très bien et que c'est à Google de déléguer aux éditeurs le pouvoir de fixation du protocole d'exclusion (ou d'inclusion), et non l'inverse. Le porte-parole

²⁵⁰ « It's rather strange for Google to be telling publishers what they should think about robots.txt, when publishers worldwide - across all sectors - have already and clearly told Google that they fundamentally disagree. » [...] « If Google's reason for not (apparently) supporting ACAP is built on its own commercial self-interest, then it should say so, and not glibly throw mistruths about, [...] Google should reflect on the fact that after 12 months of intensive cross industry consideration and active development - in which Google has been party to - publishers have identified not only the patent inadequacies of robots.txt, but more progressively have come up with a practical, open and workable solution for publishers and content aggregators. [...] So, we - once again - call upon Google to embrace ACAP and to readily acknowledge the right of content owners to determine how their content is use » <http://www.journalism.co.uk/news/gavin-o-reilly-responds-to-google-s-rejection-of-acap/s2/a531184/> (dernière visite le 21 août 2013)

des éditeurs essaye ainsi de justifier un « pouvoir sur » que les éditeurs sont selon lui en droit d'exercer, et non Google. Malgré cela, Google n'a pas adhéré à ACAP.

Désireux de calmer la dispute, le PDG de Google Eric Schmidt a précisé alors qu'il ne s'agissait que d'un problème technique, et non, pour Google, d'une volonté de « voler le contrôle de l'information » aux éditeurs²⁵¹ :

« ACAP est un standard proposé par un groupe de personnes qui essayent de résoudre le problème. Certains de nos employés travaillent avec eux pour voir si la proposition peut être modifiée de manière à pouvoir fonctionner sans remettre en cause le fonctionnement du moteur de recherche. Or pour l'instant, [ACAP] est incompatible avec la façon dont notre système opère. »²⁵²

Schmidt tente d'apaiser la situation en mettant en avant un problème d'incompatibilité à propos duquel il ne donne pas de plus amples détails techniques. Il s'agit donc de croire sur parole le Président de Google à ce sujet.

Le projet ACAP, même s'il a largement échoué, a pu, et peut toujours, étant donné précisément la controverse dont il a fait l'objet, orienter les détours qui soutiennent la chaîne d'associations. En effet, ACAP a permis à Google de savoir précisément ce que les éditeurs demandaient. Ainsi, si un nouveau protocole est mis en place un jour, il y a fort à parier que l'initiative ACAP et les débats qu'elle a provoqués seront considérés dans la formulation du nouveau projet. C'est en tout cas ce que suggère Dan Sullivan en ces termes : « Personnellement, je doute qu'ACAP devienne un jour le robots.txt 2.0 — mais je pense que certains éléments d'ACAP seront intégrés à [la nouvelle version du protocole] »²⁵³ (Sullivan, 2007).

Aujourd'hui, ACAP est au point mort. En 2011, à Berlin, le projet a été repris²⁵⁴ par l'*International Press and Telecommunications Council* (IPTC), consortium d'agences de presse et d'éditeurs de presse en ligne. Son objectif est désormais d'influencer les projets de loi et les décisions prises par la Commission

²⁵¹ <http://www.itwire.com/your-it-news/home-it/17206-acap-content-protection-protocol-doesnt-work-says-google-ceo> (Interview Eric Schmidt, dernière visite le 21 août 2013)

²⁵² « ACAP is a standard proposed by a set of people who are trying to solve the problem [of communicating content access permissions]. We have some people working with them to see if the proposal can be modified to work in the way our search engines work. At present it does not fit with the way our systems operate. » Notre traduction, disponible ici :

²⁵³ « Personally, I doubt ACAP will become Robots.txt 2.0 — but I suspect elements of ACAP will flow into that new version or a successor. » (Sullivan, 2007, notre traduction)

²⁵⁴ http://www.iptc.org/site/Home/Media_Releases/News_syndication_version_of_ACAP_ready_for_launch_and_management_handed_over_to_the_IPTC (dernière visite le 21 août 2013)

Européenne en matière de protection du droit d'auteur. D'une réflexion et d'une proposition essentiellement techniques, le projet a ainsi glissé vers la sphère politique. Même si ACAP n'est plus à l'ordre du jour, il a été intégré à un consortium qui, lui, est encore actif. Une version 2.0 d'ACAP a tout de même été définie en juin 2011 à partir du standard *Open Digital Rights Language*, à laquelle, sans surprise, les principaux moteurs n'ont pas adhéré.

Lorsque, dans le cadre de notre enquête, nous avons questionné à propos d'ACAP celui qui a remplacé Rob Jonas au poste de Directeur des partenariats « media et édition » en zone Europe, Madhav Chinappa nous a répondu que d'après lui le robots.txt fonctionnait très bien, et qu'il n'y avait pas de raison de le remplacer. Ces propos sont identiques à ceux de Jonas : le point de vue de Google sur cette question semble donc être demeuré le même.

6.3.4 Consensus léonin

Le Code Civil définit le contrat comme une « une convention par laquelle une ou plusieurs personnes s'obligent, envers une ou plusieurs autres, à donner, à faire ou à ne pas faire quelque chose », et précise : « le contrat est synallagmatique ou bilatéral lorsque les contractants s'obligent réciproquement les uns envers les autres » (*Code civil*, 1804, art. 1101-1102, p. 200). Dans le cas de Google et de la liste d'hyperliens générée par le dispositif, il semble qu'il n'y ait pas d'obligation de Google vis-à-vis des éditeurs, dans la mesure où la firme peut décider, et cela à tout moment, de ne pas référencer tel ou tel éditeur, tel ou tel contenu. Les éditeurs se retrouvent dans une situation où ils sont d'une certaine façon les obligés de Google : dès lors qu'ils publient des contenus en lignes, ceux-ci peuvent être crawlés et indexés. Certes, le protocole d'exclusion robots.txt permet à l'éditeur de demander à Google de ne pas procéder au référencement, cependant ce protocole n'a pas force de loi et Google pourrait décider, sans être hors la loi, de paramétrer ses crawlers de façon à l'ignorer. En d'autres termes, le respect du protocole dépend du bon-vouloir des concepteurs du dispositif. Ainsi, produire et publier un contenu sur le web revient à accepter qu'il puisse être crawlé, indexé et référencé, à condition qu'une telle activité n'enfreigne pas les règles du *copyright* ou du droit d'auteur. Or, nous l'avons

vu, le *copyright* et le droit d'auteur ne sont pas nécessairement contraignant pour Google, qui peut se prévaloir du *fair-use* dans le cas du premier, ou des différentes exceptions dont le second fait l'objet. Dès lors, la relation de la firme et des éditeurs peut être apparentée à un contrat unilatéral, que le Code civil définit comme une situation dans laquelle « une ou plusieurs personnes sont obligées envers une ou plusieurs autres, sans que de la part de ces dernières il y ait engagement » (*Code civil*, 1804, art. 1101-1102, p. 200). L'obligation se trouve alors du côté d'une seule des parties. Dans notre cas, elle se trouve du côté des éditeurs, qui doivent accepter les conditions proposées par la firme, laquelle n'est en rien obligée d'accepter les conditions que les éditeurs lui soumettent pourtant. C'est par conséquent d'un accord²⁵⁵ unilatéral dont il s'agit, dont la seule limite est fixée par une jurisprudence floue en termes de *copyright*. Cependant, cette limite elle-même peut parfois être dépassée, et ne semble pas empêcher la firme de soumettre les éditeurs aux conditions qu'elle leur fixe. Nous avons vu avec le cas de la presse belge comment Google pouvait faire annuler les effets d'une décision de justice confirmée en Cour d'Appel. Dans les cas mentionnés dans ce chapitre (Belgique, Allemagne, Italie, France), la firme a toujours prôné le même argument : il est possible aux éditeurs qui le souhaitent d'exclure leurs documents du champ de *crawling*, grâce au protocole robots.txt dont Google a décidé de respecter les modalités. Dès lors que les éditeurs n'implémentent pas le protocole, la firme estime qu'ils sont d'accord pour que leurs documents soient référencés selon des modalités implicites et définies par elle seule. En effet, nous venons de voir comment Google a refusé aux éditeurs la possibilité de définir eux-mêmes les modalités d'inclusion ou d'exclusion. Ainsi, la tentative ACAP, qui, comme le robots.txt, n'a pas force de loi, a été enrayée par la non-adhésion de la firme au projet des éditeurs. Et l'entente entre les parties demeure par conséquent unilatérale, prohibitive et implicite.

Pour conclure, nous pensons que le dispositif et son fonctionnement sont l'objet d'un consensus « léonin ». Cet adjectif, appliqué à un accord, désigne la différence de poids des parties dans les négociations dont l'entente est le résultat. Il est clair, en effet, que les dirigeants et les employés de Google sont ceux qui

²⁵⁵ Nous disons « accord » plutôt que « contrat » puisqu'il n'y a pas ici d'acte juridique à proprement parler.

choisissent les conditions et les modalités du référencement sur leurs services, et ce même lorsque les éditeurs se rassemblent pour proposer d'autres modalités. En reprenant les termes de l'ANT, nous dirions que les contractants peinent à jouer un rôle de *contre-actants*. Finalement, la plupart des éditeurs n'utilisent pas le robots.txt et n'implémentent pas non plus ACAP, puisque les moteurs ne le respectent pas. Aussi acceptent-ils, bon gré mal gré, par consensus, les dispositions et le résultat d'un accord unilatéral et léonin.

CONCLUSION PARTIE 2 : LE DISPOSITIF MIS A PLAT

Comme nous l'avons vu tout au long du chapitre 3, les éditeurs peuvent spécifier, renforcer, diminuer ou rompre l'association de leur cours d'action à l'action de Google. Il existe ainsi différents degrés d'enrôlement et de mobilisation, dont dépendront *in fine* la solidité de l'association et l'ampleur de la traduction. En guise de conclusion pour cette partie, nous déployons le dispositif, c'est-à-dire que nous ordonnons les principaux éléments descriptifs tirés de nos trois derniers chapitres en une même « référence circulante » (Latour, 1999a) composée de tableaux et de schémas permettant, en quelques rapides coups d'œil, de visualiser l'ensemble de l'écosystème étudié. C'est sur ce déploiement que sera basée la partie suivante, où nous nous attacherons à « ré-assembler » le dispositif. En effet, maintenant que nous connaissons ce que les éditeurs peuvent faire, tentent de faire faire et ont intérêt à faire, nous allons voir ce qu'ils font effectivement, comment ils le justifient et ce que cela implique.

Champ de manœuvres

Dans le tableau ci-dessous, nous résumons les principaux « pouvoir faire » et « pouvoir de faire » du champ de manœuvres des éditeurs, dès lors qu'il s'agit d'établir, de spécifier et de renforcer l'association de son action à l'action de Google.

Tableau 5. Champ de manœuvres des éditeurs (Etablir, Spécifier, Renforcer)

		Etablir	Spécifier	Renforcer
Search + Actualités	Pouvoir faire	- Ne pas utiliser le Robots.txt - Sitemap « Search » à minima : <urlset> <url> <loc>	- Robots.txt - User-agent: * - Disallow: (spécifier) - Allow: (spécifier) - Méta-tag <Meta Http-Equiv="Content-Type" Content="..."; charset=..."> - <Meta Http-Equiv="refresh" Content="...;url=..."> - Sitemap « Search » balises facultatives : <lastmod> <changefreq> <priority> - Exclure certains documents volumineux du champ d'indexation à l'aide du robots.txt - Utiliser les microdatas, les microformats et les RDFa.	- Privilégier les formats HTML - Augmenter la vitesse de chargement de la page (CDN) - Ecrire des documents explicites - Mettre en place une politique de tagging. - Créer des pages d'atterrissage sur des mots-clé pertinents - Positionner les liens hypertextes sur des ancres pertinentes - PageRank sculpting - Être présent sur Google+ - Qualité éditoriale - Microdonnées
	Pouvoir de faire		- AuthorRank via la balise rel=author et les comptes Google+ des journalistes - Choisir les rubriques qui apparaîtront dans le « Rich Snippet »	- Augmenter la fréquence du crawling (déconseillé) - Dénonciation des éditeurs qui ne respectent pas les recommandations - Utiliser « Tendances de recherche » - Signature numérique « rel=author »

Seulement Google Actualités	Pouvoir faire	- 3 chiffres dans l'URL - Mettre des images plus grandes que 60x60 pixels	- Mots clés Actualités <meta name="news_keywords" content="balise meta, mot-clé1, mot-clé2, mot-clé3, etc., news_keywords">	- Privilégier le format .jpeg pour les images, et un « format raisonnable » - Ne pas utiliser, ou peu, de cadres.
	Pouvoir de faire	- Créer une chaîne Youtube - Indiquer clairement les forums et espaces de commentaires (/board/ ou /forum/)	- Premier clic gratuit - Syndication source - Standout	- SiteMap Actualité - Intégrer « Le choix des rédactions »

Nous avons également vu comment un éditeur pouvait réduire, interrompre momentanément ou rompre l'association :

Tableau 6. Champ de manœuvres des éditeurs (Réduire, Interrompre, Rompre)

		Réduire	Interrompre	Rompre
Search + Actualités	Pouvoir faire	- Privilégier les formats vidéo, image, Flash, JavaScript - Exclure certains documents du champ d'indexation à l'aide du robots.txt - Méta-tag <Meta name="googlebot" Content="...", ..."> <Meta Name="google" Value="notranslate"> - Jeux de mots, sous-entendus - Faire des liens sur des ancrs non-pertinentes - Empêcher Google de proposer un « cache » <Meta name="googlebot" Content="noarchive">	- Méta-tag <meta name="robots" content="noindex, follow">	- Robots.txt (rupture générale) User-agent: * Disallow: / - Méta-tag (rupture localisée) <meta name="robots" content="noindex, nofollow"> ou <meta name="robots" content="none">
	Pouvoir de faire	- Diminuer la fréquence du crawling		
Seulement Actualités	Pouvoir faire	- Faire payer ses contenus, renvoyer vers une page de paiement ou d'abonnement		
	Pouvoir de faire			

Si nous nous situons à présent du côté de Google, voici le « pouvoir de faire faire » que nous avons identifié (c'est-à-dire les éléments que Google tente d'influencer, mais sur lesquels l'éditeur garde la main) :

Tableau 7. Champ de manœuvres de Google (Pouvoir de faire faire)

Pouvoir de faire faire		
Search + Actualités	Technique	- HTML suggestions : influence sur les caractéristiques techniques de documents - AuthorRank : influence sur la mise en œuvre d'une signature numérique, reliant les contenus produits aux comptes Google+ de leurs auteurs' - Schema.org : influencer sur le choix par les éditeurs du format « microdata » plutôt que les microformats et les RDFa.
	Editorial	- Tendance de recherche : influence sur le choix des mots employés et des sujets traités - Rich Snippet : Influence sur le choix du nom des rubriques et sur leur hiérarchie
Actualités	Technique	- Influencer pour le choix du HTML - Influencer la taille des images (minimum 60x60 pixels) - Influencer sur le choix du format jpeg - N'implémenter qu'un nombre de balises « raisonnable »
	Editorial	- Ne pas publier, ou peu, de didacticiels, éditoriaux, offres d'emploi, contenus exclusivement informatifs (prévisions météo, informations boursières, etc.) - Ecrire ni trop long ni trop court - Disposer d'un processus formel de révision éditoriale - Titre : 22 mots maxi, ne pas utiliser de préfixe dans le texte d'ancrage - Privilégier l'expertise - Privilégier les articles signés, reliés à une biographie de chaque auteur, et ainsi une responsabilité en nom propre de ceux qui traitent l'information - Privilégier le bon usage de la langue
	Economique	- Proposer gratuitement les contenus, ou, au moins, le « Premier Clic Gratuit » avec un minimum de cinq articles par jour - Inviter à extraire les contenus payants du champ de Google Actualités, sans quoi l'ensemble des pages du site risqueraient de voir leur référencement pénalisé - Ne pas revendre à des annonceurs la possibilité de transmettre leur message sur Google Actualités sous forme publi-rédactionnelle

Nous pouvons également identifier le « pouvoir sur » de Google : un pouvoir au sens plus classique du terme, disciplinaire, fonctionnant sur le principe de la sanction, de la récompense et de la limitation :

Tableau 8. Champ de manœuvres de Google (Pouvoir sur)

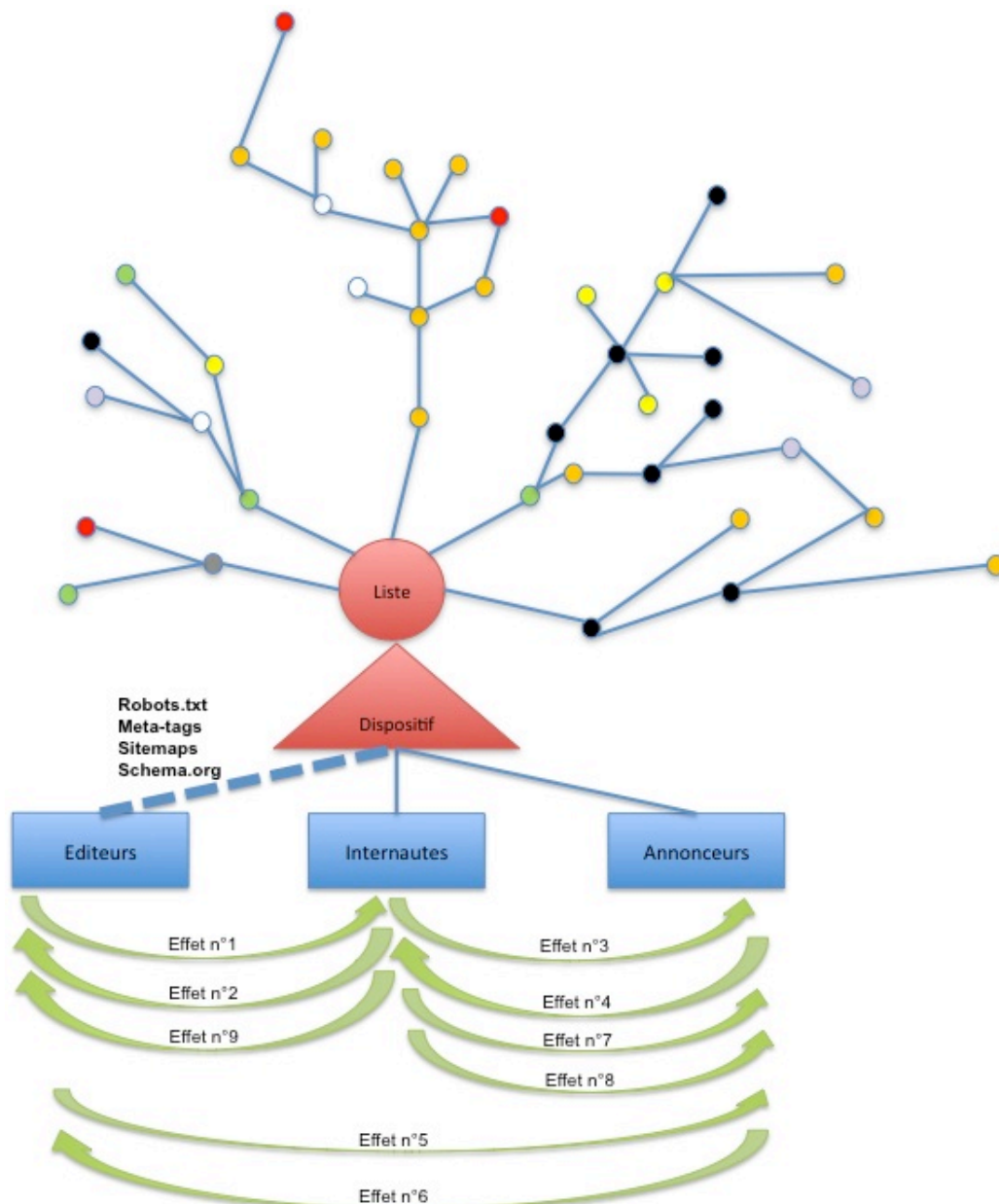
Pouvoir Sur	Limitation	Sanction	Récompense
Search + Actualités	- Minimum de cinq pages par jour pour le Premier clic Gratuit - Limitation de l'usage du Standout à sept articles par semaine maximum	Désindexer un site	- Afficher les Rich Snippets - Considérer le « rel=author »
Seulement Actualités	- 10 mots clefs maximum pour le Méta-tag « News Keyword » - Article de 10 caractères au moins et tire comprenant en tire 2 et 22 mots. - Limitation du nombre d'articles du Choix des rédactions à cinq - Intégration obligatoire d'au moins un nouvel article toutes les 48 heures dans le choix des rédactions - Mettre au minimum trois articles dans Le choix des rédactions	Ne pas indexer ou désindexer un site	- Insérer dans l'index - Insérer dans « Le choix des rédactions » - Manipuler les résultats en faveur de l'éditeur²⁵⁶

²⁵⁶ Ce n'est pas parce qu'ils ont ce pouvoir qu'ils le font. Mais rien ne le leur interdit, et personne ne sait s'ils le font ou non. Ils ont donc ce pouvoir.

Lieu d'échange et d'opportunités

Si nous reprenons à présent les Figures 26, 27 et 28 du chapitre 5 (cf. 5.2.4 et 5.2.5), nous obtenons une représentation du dispositif en termes d'effets de réseaux et d'opportunités de profit, du point de vue de Google. Cela nous permet, d'un seul coup d'œil, de situer les éditeurs, ainsi que les forces et les intérêts qui enserrant la production du méta-énoncé et sont susceptibles de performer à la fois le fonctionnement du dispositif et la qualité de ce qui y est produit et échangé.

Figure 30. Le dispositif d'infomédiation schématisé



Terrain de négociations et d'entente

Dans le chapitre 6, nous avons sondé les débats dont fait l'objet le lien entre les éditeurs et Google, et donc les modalités d'enrôlement/mobilisation des éditeurs dans le dispositif (représentées par le lien en pointillés sur notre schéma précédent). Nous avons vu comment l'entente entre les parties avait lieu, sous quelles conditions, ainsi que les discussions dont ces conditions et ces modalités faisaient l'objet. Nous avons montré, grâce à l'exemple de la tentative ACAP, comment, et à quel point, Google avait pour l'instant le pouvoir de déterminer la nature du procédé par lequel les éditeurs s'enrôlent et les termes et les conditions de l'accord qui les lie à la firme et aux actions des autres acteurs enrôlés dans le dispositif.

Maintenant que nous avons tracé en quelque sorte le squelette du dispositif (Fig. 30), nous allons y placer, si l'on peut dire, *la matière vivante*. Nous allons rendre compte pour cela dans la Partie 3 de la manière dont les éditeurs se déplacent étant donné ce champ de manœuvres, ces intérêts, ces opportunités et ces conditions d'entente. Comment actualisent-ils, très concrètement, la situation décrite ici ? Et que devient, dans le cas où un éditeur cherche à maximiser le trafic provenant de Google, l'organisation de presse (chapitre 7), la mise en récit de l'actualité (chapitre 8) et les pratiques journalistiques (chapitre 9) ? A partir de ces questions, nous chercherons à « ré-assembler » le dispositif autour de ce champ de manœuvres, de ces liens, de ces effets de réseaux, de cet arbre de navigation, de ces opportunités de profits et de ce consensus léonin. Autrement dit, maintenant que nous avons schématisé le *squelette* du processus de communication étudié, nous tâcherons de comprendre comment la *chair* s'agence « autour de », « par » et « dans » ce squelette.

Partie 3 – Reconfiguration médiatique : effets et limites du dispositif

« Loin de se trouver, face aux événements et aux discours qui animent le corps social, dans une position extérieure et autonome, le journaliste se situe en réalité à un carrefour de contraintes avec lesquelles il est plus ou moins forcé de composer. »

(Derville, 1999, p. 171)

« SEO is applying pressure to these standards not in the perceived interest of the reader or because of publication constraints (as once may have been the case), but in the interests of a third-party commercial arbiter in online distribution: Google. A new conformity in the language and “aboutness” of news online is emerging, mediated by Google. Google stands between those terms used by the reading public and journalist alike, implicitly regulating the “market place of ideas” as it goes. Relevance as a quantitative measure is challenging what people feel qualitative “relevance” to be. »
(Dick, 2011, p. 475)

Dans cette partie, nous analysons le processus de production journalistique à partir de la mise à plat du dispositif effectuée dans la partie précédente, des entretiens menés auprès de dirigeants et d’employés œuvrant pour des sites issus de titre de presse imprimée (SIPI) et de nos observations en ligne. Dans le chapitre 7, nous nous intéresserons à l’organisation de presse : comment et pourquoi développe-t-elle une stratégie SEO ? à quels nouveaux métiers cela donne lieu et comment les employés coopèrent-ils autour de ces questions ? comment les différents acteurs négocient-ils selon leurs intérêts propres, leurs objectifs, leurs savoir-faire ? Dans le chapitre 8, nous nous intéressons aux effets des stratégies SEO sur la structure des sites et des pages, ainsi que sur l’utilisation des balises *méta-tags* et des liens hypertextes et sur les outils et les partenaires associés à la valorisation des contenus. Nous verrons ainsi comment, concrètement, le résultat du processus de production journalistique est *travaillé* par la question du référencement. Enfin, dans le chapitre 9, nous nous intéressons à la manière dont les pratiques et les valeurs des journalistes sont

négociées étant donné les stratégies de référencement. Que leur demande-t-on de faire ? Dans quelle mesure sont-ils libres d'arbitrer entre leurs propres visées et celles du référenceur ? A quel point ces visées sont-elles conciliables et conciliées ? Quels sont les points qui posent problème ? Quel traitement de l'actualité résulte des mises en tensions dévoilées ?

Notre terrain est composé de deux sous-ensembles (cf. 1.7). Le premier se situe au niveau mezzo et s'appuie sur les 16 entretiens semi-directifs menés entre septembre 2011 et mars 2012 auprès de dirigeants d'entreprises de presse, de journalistes, d'employés en charge du référencement, de consultants SEO et d'un formateur aux pratiques d'écriture journalistique. En plus de ces entretiens, nous avons effectué en octobre 2012 une observation de 252 articles mis en ligne sur 6 sites d'information politique et générale adossés à des titres de presse imprimée, et nous avons assidument observés ces sites à la fois « devant » et « derrière », c'est-à-dire que nous avons observé ce que l'internaute *voit* grâce à son navigateur et au moteur de recherche et, d'autre part, ce que le navigateur et le moteur *lisent* dans le code HTML et *traduisent* (i.e. interprètent et déplacent) pour l'internaute. Le deuxième ensemble de notre terrain est monographique. Nous avons observé le site *Lexpress.fr* durant l'année 2012 et rencontré, entre septembre 2011 et mars 2012, 19 employés travaillant à son élaboration, son fonctionnement et sa monétisation. En nous appuyant sur ces entretiens, nous décrirons certaines spécificités de Roularta Media France, allant ainsi de l'échantillon à variation à une étude monographique, et d'observations générales à l'analyse descriptive d'un cas en particulier. Les chapitres 7 et 8 s'achèveront ainsi par un zoom nous permettant d'interroger et de préciser les tendances observées, de mieux décrire les mutations induites par la stratégie de référencement et d'en dévoiler certaines concernant RMF en particulier. Quant au dernier chapitre (chapitre 9), il sera exclusivement consacré au groupe RMF et au site web *Lexpress.fr*, pour des raisons que nous justifierons dans son introduction.

CHAPITRE 7 – L'ORGANE DE PRESSE COMME ESPACE

STRATEGIQUE

« Le changement en général et le changement dans les organisations en particulier sont des objets difficiles à appréhender. Qu'appelle-t-on "changement" en effet ? L'évolution des structures, celles des identités professionnelles, la constitution des règles, les décisions des directions, etc. ? Dira-t-on que le changement vient des contraintes extérieures à l'entreprise, de l'environnement, de la concurrence, des technologies, ou au contraire n'y a-t-il de changement que lorsque les acteurs concernés ont transformé leur système de représentations et de relations, le sens qu'ils donnent à leurs actions ? Le changement peut-il être imposé ou bien ne faut-il appeler "changement" que celui qui a entraîné des modifications faites par les acteurs eux-mêmes ? »

(Bernoux, 2010, p.7)

Pour être un véritable « organe » de presse, c'est-à-dire une « institution formée d'un ensemble d'éléments coordonnés entre eux et remplissant des fonctions déterminées » (CNRTL), l'entreprise propriétaire d'un site d'information politique et générale doit réussir à ce que les individus et les objets, les valeurs, les inscriptions, les représentations qui la composent se réunissent autour d'intérêts communs et d'un cours d'action relativement unifié. Il s'agit d'établir certaines règles et certaines logiques de fonctionnement, explicites ou non, permettant à la production journalistique d'avoir lieu et d'être viable. Nous nous intéresserons ici à ceux dont les scripts *interfèrent* et *composent* le cours d'action de ce que nous avons considéré dans la Partie 2 comme un réseau stabilisé et nommé « éditeur ». Dresser l'ensemble de la chaîne d'attachements permettant à l'éditeur d'agir déborderait largement le cadre de ce travail, aussi resterons-nous attachés à la seule question du référencement : Quels obstacles-problèmes menacent le réseau tout en le conduisant à se reconfigurer ? Quels sont les acteurs concernés ? Comment leurs actions sont-elles coordonnées ? Quelle est l'ampleur des traductions opérées ? Pour répondre à ces questions, nous analyserons la manière dont les acteurs justifient la décision de mettre en place une stratégie de référencement, tout en montrant à quelle situation économique celle-ci prétend répondre. Dans une seconde sous-partie, nous nous intéresserons au référenceur, c'est-à-dire à celui qui a été embauché dans le but de maximiser l'apport

de trafic provenant des moteurs de recherche. Nous montrerons qu'il est un « centre de traduction » : interface à la fois entre Google et l'éditeur et, chez l'éditeur, entre les dirigeants, les équipes techniques, les journalistes et les équipes commerciales. Après quoi nous expliquerons en quoi la rationalisation du SEO peut dans certains cas s'apparenter à une contestation du script de Google, dont le but est d'exercer une forme de *pouvoir* sur l'infomédiation des contenus. Nous aurons ainsi abordé les mutations qui ne concernent pas directement le site ou la politique éditoriale, mais l'organisation de leur production, tandis que nous consacrerons les chapitres 8 et 9 aux mutations de ce qui est effectivement mis en récit et en circulation. Dans la dernière sous-partie du chapitre 7, nous reprendrons notre monographie où nous l'avons laissée à la fin du chapitre 2, afin de voir comment le travail est structuré chez Roularta Media France en particulier, et quels y sont les tenants et les aboutissants organisationnels de la stratégie de référencement.

7.1 Rationalisation du référencement

Nous avons questionné les dirigeants et les employés d'entreprises propriétaires de sites issus de la presse imprimée (SIPI) ayant mis en place des stratégies de référencement afin de comprendre comment ils se représentaient leur relation avec Google, et, d'autre part, de quelle manière ils justifiaient la mise en œuvre de changements visant à optimiser la visibilité de leurs documents. Dans cette sous-partie, nous développons le sujet de la dépendance des éditeurs à l'action de Google. Nous expliquerons en particulier comment la rationalisation du référencement est destinée à répondre à un besoin urgent de financement et à atténuer les symptômes d'une situation encore instable.

7.1.1 Dépendance

Les acteurs interrogés nous ont dit être conscients d'une forme de dépendance indépassable unissant le site pour lequel ils travaillaient à Google. Etant donné le choix d'un modèle axé sur le financement par la publicité facturée au coût par affichage, le trafic apporté par Google Search et Google Actualité est nécessaire à la

viabilité des SIPI. Dès lors, les dirigeants, journalistes, techniciens, n'imaginent pas y renoncer.

« Nous ne pouvons pas renoncer à la source Google, car nous nous couperions d'un trafic considérable. On se tirerait une balle dans le pied. Il faut absolument ce trafic. C'est vital pour notre modèle. »

Lefigaro.fr — Rédacteur en chef — Luc de Barochez

« Les sites médias dépendent de Google. Un site média, celui du Parisien mais aussi ceux des principaux concurrents, c'est en moyenne 30 à 50% de trafic qui arrive de Google. »

Leparisien.fr — Référenceur — Guillaume Giraudet

« La Google dépendance, à mon avis, elle sera toujours présente, parce qu'on fonctionne tous, si on refait le lien avec la pub, sur ce mécanisme : c'est la masse, la puissance, qui fait la crédibilité et qui génère du chiffre. »

Lexpress.fr — Directeur Acquisition d'audience — Marc Bougué

Le Président directeur général de 20minutes.fr va jusqu'à souhaiter que Google « pille » ses contenus autant que ce sera possible, dans la mesure où cela augmentera le trafic vers ses pages et la profitabilité du site :

« A partir du moment où nous sommes sur un modèle de gratuité, tout ce qui nous procure de l'audience va dans le bon sens. Google, diraient certains, pille nos contenus, mais moi je dis tant mieux ! L'important, c'est que ça nous revienne d'une manière ou d'une autre. Nous sommes dans un modèle de gratuité, et dans une culture internet, donc nous ne cherchons pas à garder tel ou tel contenu pour le monétiser. On ne monétise pas les contenus nous, mais l'audience... C'est ça notre métier. Et Google, c'est un tiers de notre audience ! »

20minutes.fr — Président directeur général — Pierre-Jean Bozo

Nous retrouvons ici un modèle économique similaire à celui du titre papier gratuit *20 minutes*, distribué au plus grand nombre de personnes possible dans le but de maximiser l'exposition des encarts publicitaires et, par-là même, les revenus provenant des annonceurs. Ainsi, comme l'avait déjà remarqué Yannick Estienne, les stratégies des sites d'information et des titres de presse quotidienne gratuits sont très proches, basées sur « de nouvelles conceptions de l'information étroitement liées à la façon dont le marketing pense le produit éditorial » (Estienne, 2007, p. 88).

Les quotidiens régionaux n'imaginent pas non plus se priver du trafic envoyé par Google vers leurs pages web. Notamment, les outils Google Search et Google

Actualités leur permettent d'atteindre un lectorat parfois éloigné de leur zone géographique et qui, parce que ce lectorat ne connaît pas ou très peu le titre, n'aurait pu arriver par aucun autre moyen qu'un moteur de recherche ou un agrégateur.

« Nous sommes dans une stratégie de maximisation d'audience, donc nous n'allons pas nous priver du trafic que peut, ou pourrait, nous apporter Google. De plus, nous atteignons grâce à Google des lecteurs à Strasbourg ou en Bretagne, qui ne nous connaissent pas, et qui ne seraient pas venus sinon. C'est donc un outil qui nous permet d'étendre notre lectorat potentiel et d'augmenter considérablement nos recettes publicitaires. »

La Dépêche du Midi — Journaliste, Resp. desk numérique — Philippe Rioux

La présence sur Google permet aux sites adossés aux quotidiens régionaux de dépasser la seule dimension régionale de leur activité. Sur le web, *La Dépêche du Midi* devient l'expert national (et mondial) des sujets d'actualité de sa région. Le web rompt ainsi avec le papier en débordant son périmètre. Cela peut être rapproché de la stratégie des sites adossés à des titres de presse magazine hebdomadaire, pour qui la présence sur Google constitue une façon de rompre avec la temporalité de publication du newsmagazine en s'instituant comme des sites traitant l'information en continu.

« L'inscription dans le quotidien, pour nous, elle n'est pas évidente. Sur un site comme lemonde.fr, adossé à un quotidien, les gens y vont naturellement tous les jours. Alors que L'Express, même si ça paraît évident que c'est en continu, ça ne l'est pas encore pour tout le monde, ce qui fait que le SEO c'est hyper important. Être présent tous les jours sur les requêtes des internautes, ou bien sur Google News, ça nous permet de nous positionner comme acteur de l'immédiateté au même titre que les autres. Et petit à petit, les gens comprennent qu'on est sur de l'info en continu, comme lemonde.fr. »

L'express.fr — Directeur Acquisition d'audience — Marc Bougué

Les chiffres exacts des structures d'audience ne nous ont pas été communiqués, cependant nous avons réussi à obtenir oralement certaines informations à leur sujet, nous les décrivant comme comportant au moins 20% depuis Google Search et au moins 10% depuis Google Actualités. Ainsi, chacun des acteurs interrogés nous a dit que le trafic provenant de Google représentait au moins un tiers de son trafic global. De tels chiffres étaient présentés comme la preuve que l'association du cours d'action de l'éditeur au cours d'action de Google était *obligatoire*.

Un des arguments également mis en avant concernait la durée de vie des articles du papier. Le site web permet en effet à un article d'être lu plusieurs jours/mois après sa publication, et Google joue un rôle important dans cet allongement de la durée de vie d'un article. Au *Figaro*, au *Point*, à *L'Express* et à la *Dépêche du Midi*, il nous a été expliqué que plus on s'éloignait de la date de publication d'un article, et plus la part de trafic provenant de Google vers cet article augmentait. Ainsi, le dispositif permet aux articles vieillissant de continuer à être lus et à générer des bénéfices. Cela nous a également été confirmé par le Délégué général du Syndicat de la Presse Magazine.

La dépendance suscite certaines critiques, aussi bien chez les dirigeants que chez les journalistes, craignant en général que leur survie ne dépende de leur référencement alors que les dirigeants et les employés de Google peuvent théoriquement choisir demain d'exclure un éditeur du champ de *crawling* ou bien de paramétrer l'algorithme de manière à nuire, volontairement ou non, à la visibilité des documents. C'est pourquoi, en même temps qu'ils nous expliquaient ne pas pouvoir se passer de la part de trafic apportée par Google, les acteurs nous disaient qu'ils cherchaient à ce qu'elle ne soit pas non plus trop importante, nous présentant une telle dilution comme une « force » et un gage de « liberté » :

« Nous ne sommes pas dépendants à 100%. Moins de 40% de nos lecteurs viennent de Google, ce qui est notre grande force. Ça nous rend libres. Si un jour nous avons un problème avec un algorithme, so what ? ça impactera seulement un tiers de l'audience et ça nous laissera le temps de nous retourner. Ça peut être une faiblesse, mais en même temps c'est une force. Le problème qu'il faut comprendre, c'est que Google peut désindexer vos contenus ou bien changer son algorithme. Ça n'arrivera sûrement pas, mais il faut quand même comprendre que ça peut arriver et agir en fonction. »

20minutes.fr — Président directeur général — Pierre-Jean Bozo

L'enjeu est présenté à chaque fois comme étant de réussir, grâce à la stratégie de référencement, à attirer le plus grand nombre possible de lecteurs sur le site via Google Search et Google Actualités, mais ensuite à garder ce trafic, lui faire visiter plusieurs pages, le fidéliser et le qualifier. Il s'agit de transformer l'utilisateur des services de Google en un lecteur assidu et fidèle, qui n'aura plus besoin d'utiliser de

moteur ou d'agrégateur pour s'informer. Le but est ainsi de « prendre » des internautes à Google plutôt que de collaborer durablement avec la firme.

Les journalistes, quant eux, sont assez fatalistes. Dès lors qu'on aborde la question de la dépendance de leur site vis-à-vis de Google et de la rationalisation du référencement, ils ne voient pas comment il serait possible, en tout cas à court terme, de faire autrement.

« On est très dépendants des moteurs de recherche, peut-être trop, sûrement trop. Parce que si demain il n'y a plus Google, je ne vois pas d'où pourra venir l'audience nécessaire à notre fonctionnement. Donc c'est un allié, ça c'est sûr, mais c'est peut être parfois un peu trop un allié, et parfois aussi un peu trop le seul allié possible. »

Lepoint.fr — Journaliste (Economie) — Marc Vignaud

« J'ai deux patrons : mon rédacteur en chef et Google. [...] Travailler sur le web, c'est en quelque sorte aussi, à mon avis en tout cas, travailler pour Google. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Certains référenceurs, dont pourtant cette dépendance constitue la raison d'être de leur embauche dans l'organisation, voient également d'un mauvais œil une trop grande sujétion à l'action de Google :

« Je trouve ça hyper inquiétant la dépendance à Google, même si c'est ça qui fait mon métier. Parce que ce n'est pas Google qui s'adapte à toi mais nous qui nous adaptons à Google, donc c'est très particulier. En fait, on est à la merci de Google. Heureusement, les grandes marques sont favorisées. Je ne pense pas que Google nous mette des pénalités ou nous défavorise volontairement, quand on est une marque comme Le Point. Mais bon, on a toujours une inquiétude. [...] Google est une menace. Quelque part, les internautes n'arrivent pas directement sur ton site. Google sert d'intermédiaire. C'est comme des gens qui achètent en grande surface leurs fruits et légumes plutôt que d'aller directement chez le producteur. C'est cracher dans la soupe, de dire ça, et dans la soupe qui me fait vivre en plus, mais pour moi c'est clair : Google est une menace. »

Lepoint.fr — Référenceur — Laure Aubry

La firme Google est ainsi considérée comme un allié obligé, mais aussi comme une menace potentielle. L'association semble avoir lieu à défaut d'une meilleure solution, dans le but à la fois d'augmenter le trafic et, dans certains cas, de rompre avec l'image que les sites issus de la presse régionale et de la presse hebdomadaire

héritent du titre papier, pour finalement se présenter comme des sites d'information en continu adressés à tous les internautes indifféremment. Par ailleurs, les éditeurs cherchent à diminuer leur dépendance et présentent leur objectif comme étant, non pas de collaborer sur le long terme, mais idéalement de fidéliser le trafic provenant des sites Google pour que les internautes en question n'utilisent plus ces services.

7.1.2 Situation d'urgence

En plus de répondre au besoin des internautes, Google répond donc aux difficultés des éditeurs à établir un modèle viable (Attias, 2007) en constituant pour eux un réservoir de trafic dont il est possible, théoriquement et moyennant un certain coût, d'optimiser l'apport. Lorsque nous avons interrogé les employés des sites nés en ligne *Mediapart*, *Owini*, *Rue 89*, dont les modèles économiques ne sont pas adossés à l'affichage publicitaire, le discours au sujet de la dépendance vis-à-vis de Google n'était pas le même : le référencement n'était pas perçu comme contraignant ni même, dans le cas particulier de *Mediapart* et de son modèle payant, comme nécessaire. A l'inverse, les entreprises choisissant de mettre en place une stratégie de SEO le font parce qu'elles sont « de plus en plus soumises à des logiques commerciales de maximisation des audiences » (Ouakrat, 2011, p. 18).

« Les agences de pub sécurisent : elles vont à ceux qui ont la meilleure audience. Donc si tu n'es pas dans le ranking des premiers ou des quatre premiers, t'es mort : personne ne mettra de publicité chez toi, enfin beaucoup moins d'agences en tout cas. L'enjeu, il est là avec Google. Il faut atteindre plus d'audience que les autres. [...] [Les éditeurs] se rendent compte que faire que du ranking et de l'optimisation ça ne leur donnera peut-être pas de la valeur ajoutée, mais ça leur permettra d'être dans le top. C'est pour ça qu'ils le font. C'est même juste pour ça. »

Syndicat de la Presse Magazine — Délégué général — Xavier Dordor

Le nombre de visiteurs semble ainsi compter davantage que la qualification de l'audience ou la qualité du pacte de lecture, tandis que le besoin de maximiser en valeur absolue l'audience explique l'importance pour les SIPI d'être aussi bien référencés que possible. Cependant, une telle situation peut être imputée au manque de maturité et d'équipement du marché publicitaire quant au ciblage des campagnes. En effet, si les éditeurs parvenaient à mieux valoriser la qualification de leur audience

auprès des annonceurs, plutôt que le seul nombre de visiteurs et de pages vues, ils seraient moins dépendants du lectorat non qualifié provenant de Google. Mais pour l'instant, étant donné une facturation au coût par affichage (CPM), le mode de calcul des tarifs publicitaires est corrélé au volume du trafic, ce qui explique le choix de rationaliser le SEO.

Par ailleurs, malgré cette situation d'urgence et de dépendance, les dirigeants des SIPI n'en sont pas moins conscients de la dépendance de Google vis-à-vis d'eux et en particulier des effets de réseaux croisés que leur présence génère (chapitre 5). Dès lors, ils ne sont pas prêts à accepter sans condition le référencement de leurs articles et la situation est perçue comme une solution de court terme répondant à la situation d'urgence provoquée par un marché non mature, qui n'a pas vocation à perdurer, du moins pas dans ces conditions. C'est pourquoi les négociations liées aux modalités de reprise des contenus, décrites dans le chapitre 6, continuent :

« Pour moi, Google est un réseau de distribution physique. Le problème de fond, c'est que Google est monopolistique et que, comme dans tous secteurs, vous avez un certain nombre d'acteurs qui font la chaîne de valeur de ce secteur. Il faut travailler en bonne intelligence les uns avec les autres pour que chacun ait sa part, puisse vivre correctement et alimenter la chaîne de valeur. Sur le digital, il y a Google et Apple qui brassent des milliards. Mais sans les éditeurs, Google et Apple ne peuvent pas vivre. Sans des contenus à valeur ajoutée, ils ne peuvent pas vivre. Sauf que l'éditeur aujourd'hui ne s'y retrouve pas dans la chaîne de valeur. Et si l'éditeur meurt, parce que c'est ça qui va se passer s'il ne se passe pas d'autre chose, Google et Apple finiront par mourir eux aussi. Donc la situation ne peut pas rester comme elle est. Il faut dialoguer. Il faut négocier. Pour l'instant, l'écosystème n'est pas fixé. »

Lexpress.fr — Editrice — Sophie Gohier

7.1.3 Choisir d'internaliser

Dans le contexte d'urgence et de dépendance décrit ci-avant, les principaux SIPI français ont décidé rationaliser la mise en visibilité de leurs documents sur les moteurs en embauchant un référenceur²⁵⁷. Une telle évolution correspond à ce que Pablo Boczkowski nomme la « phase de rationalité » (Boczkowski, 2010, p. 48),

²⁵⁷ Le nom utilisé pour désigner l'employé en charge d'optimiser la visibilité des pages sur Google peut varier selon les entreprises. Nous le désignons, par souci de clarté, par un vocable unique : « référenceur ».

c'est-à-dire qu'elle correspond au moment où il ne s'agit plus pour un organe de presse de montrer qu'il est présent sur le web mais également d'agir de manière à ce que sa présence soit profitable. Selon Virginie Clève, experte du référencement et employée par *Radio France* en qualité de Responsable du Pôle Marketing Digital, les compétences nécessaires à la mise en place d'une stratégie SEO pour un site d'information font l'objet en France d'une pénurie. L'embauche d'un référenceur représente par conséquent un coût que seules certaines entreprises ont les moyens de supporter — mais étant donné le caractère quotidiens des tâches à mener, il est absolument nécessaire d'internaliser (Clève, 2010, 2011)²⁵⁸.

Les principaux SIPI français ont opté pour une stratégie d'internalisation qui n'était pas toujours justifiée *a priori* du point de vue économique. Au *Point*, par exemple, le référenceur engagé a eu l'impression que son embauche était liée à une intuition d'avantage qu'à un calcul rationnel, et que ses chefs n'avaient réalisé le bien-fondé de son embauche que lorsque le trafic s'était effectivement mis à augmenter.

Lorsqu'un référenceur arrive dans une entreprise de presse, il a besoin d'une période d'acclimatation et d'observation pour élaborer un plan d'action. Une fois les premières mesures techniques lancées, il doit former les journalistes, procéder à une veille constante quant aux mises à jour effectuées par Google et aux outils d'optimisation, et, enfin, effectuer des tests dans le but d'ajuster, selon le *feed-back* obtenu, sa stratégie. C'est pourquoi l'internalisation semble nécessaire, y compris du point de vue des agences :

« Nous pouvons mettre en place une solution lors de la création du site, ou bien, si le site existe déjà, faire un audit et donner des reco[mandations]. Mais ensuite, si on s'en va, c'est pas certain que ça va servir à quelque

²⁵⁸ « Trouver un référenceur capable de propulser durablement un site en home de Google Actualités est [...] quasi mission impossible. Cette compétence est à mon avis presque impossible à obtenir lorsque l'on est consultant en agence du fait des barrières prestataire/client et lorsque l'on sait que 80% environ des référenceurs travaillent en agence, on peut déduire qu'il ne s'agit que d'une dizaine ou d'une vingtaine de personnes à Paris et peut être une cinquantaine en France au maximum. [...] Cela explique que beaucoup de sites médias / presse parmi les plus connus peinent encore à émerger, voire n'émergent pas du tout : il ne suffit pas d'avoir un excellent référenceur à sa disposition, il faut avoir un spécialiste de Google Actualités. Vous voulez du trafic gratuit via Google Actualités ? Préparez-vous à sortir votre carnet de chèque ! Le référenceur qui vous apportera les connaissances nécessaires connaît la rareté de sa compétence et ne se vendra pas pour rien. Et le temps qu'il faudra passer en spécifications, développement et formation des rédacteurs de contenu sera lui aussi loin d'être bon marché... Alors avant de vous lancer dans un projet d'optimisation pour Google Actualités, calculez bien le ROI à en attendre. Dans certains cas cela sera très rentable, dans d'autres, pas du tout. » (Clève, 2010)

chose. A partir de là, deux solutions : ou bien ils nous gardent comme consultants réguliers, ou bien ils embauchent quelqu'un qui se chargera de suivre les recos et d'en faire de nouvelles. »

Resoneo — Consultant — Jean-Christophe Baudrier

Le PDG de *20 minutes* Pierre-Jean Bozo justifie également l'internalisation en mettant en avant la nécessité de pouvoir réagir très rapidement en cas de déclassement dans les résultats :

« On avait d'abord un prestataire suédois, qui était le même pour tous les sites du groupe²⁵⁹, puis on a embauché à l'interne quelqu'un ici, à Paris, en 2008. Quand on a des problèmes avec Google, c'est impératif d'avoir quelqu'un à l'interne qui soit capable tout de suite de passer un petit coup de fil pour leur demander ce qui ne va pas. Ça ne se soustraite pas à un certain niveau, ce genre d'opération, question de réactivité. Vous imaginez un consultant suédois qui est à Stockholm et qui appelle les USA puis qui fait un retour en France ? Attendez, faisons simple... »

20minutes.fr — Président directeur général — Pierre-Jean Bozo

Il est intéressant de constater ici qu'il est fait mention d'une relation hors-ligne entre l'éditeur de presse et Google. Ainsi, il semble que la marge de manœuvres décrite dans le chapitre 4 soit agrémentée par un champ d'action hors-ligne qui consiste à contacter directement, et rapidement, les employés de Google en cas de déclassement. Nous verrons que c'est également le cas pour *L'Express* (cf. 7.4.4).

7.2 Le référenceur comme centre de traduction

Afin de définir une stratégie de maximisation du trafic, le référenceur capitalise ce qu'il sait du fonctionnement du moteur, utilise les outils de description et d'analyse à sa disposition, le langage HTML, les conseils donnés par Google, le contenu du site pour lequel il travaille, les objectifs de ses dirigeants, les vellétés de ses collègues, les analyses publiées par des experts, les compte-rendu de mesure d'audience, le classement de l'OJD, l'observation des sites concurrents, et, bien sûr, les listes de résultats et ce qu'elles lui apprennent à propos de la situation à optimiser. Ainsi, le référenceur doit, pour arriver à ses fins, se saisir des lignes d'actions

²⁵⁹ Le propriétaire de *20 minutes* est le norvégien Schibsted, qui a choisi de consulter un prestataire suédois pour tous les sites détenus par le groupe.

décrites dans le chapitre 4, dans l'environnement économique décrit dans le chapitre 5 et en acceptant, au moins partiellement, les conditions et les modalités de contractualisation décrites dans le chapitre 6. A chaque mise en œuvre, ce « centre de traduction » (Callon, 2006) fusionne avec le réseau de lignes d'action qui permet, mais aussi qui autorise et peut conditionner, son action. Avant de décrire et d'analyser les négociations et les ajustements nécessaires à une telle fusion (chapitres 8 et 9), nous dressons ici le portrait du centre de traduction lui-même : Qui est le référenceur ? D'où vient-il ? Quelles sont ses valeurs ? Quelle est sa place dans l'organisation de presse ? Les réponses à ces questions nous permettront de comprendre à la fois sa capacité à mobiliser le réseau, les limites de cette capacité et, enfin, les obstacles-problèmes qu'il cherche à résoudre ou à contourner.

7.2.1 Profils hétérogènes et fonction hybride

Les référenceurs que nous avons rencontrés avaient tous moins de trente-cinq ans. Parmi eux, il y avait une seule femme (*Lepoint.fr*) et le plus jeune avait 23 ans lors de l'entretien (*Leparisien.fr*). Ils ne se sont jamais définis comme des informaticiens. Plusieurs d'entre eux n'ont d'ailleurs pas fait d'études d'informatique. Le référenceur du site *lefigaro.fr* est diplômé d'une école de commerce et ses compétences en SEO sont le fruit d'un intérêt et d'une expérience personnels. Le référenceur du site *Lepoint.fr* a une maîtrise de lettres modernes et une formation en webmarketing. D'autres référenceurs ont fait des études spécialisées. C'est le cas du référenceur de *Lexpress.fr*, initié au SEO lors de ses études à l'*Ecole Technologie Informatique Communication*. Selon les référenceurs eux-mêmes, l'hétérogénéité de leurs parcours est due à la nouveauté de leur métier, ainsi qu'à la triple compétence dont celui-ci fait l'objet :

« Parfois c'est technique, et d'autres fois, quand il s'agit de choisir sur quels mots-clés se positionner, on se rapproche du côté marketing, littéraire presque, sémantique. Parfois enfin, quand il s'agit d'écrire d'une certaine façon, la fameuse écriture pour Google, c'est de l'éditorial pur. C'est pour ça qu'aucune formation n'est parfaite pour faire du SEO et, qu'en même temps, toutes y conduisent. »

Resoneo – Consultant — Jean-Christophe Baudrier

Les différents parcours peuvent aboutir à des visions différentes du métier, et à des tensions, dès lors qu'il s'agirait de privilégier l'un ou l'autre des trois aspects mentionnés. Par exemple, le référenceur du site *Lepoint.fr* voit dans son métier une pratique liée au travail des journalistes davantage qu'à la publicité et à la monétisation :

« J'ai un rôle qui est un peu liée à la publicité, c'est vrai, mais je ne veux pas non plus travailler pour la publicité. Ça ne m'intéresse pas. Moi j'ai fait des études de lettres, et donc je me sens plus proche des journalistes. C'est avec eux, et pour eux, que je travaille. »

Lepoint.fr – Référenceur — Laure Aubry

Certains référenceurs, par exemple au *Point* et au *Figaro*, sont également chargés de la mesure d'audience et, en particulier, de mettre en œuvre les outils de mesure et la « Déclaration Systématique de Fréquentation » de l'OJD. Cette double compétence leur permet d'être aux premières loges pour mesurer les effets du SEO. En revanche, ces référenceurs nous ont dit que la mesure d'audience était malgré tout une tâche distincte du SEO et nécessiterait, pour être parfaitement menée à bien, un employé à temps plein. Ainsi, le temps consacré à la mesure d'audience par le référenceur du site *Lefigaro.fr* l'empêche selon lui d'avoir assez de temps pour mettre en place une stratégie de référencement optimale.

7.2.2 Rémunération

Le salaire des référenceurs interrogés varie entre 37 000 euros et 52 000 euros brut par an. À expérience équivalente, les référenceurs sont donc en général mieux payés que les journalistes, d'après les informations obtenues auprès des dirigeants qui ont accepté d'aborder ce point avec nous, jugé « délicat ». Un tel salaire est justifié par la rareté des compétences requises (Clève, 2010). Les référenceurs, quant à eux, nous ont dit que ces salaires étaient en-deçà de ce qui leur aurait été offert dans le secteur du e-commerce. En conséquence, deux d'entre eux ont précisé qu'il s'agissait d'un choix de leur part lié à d'autres critères que le seul montant de la rémunération : être moins payé mais travailler pour un site de presse, car la nature du contenu et le rôle social lié au traitement de l'actualité les intéressaient davantage.

« Je pense que c'est parce que le SEO explose, on se fait chasser de partout, il y a beaucoup moins de gens qui font du SEO que de personnes qui cherchent, donc c'est une compétence rare, à l'inverse des journalistes qui sont nombreux à chercher du travail. [...] Dans le e-commerce, je serais sûrement mieux payée, donc si je suis là, c'est d'abord pour ce qu'on fait, je veux dire ce qu'on fait ensemble, avec les journalistes, qui m'intéresse plus que de vendre des voyages. »

Lepoint.fr – Référenceur — Laure Aubry

Les référenceurs n'ont pas de part variable dans leur salaire, ce qui est en général le cas pour leurs homologues des sites de e-commerce dont cette part est mesurée à partir d'objectifs fixés au préalable en termes d'apport de trafic. En plus d'être due à deux cultures d'entreprise très différentes, cette absence est liée au fait que les référenceurs des sites de presse n'ont pas d'objectifs en termes de trafic. Aucun ne semble pourtant redouter l'idée qu'on puisse lui fixer des objectifs, mais tous nous ont dit apprécier l'absence de pression. D'autre part, certains référenceurs complètent leurs revenus par le salaire qu'ils tirent d'une activité annexe. Quatre de ceux que nous avons interrogés avaient en effet déposé un statut d'auto-entrepreneur et travaillaient à leur compte pour des clients dont les sites n'étaient pas concurrents des sites des éditeurs qui les avaient embauchés.

7.2.3 Place dans l'organisation

Les référenceurs jouissent d'une certaine liberté d'action, d'une absence de pression et d'une autonomie dans l'entreprise. Ils ont des compétences techniques que personne d'autre qu'eux ne possède et en tirent une forme d'indépendance, mais aussi une forme de *pouvoir*. Une telle expertise et la maîtrise d'une « zone d'incertitude » sont à l'origine de ce que les sociologues des organisations Michel Crozier et Erhard Friedberg identifient comme une des quatre principales sources de pouvoir dans l'entreprise, « découlant de la maîtrise d'une compétence particulière et de la spécialisation fonctionnelle » (Crozier et Friedberg, 1977, p. 83).

« Quand je dis un truc, puisque personne ne comprend comment marche le SEO, les gens ne me posent pas trop de questions et font ce que je leur recommande — à moins que les journalistes ne s'y opposent formellement. »

Leparisien.fr - Référenceur — Guillaume Giraudet

« C'est très flou pour tout le monde, et pour moi aussi parfois, mais pour moi c'est quand même moins flou. Et c'est grâce à ce flou que, dans la boîte, j'ai mon indépendance et ma marge d'action. »

Lepoint.fr – Réferenceur — Laure Aubry

Le référenceur parle « un autre langage » souvent très peu compris, ou même pas du tout, par ses collègues et ses chefs, qui le laissent par conséquent agir comme il le souhaite du moment que cela ne remet pas en cause leurs propres cours d'actions (ce qui, nous le verrons, peut arriver). Pour autant, cette autonomie ne signifie pas que le référenceur est un électron libre dans l'entreprise. Il est attaché en général à une division chargée de l'acquisition et de la mesure d'audience, elle-même rattachée à la division marketing et située ailleurs (une autre pièce, un autre étage) que la rédaction, ce que plusieurs d'entre eux nous ont dit regretter :

« Ce serait plus logique d'être au même étage que celui où les journalistes travaillent. Parce que moi, en tout cas, je pourrais être plus utile à leur niveau, alors que la pub, je n'ai rien à faire avec eux. »

Lepoint.fr — Réferenceur — Laure Aubry

Le référenceur n'a pas d'ascendant hiérarchique sur les équipes techniques ou la rédaction. Il est par conséquent en perpétuelle phase de négociations avec les autres employés et la direction dans le but de faire accepter ses recommandations. Il tente de convaincre ses partenaires du bien-fondé de ce qu'il propose, en tâchant de les *intéresser* à l'apport de trafic potentiel que peut apporter Google au site et de leur faire comprendre qu'il est pour cela nécessaire de mettre en place une stratégie globale. Sans l'*enrôlement* (j'accepte la définition qu'il donne de mon rôle dans l'entreprise) et la *mobilisation* (je fais en sorte d'agir en fonction de ce qu'il attend d'un tel rôle), au moins à minima, le référenceur échouera. Le référenceur doit donc mener à bien un travail de persuasion, ce qui peut être rendu compliqué par le fait que peu de monde dans l'entreprise, ou même personne, ne comprend exactement ce qu'il fait. C'est pourquoi le référenceur est un « centre de traduction » : s'il veut être compris, il doit *traduire* son projet dans la langue des autres, et rendre compatibles les actions qu'il propose avec les buts poursuivis par ses partenaires, et avec leurs intérêts, leurs marges de manœuvres, leurs représentations, sans quoi son action échouera. Il ne peut arriver à ses fins tout seul.

7.3 Le SEO comme négociation

Le référenceur n'est pas simplement un centre de traduction à l'intérieur de l'entreprise, mais également une interface entre l'organisation de presse et l'infomédiaire. Il cherche à *traduire* dans une langue comprise (et appréciée) par les crawlers, l'indexeur et l'algorithme le travail des journalistes et des équipes techniques et commerciales, de manière à ce que les cours d'action des ingénieurs de Google et des éditeurs de contenus puissent se rencontrer et que la liste puisse être générée de façon à satisfaire les employés de Google, qui ne désindexeront pas le site, ainsi que ses collègues et ses supérieurs, qui verront leurs intérêts satisfaits et continueront d'adhérer à la stratégie qu'il aura définie. Le travail du référenceur consiste ainsi à ajuster la médiation technique de manière à ce qu'elle satisfasse l'ensemble des actants en présence, sans quoi son rôle pourra être remis en cause, soit parce que Google aura jugé ses pratiques abusives et sanctionné le site qu'il représente, soit parce que ses supérieurs et/ou ses collègues auront jugé son action inutile ou nuisible à leurs intérêts et décidé de ne plus y collaborer.

Ainsi, l'action du référenceur, en plus d'être une négociation directe dans l'entreprise, peut être assimilée à une forme de négociation avec les employés de Google qui, elle, n'est pas directe, mais *médiée* par la technique. Quand le référenceur se saisit de son champ de manœuvres, il ne le fait toutefois pas toujours conformément aux préconisations de Google, ni forcément dans l'intérêt de Google. Il le fait également pour satisfaire les intérêts de l'entreprise qu'il représente et les siens propres. L'enjeu pour lui est de réussir à positionner ses contenus en haut des résultats sans que Google ne décide de le punir. Dès lors, il flirte avec les limites, actionne différents leviers, joue avec les règles et cherche, en faisant preuve de créativité, à repousser les bornes de son champ d'action, sans jamais en arriver au point où Google décidera d'exclure son site.

Dans cette sous-partie, nous procéderons en premier lieu à un tour d'horizon afin de montrer comment la limite entre ce qui est acceptable ou non en matière de SEO est négociée entre Google et les référenceurs. Nous reviendrons ensuite à notre terrain d'étude et à ce que nous pouvons dire des référenceurs œuvrant pour les SIPI français. Enfin, nous montrerons que ce sont finalement les caractéristiques de

l'ensemble du dispositif qui sont négociées, dès lors que Google s'adapte aux stratégies de SEO autant que les référenceurs adaptent leurs stratégies à l'action de Google.

7.3.1 Chapeaux noirs et chapeaux blancs

Il est d'usage de différencier parmi les référenceurs les « *white hats* » (chapeaux blancs) et les « *black hats* » (chapeaux noirs), c'est-à-dire ceux qui auraient un comportement allant dans le sens de ce que préconise Google, et ceux qui n'agiraient que dans le sens de leur seul intérêt en tentant de manipuler les résultats. Cependant, la limite entre *black* et *white* est floue. Le *black hat* change notamment de définition selon qu'on se place du côté de Google ou de celui des référenceurs : « Pour la société Google et les référenceurs, ce terme recouvre l'ensemble des procédés qui biaisent les algorithmes de moteurs de recherche. Pour les autres, il rejoint celui de nuisance, de spamming, de messages et commentaires non sollicités sur leurs supports, Matt Cutts²⁶⁰ parle de ravage ou brûlure. Pratiquement, on remarque que le terme englobe un large panel de pratiques qui pour la plupart visent au profit économique. Ainsi, sur les forums dédiés au *black hat SEO*, se mêlent les personnes souhaitant favoriser l'indexation de leur site internet par tous moyens, mais aussi des individus cherchant le bénéfice par les manières les plus illégales qui soient, tels que le piratage de sites internet, de cartes de crédit, aussi, il n'est pas étonnant que cette activité ait mauvaise presse » (Boutet et Amor, 2010, p. 195).

Quand on se rend sur des sites dédiés au *black hat*, ce n'est pas toujours l'enjeu économique qui est mis en avant, mais celui de lutte, de contradiction et de refus du pouvoir exercé par Google.

« Le "Black Hat" pour moi ça veut dire faire tout ce qui est en mon pouvoir pour outrepasser les règles. Les règles sont mises en place pour s'assurer que les gens soient subordonnés à des lois et des régulations instituées par des corporations géantes et des systèmes qui prennent la forme d'un gouvernement. »²⁶¹

²⁶⁰ Matt Cutts est ingénieur, en charge chez Google de la lutte anti-spam, et porte-parole régulier de la firme à propos de la question des bonnes et des mauvaises pratiques en matière de référencement.

²⁶¹ <http://www.blackhatworld.com/blackhat-seo/blackhat-lounge/333179-we-future-why-we-all-here.html> (dernière visite le 22 août 2013)

« Pour ce qui concerne les règles des moteurs de recherche, j'ai comme projet de les imprimer sur du papier toilette. En d'autres termes, il faut appliquer la règle du "pas vu, pas pris" et rien de plus. Google et aucun autre n'a le droit de nous dicter comment faire pour fourguer des pages (plus ou moins pertinentes je l'accorde) dans l'index. »²⁶²

« Le black hat est le dernier pragmatiste. Il ne se sent pas concerné par l'éthique du spamming. Il voit Google sous son jour véritable — une corporation globale qui cherche à maximiser ses profits et les dividendes de ses actionnaires. Google Inc. n'est pas le gouvernement — leurs préconisations sont guidées par des motivations liées au profit et non par l'éthique ou la règle de droit. »²⁶³

Les référenceurs *black hat* innovent pour contourner ce qui semble être imposé. Ils essayent ainsi de faire faire au dispositif ce qui va dans le sens de leurs propres intérêts. En cela, ils peuvent être rapprochés des *hackers* : « ardent (et excentrique) programmeurs capables de manipuler les machines de manière brillante et non-orthodoxe » (Nissenbaum, 2004, p. 196). Ils œuvrent en général dans l'anonymat, et comme les premiers hackers, « ont en général peu de respect pour les règles contraignantes que les administrateurs aiment imposer, et qu'ils cherchent à contourner » (Stallman, 2001).

S'il n'est pas aisé de définir le SEO *black hat* du point de vue de ceux qui le pratiquent, c'est, en premier lieu, parce que leur objectif est précisément d'être pris pour des référenceurs qui respectent les préconisations de Google. Ainsi, un référenceur *black hat*, c'est d'abord un référenceur qui se fait passer pour un *white hat*. Il ne respecte pas *pratiquement* les préconisations de Google tout en les respectant *apparemment*²⁶⁴.

Une autre des ambiguïtés majeures quant à la définition des pratiques *black hat* vient du fait que ces pratiques ne désignent pas toujours ce qui est défendu par les porte-parole de Google, mais parfois seulement ce que Google n'avait pas anticipé. Les référenceurs font preuve de créativité à partir de ce que les employés de la firme ont prévu, décidé et conçu. « Les personnes qui s'aventurent hors des sentiers battus

²⁶² <http://www.visibilite-referencement.fr/blog/laurent-bourrelly-black-hat-seo> (dernière visite le 22 août 2013)

²⁶³ <http://searchengineland.com/the-secret-life-of-the-black-hat-seo-59472> (dernière visite le 22 août 2013)

²⁶⁴ « Je pense qu'il faut maîtriser les bases du référencement white hat déjà, sinon on va droit dans le mur. Parce que dans tous les cas le Black Hat devra faire croire à Google qu'il est le plus white hat des white hats. », source : <http://www.plateforme-redacteur.fr/levolution-du-black-hat-seo-au-travers-de-lautomatisation> (dernière visite le 22 août 2013)

sont souvent taxées de *black hat* SEO sans pour autant forcément nuire » (Boutet et Amor, 2010, p. 180). C'est pourquoi nombre d'entre eux ont le sentiment de ne pas avoir mérité leur mauvaise réputation²⁶⁵.

Une sorte de frontière éthique permettrait de délimiter ce qui est *black hat* et ce qui est *white hat*, mais c'est sur la définition même de cette frontière que les différents acteurs ne s'entendent pas. Des dizaines de chartes existent²⁶⁶, plus ou moins contraignantes, et qui n'engagent personne. C'est pourquoi, selon certains référenceurs, il n'y a finalement que Google qui ait véritablement le pouvoir de fixer la limite :

*« C'est quand même les moteurs de recherche qui décident ce qui est Black Hat de ce qui ne l'est pas. Jusqu'à peu, la vente de liens n'avait rien de Black Hat. »*²⁶⁷

*« Les Black Hat SEO n'ont pas peur de se faire blacklister un nom de domaine et ça ne les dérange pas d'aller poser des liens partout, même là où ça peut gêner le webmaster. Après on peut aussi fixer une limite qui serait celle des règles des moteurs de recherche (et en particulier Google). »*²⁶⁸

D'autres observateurs, enfin, prétendent que les *black hats* sont ceux qui sont prêts à tout pour augmenter la visibilité de leurs documents, sauf peut-être à transgresser la loi édictée par le gouvernement du pays dans lequel ils agissent. Selon ces observateurs et ces praticiens, l'immense majorité des référenceurs ne respectent pas scrupuleusement les préconisations de Google, mais ne peuvent pas pour autant être qualifiés de *black hats*. Il existerait donc une couleur de chapeau intermédiaire, les chapeaux gris, ou *grey hats* :

²⁶⁵ « Je dirais que la plupart des gens ont une vision complètement erronée de nous. Nous ne sommes pas des méchants qui sabotons le travail d'autrui. Nous avons juste une approche différente de la manière dont on peut arriver à bien se classer sur les requêtes des moteurs de recherche. », source : <http://www.visibilite-referencement.fr/blog/xxx-black-hat-seo> (dernière visite le 22 août 2013). « « Ça devient un peu plus subtil lorsqu'on isole le terme Black Hat qui émane directement du hacking, ayant pour objectif principal de nuire à autrui. Même si un des leviers à disposition du Black Hat SEO consiste à faire baisser les concurrents dans les résultats de recherche, l'objectif principal vise à valoriser ses propres plateformes ; ce qui n'inclut pas forcément de nuire aux autres. A ce niveau, le terme Black Hat SEO est totalement empreint d'un concept qui ne correspond pas à la réalité. » source : <http://www.visibilite-referencement.fr/blog/laurent-bourrelly-black-hat-seo> (dernière visite le 22 août 2013)

²⁶⁶ Par exemples : http://www.bruceclay.com/eu/web_ethics.htm ; <http://www.seoptimise.com/blog/2010/01/30-black-hat-seo-techniques-you-can-use-ethically.html> (dernière visite le 22 août 2013)

²⁶⁷ <http://www.visibilite-referencement.fr/blog/xxx-black-hat-seo> (dernière visite le 22 août 2013)

²⁶⁸ <http://www.visibilite-referencement.fr/blog/mar-le-parle-de-blackhat> (dernière visite le 22 août 2013)

« [Le black hat] c'est le fait de "pousser à bout" les moteurs de recherche tout en essayant de placer un site en première position par TOUS les moyens jusqu'à une certaine limite (que certains n'ont pas) : la loi. Si je disais que le black hat seo était le fait de ne pas respecter les guidelines de Google, alors tous les référenceurs le seraient. Il y a donc une couleur de chapeau intermédiaire entre le black et le white : le grey hat SEO. »²⁶⁹

Comme nous le voyons dans les citations ci-dessus, les limites du champ de manœuvres SEO sont donc négociées par les référenceurs qui, dans certains cas, peuvent être appelés *black hats*, mais dont le but est en réalité de mieux contrôler et d'optimiser la mise en visibilité de leurs documents sans laisser le dispositif fonctionner exactement comme ses concepteurs l'ont prévu. Ils greffent alors leurs propres cours d'action et leur propre créativité sur le cours d'action et la créativité des ingénieurs de Google, dans le but de défendre leurs intérêts (cf. 8.2).

7.3.2 Les chapeaux gris de la presse en ligne

La plupart des référenceurs interrogés dans le cadre de notre étude nous ont dit qu'ils étaient des « chapeaux gris », c'est-à-dire qu'une partie de leurs actions allait dans le sens des préconisations de Google tandis que les autres allaient dans un sens contraire ou non prévu. Les référenceurs nous ont également dit que leur autonomie dans l'entreprise leur permettait de procéder à certaines actions pouvant être considérées comme relevant de pratiques *black hat* sans avoir nécessairement à en informer leur hiérarchie. Ainsi, ils sont seuls juges de ce qui est acceptable ou non, seuls à connaître de manière approfondie les préconisations de Google et leur marge de manœuvres, ainsi que les risques de sanction dans le cas de l'utilisation de méthodes déconseillées par la firme. Ceux-là arbitrent entre le risque et l'opportunité, tout en jouant sur le fait qu'ils travaillent pour des sites adossés à des marques fortes, dont, supposément, Google a besoin pour offrir à ses utilisateurs un service de qualité. Seul le référenceur du Figaro nous a dit qu'il ne pourrait en aucun cas être apparenté à un *black hat* :

²⁶⁹ <http://www.visibilite-referencement.fr/blog/maximedit-blackhat-seo> (dernière visite le 22 août 2013)

« Moi je suis un pur white hat. Je suis un chevalier Jedi. Le jour où ça marchera à 200%, je trouverai ça plus glorifiant de me dire qu'on a fait quelque chose de propre, et que ça a payé. »

Lefigaro.fr — Réferenceur — Mickaël Vuillaume

Les référenceurs nous ont tous dit qu'ils n'agissaient pas dans le sens contraire des intérêts du lecteur, conformément à ce qu'ils nomment eux-mêmes leur « déontologie » ou leur « éthique ». Leur principal objectif est de concilier les visions des porte-parole de Google et des journalistes quant à ce qui constitue l'intérêt des internautes. S'ils ne faisaient rien, certains éditeurs dont le contenu n'a pas de valeur ajoutée mais qui pratiquent un SEO intensif pourraient potentiellement remonter dans les classements au-dessus des contenus produits par les journalistes.

« D'après notre ami Matt Cutts, tu fais un contenu de bonne qualité et les links viennent tous seuls. C'est faux. Si tu ne fais pas de SEO, les links ne viendront pas, et tu apparaîtras derrière ceux qui ont fait du SEO, parce que c'est comme ça que ça marche, un point c'est tout. »

Resoneo – Consultant Senior — Julien Crenn

Ainsi, certains référenceurs estiment que s'ils ne faisaient pas de SEO, et s'ils n'allaient pas parfois dans le sens contraire ce qui est préconisé par Google, les sites dont le contenu est de piètre qualité mais qui pratiquent le *black hat* de façon intensive réussiraient à tromper l'algorithme et à se retrouver en haut des listes de liens. Désobéir à Google pourrait donc signifier dans certains cas servir l'internaute et, étant donné les effets de réseaux, servir aussi Google puisque la firme a intérêt elle-même à satisfaire ses utilisateurs. Une telle observation est confirmée par les travaux des économistes Ron Berman et Zsolt Katona selon lesquels le SEO peut profiter à un moteur de recherche et à ses utilisateurs dès lors que la valeur que le site attribue au trafic est proche de la valeur que les internautes attribuent au contenu (Berman et Katona, 2011). Dans le cas contraire, les économistes affirment que les pratiques SEO pénaliseraient aussi bien le propriétaire du moteur que ses utilisateurs. Ainsi, si un éditeur produit un contenu valorisé par les internautes et pratique le SEO pour remonter dans les résultats, y compris certaines méthodes *black hat*, cela pourrait avoir un effet positif pour l'ensemble des acteurs engagés.

7.3.3 Tours et détours du dispositif

Les ingénieurs de Google voudraient idéalement que les éditeurs fassent comme si le moteur de recherche n'existait pas²⁷⁰. Ainsi, les concepteurs pourraient paramétrer les crawlers, l'indexeur et l'algorithme sans que cela ait d'influence sur les sites et les contenus. Et dès lors, les caractéristiques de la liste seraient le résultat de leur seule interprétation. Cependant, étant donné le réservoir de trafic que constitue le moteur, certains éditeurs peuvent modifier leur cours d'action précisément dans le but d'y être référencés le mieux possible. C'est pourquoi les employés de Google vont chercher à modifier les critères de manière à minimiser l'impact de ce qu'ils désignent comme étant des pratiques *black hat*, tout en menaçant de punir ces pratiques par un biais systématique.

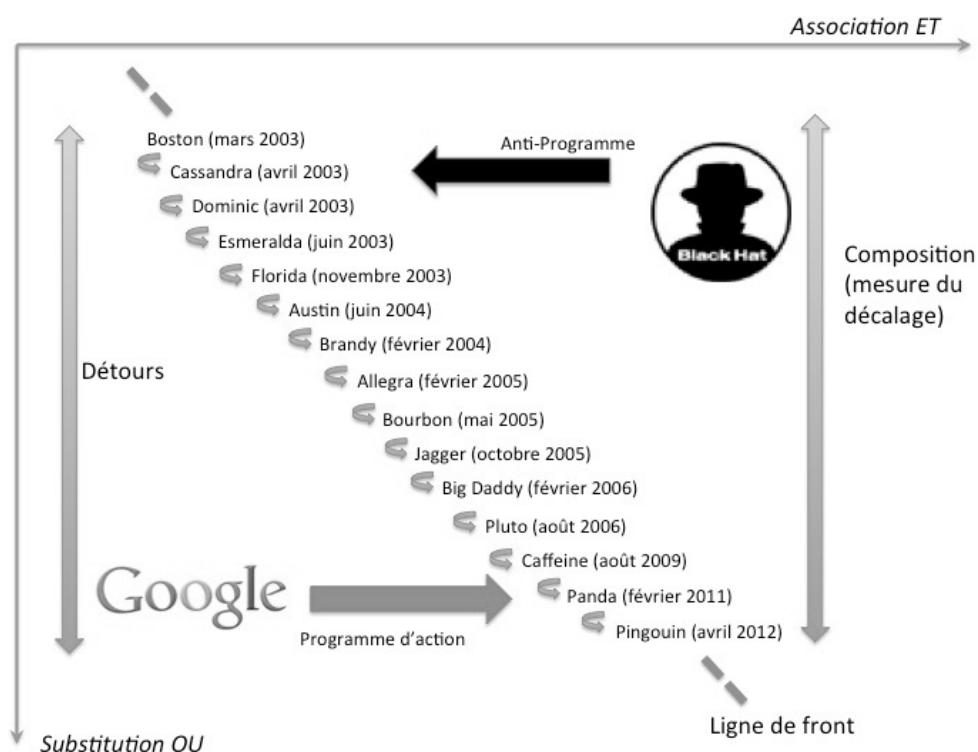
Si nous représentons schématiquement les principales modifications d'algorithme de ces dix dernières années (Fig. 32), nous obtenons un « graphe sociotechnique » (Latour *et al.*, 1991 ; Latour, 2010), où nous voyons schématiquement comment Google, pour continuer à associer un maximum d'acteurs au cours d'action des concepteurs, ainsi que pour défendre sa propre définition de l'échelle de pertinence et des intérêts des internautes, fait subir aux caractéristiques du dispositif plusieurs détours. A chaque changement d'algorithme, le dispositif devient un « nouveau dispositif », qui se substitue à l'ancien (axe vertical : OU), pour maintenir l'association de Google à ses partenaires *white hat* et éventuellement en rallier de nouveaux (axe horizontal : ET). La liste de liens est alors produite *autrement*, parce que le crawler, l'indexeur, l'algorithme, ou les trois à la fois, ont changé. La médiation technique au sens de l'ANT, et en particulier l'étape de *composition* (cf. 1.3.5), n'est plus la même. Pourtant, l'objectif de cette médiation, lui, n'a pas changé, en cela qu'il s'agit toujours pour les concepteurs de hiérarchiser les documents selon les critères de pertinence qu'ils auront définis. En effet, si les méthodes SEO non voulues par la firme se répandent, les listes de liens ne

²⁷⁰ « “Idéalement, personne ne devrait jamais avoir besoin d'apprendre le SEO” explique Matt Cutts. “Mais le fait est qu'il existe des gens qui vont essayer de se promouvoir eux-mêmes, aussi voulons-nous prendre part à la conversation et dire “Voici certaines choses éthiques que vous pouvez faire. Voici certaines choses qui sont très risquées. Evitez ces dernières.” » (Levy, 2011)

correspondront plus à ce que les concepteurs jugent pertinent, et ceux-là auront perdu en partie leur *pouvoir* de hiérarchiser les pages web. C'est donc bien pour rester fidèle à leur projet, leur vision, leur script, que les concepteurs modifient le dispositif.

Plus les modifications d'algorithme seront nombreuses et plus la matérialité du dispositif s'éloignera de ce qu'elle était au départ, tout en diminuant la portée des actions *black hat*. Plus les actions des éditeurs *black hat* seront contrées efficacement par Google et moins ceux-ci occuperont de place dans les résultats, permettant à la firme de mieux s'associer aux éditeurs *white hat* qui verront leurs efforts récompensés et seront ainsi encouragés à continuer d'agir conformément à ce qui leur a été recommandé par les concepteurs.

Figure 31. Graphe sociotechnique : modifications d'algorithme (2003-2012)



Google se heurte aux *black hat* dont les pratiques constituent autant d'épreuves susceptibles d'interrompre son programme. En prenant en compte un nouveau signal d'entrée, c'est-à-dire en ajoutant à l'algorithme un critère doté d'une pondération diluant les pondérations des critères jusqu'ici pris en compte, Google contourne les anti-programmes et compose un nouveau dispositif — toujours dans le même but, au nom du même projet. Ici, les noms et dates correspondent à des modifications d'algorithme opérées par Google pour contrer les pratiques *black hat*.

Matt Cutts, ingénieur en charge de la lutte *anti-spam* chez Google, a deux fonctions principales. D'une part, il propose des modifications de l'algorithme en fonction de ce qu'il sait (et qu'il définit lui-même comme étant) des pratiques *black hat*. D'autre part, il communique à propos des mesures visant à enrayer ces pratiques, et, plus généralement, à propos de tout ce qui concerne l'algorithme d'appariement. Le fait même d'appeler son action « lutte *anti-spam* » n'est pas neutre, puisque le vocable « *spam* » désigne en général les emails non sollicités. Ainsi, appeler « *spam* » le *black hat* revient à en désigner la pratique comme un élément qui gênera le confort des utilisateurs. Le discours tenu par Matt Cutts s'ajoute au reste du dispositif et participe, autant que les changements d'algorithme, de l'empilement des détours. Certains référenceurs interrogés dans le cadre de notre enquête prétendent d'ailleurs que les modifications d'algorithme ne sont parfois rien d'autre que des opérations de communication. Ainsi, selon le référenceur de *Lexpress.fr*, la première version de la modification « Panda »²⁷¹ n'aurait été qu'une opération de communication qui aurait permis à Google, sous prétexte d'avoir modifié son algorithme, d'évincer manuellement certains sites des classements.

« De nombreux sites qui auraient dû être punis par Panda ne l'ont pas été. En fait, seulement les sites les plus connus ont été punis. Donc je crois que Panda n'est pas un vrai changement d'algo. C'est une opération de com' pour pouvoir relayer à la main les sites et dire : ce n'est pas notre faute, nous n'avons rien fait pour punir expressément ces sites-là, mais nous avons modifié la technique, et ces sites ont été punis. Alors qu'en fait, ce sont eux qui ont puni ces sites un par un, à la main. »

Lexpress.fr — Référenceur — Renaud Perrin

Cette citation donne un exemple de ce que peuvent être les soupçons des experts vis-à-vis des modifications d'algorithme qui, même si elles sont commentées publiquement, ne sont jamais expliquées dans le détail par les porte-parole de Google. Ce n'est pas simplement un signal d'entrée et une pondération qui sont ajoutés à l'algorithme, mais également un *énoncé*. Tant que la lumière n'est pas faite à son sujet, l'actant qui s'ajoute ici au dispositif est un hybride composé indifféremment de discours et de technique. Il suscite des spéculations et des suppositions susceptibles

²⁷¹ De nombreuses autres versions de cette modification ont ensuite été implémentées par Google.

de modifier à nouveau les cours d'action, ou bien au contraire de les garder inchangés. Le dispositif a été transformé, soit parce qu'il a été modifié techniquement, soit parce que l'énoncé qui l'accompagne a été révisé ou complété.

Ces modifications constantes permettent à Google de fortifier son projet en se liant plus fortement à ceux qui respectent ses préconisations, récompensés dans les classements par les places autrefois occupées par les *black hat*. Par exemple, à la suite de la modification « Panda », les fermes de contenus *AssociatedContent.com*, *FindArticles.com*, *EZineArticles.com*, *Articlesbase.com* ont largement perdu en visibilité²⁷². A l'inverse, des sites comme *Newyorker.com*, *Usnews.com*, *Morningstar.com* ont gagné plus de 20% de visibilité²⁷³. Les sites de *Fox News*, *ABC News*, *ESPN*, *The New York Times* et *Yahoo news* ont eux aussi bénéficié de la modification « Panda », tandis que certains sites qui n'avaient pourtant rien de « spams », comme celui du *British Medical Journal*, ont été impactés²⁷⁴.

*

Nous avons décrit ci-avant l'activité du référenceur dans une organisation de presse, le pourquoi de son embauche et la manière dont il peut jouer le rôle de « centre de traduction », influençant la nature et le résultat du processus d'infomédiation, autant que la nature du dispositif influence son action. Nous avons vu que les éditeurs, motivés par la maximisation potentielle de leur trafic, pouvait choisir de se conformer aux préconisations de Google (*white hat*), ou bien d'agir d'une façon non conforme à ces préconisations (*black hat*). Dans le cas particulier des éditeurs de presse en ligne, il se trouve, selon les informations que nous avons récoltées, qu'un arbitrage est effectué entre le respect des préconisations et les pratiques allant à leur encontre (*grey hat*). Avant de voir, dans les chapitres 8 et 9, les effets concrets de l'action du référenceur, nous donnerons ci-après certaines informations quant aux particularités de la stratégie de référencement mise en œuvre sur le site Lexpress.fr.

²⁷² <http://searchengineland.com/winners-losers-from-googles-webspam-update-119493> (dernière visite le 22 août 2013)

²⁷³ *Ibid.*

²⁷⁴ http://news.cnet.com/8301-31921_3-20054797-281.html (dernière visite le 22 août 2013)

7.4 Organisation stratégique de L'Express

Nous pouvons désormais reprendre notre monographie où nous l'avions laissée à la fin de la Partie 1 (chapitre 3) en poussant la porte de Roularta Media France afin de voir ce qu'il s'y passe, ce qu'il s'y négocie, quelles représentations ont cours et quelles stratégies y sont mises en place.

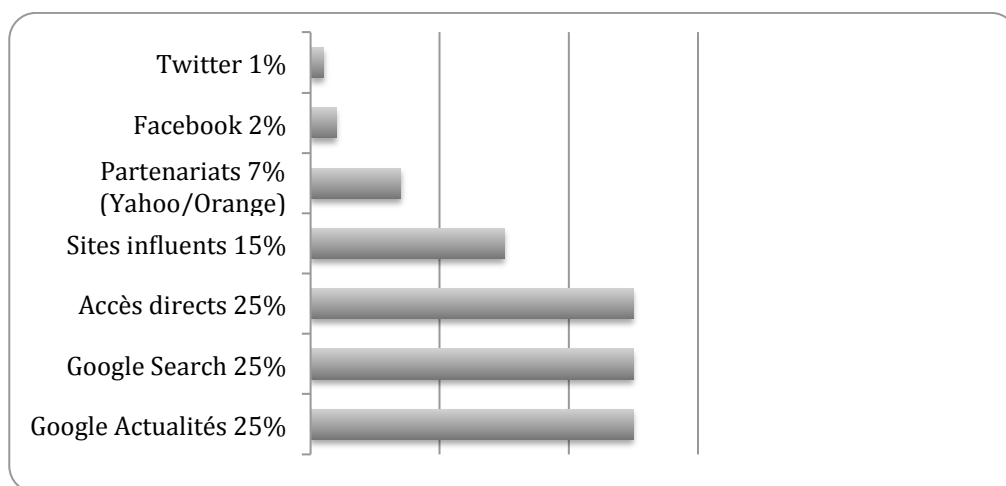
7.4.1 Réservoir de trafic, vitrine

Comme la plupart des SIPI concurrents et comparables, Lexpress.fr est une plate-forme à trois versants sur laquelle les internautes sont subventionnés intégralement (gratuité du service). Plus ils seront nombreux, plus le nombre de pages affichées augmentera et plus les effets de réseaux croisés sur le versant publicitaire seront forts, entraînant une augmentation des recettes versées par les annonceurs à qui les espaces sont facturés au coût par affichage. L'internaute peut ainsi consulter les articles sans avoir ni à payer ni à s'identifier. Le titre vend ensuite l'attention de son lectorat à des annonceurs via des encarts publicitaires facturés au coût par affichage. Les internautes voient également apparaître des liens sponsorisés, rémunérés au coût par clic (c'est alors le trafic qui est vendu et non l'attention), via notamment le réseau de syndication Ligatus et également, plus rarement, des modules AdSense. Contrairement à la publicité par bannières, gérée par la régie interne, la publicité par liens sponsorisés est sous-traitée à deux régies externes concurrentes, spécialisées en liens sponsorisés, Ligatus faisant l'objet d'un partenariat privilégié avec tous les éditeurs du Geste par rapport à AdSense. La part des revenus provenant des liens sponsorisés demeure toutefois négligeable comparée à celle des bannières.

Au final, d'après les estimations de Jean-Marie Charon et Patrick Le Floch, chaque visiteur unique du site Lexpress.fr rapporte en moyenne 2 euros à RMF, chiffre supérieur au coût de revient d'un concurrent pourtant comparable comme le *Nouvel Obs*, à qui chaque visiteur rapporterait environ 1 euro selon ces mêmes estimations, et largement supérieur au coût de revient des titres nés en ligne, comme *Rue89*, à qui un visiteur rapporterait à peu près 0,33 centimes (Charon et Le Floch, 2011, p.71-72). Etant donné un tel modèle, le référencement sur Google Search et Google Actualités est capital pour Lexpress.fr, dans la mesure où le dispositif

constitue un réservoir apportant entre 40 et 50% du trafic du site, lequel se situait en 2012 entre 5 et 5,5 millions de visiteurs uniques par mois.

Figure 32. Structure d'audience approximative de *lexpress.fr* (mai 2012)²⁷⁵



Source : Auteur

Le trafic apporté par Google est monétisé directement étant donné un système de facturation essentiellement basé sur le coût par affichage. Dans ces conditions, l'objectif des dirigeants de RMF est de maintenir un coût élevé pour ses bannières publicitaires.

Il y a une situation de dépendance à Google, en même temps qu'un climat de défiance étant donné la concurrence de la firme sur le marché publicitaire. Nous retrouvons l'idée exprimée plus haut selon laquelle le dispositif répond à une situation temporaire. Selon l'Editrice Sophie Gohier, les négociations avec Google doivent continuer, car aucune issue jugée satisfaisante et stable n'a pour l'instant été trouvée. Du point de vue du rédacteur en chef, Eric Mettout, il existe des arguments recevables aussi bien du côté des éditeurs les plus virulents à l'encontre de Google que des porte-parole de la firme (Mettout, 2012). Eric Mettout nous a également expliqué que s'opposer à Google de manière forte pouvait permettre éventuellement d'être mieux considéré et finalement mieux référencé sur les services de la firme, et

²⁷⁵ Il s'agit d'une approximation, car nous n'avons pas eu hélas aux chiffres exacts. Mais certains acteurs ont tout de même bien voulu nous communiquer l'approximation ici représentée.

donc que c'était nécessaire, tout en nous expliquant que la firme devait être considérée comme un allié auquel l'éditeur avait le *devoir* de s'adapter :

« Montrer les dents, ça ne va rien changer. Après, il y a un rapport de forces, c'est clair. Plus on est fort contre quelqu'un comme Google, et plus il nous considère et mieux on sera traité. C'est de cet ordre là. Après, moi je ne changerai pas d'avis sur le fait que Google n'est pas un ennemi, mais un allié. Si les gens passent par Google pour accéder à l'info, nous on s'adapte à Google, parce que notre métier c'est de trouver des infos, de les rendre visible et de les mettre à disposition du lecteur, c'est tout. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

Lexpress.fr ne vend pas seulement aux annonceurs le nombre de visiteurs et leur ciblage, mais également sa présence sur Google Actualités. Nous l'avons vu dans le chapitre 4, seuls certains sites peuvent être référencés sur le moteur vertical, à condition de répondre à certains critères et d'être intégrés, après soumission d'un formulaire, au champ d'indexation. Or un membre de l'équipe du marketing digital nous a expliqué que cette présence était un « plus » pour les annonceurs.

« On référence sur Google Actualités les pubs de l'annonceur, pas seulement sur Google Search. Donc ce qui est très intéressant, c'est qu'on ne vend seulement à l'annonceur de l'audience pure et dure, mais aussi le droit d'être traité sur un centre de contenus qui est pertinent avec ses enjeux. [...] Il achète le droit d'être passé dans notre moulinette avec tout ce que ça lui donne : la communauté, le référencement, un milliard de raisons intéressantes du point de vue business et chaîne de valeur. »

RMF — Responsable Marketing et Développement — Laurent François

Les annonceurs, en plus de payer pour la visibilité de leurs annonces sur le site de *L'Express* achètent la visibilité des pages de Lexpress.fr sur le dispositif d'infomédiation. Ils se positionnent au niveau « n+2 » de l'arbre de navigation.

7.4.2 Mutualisation des coûts, du *P&L* et du management

Une des originalités de *L'Express* est de mutualiser les coûts et les recettes afférents aux différents sites du groupe. L'éditeur actionne ainsi sa marge de manœuvre économique en jouant sur les synergies :

« Nous pouvons avoir cette qualité aujourd'hui parce nous mutualisons. Au Nouvel Obs, ils n'ont que deux sites, et ils ont donc des coûts structures beaucoup plus importants. L'avantage majeur de mutualiser les coûts comme

nous le faisons est de s'appuyer à fond sur les fonctions transverses, notamment l'acquisition d'audience, le référencement sur Google. C'est leur expertise qui prime. C'est un vrai avantage. Après, l'inconvénient, quand vous avez une marque comme L'Express, forte par rapport aux autres marques, c'est que tout le groupe considère qu'il n'y en a que pour cette marque, et L'Express considère qu'il paye pour les autres. Ce sont des problématiques humaines classiques. »

Lexpress.fr — Editrice — Sophie Gohier

En plus de mutualiser les coûts relatifs aux fonctions transverses, dont la stratégie de référencement fait partie, une des particularités notables de l'ensemble des activités web de Roularta Media France, est d'avoir un poste « Pertes et Profits » (P&L dans la citation ci-dessous, pour *Profit and Loss*) mutualisé :

« La spécificité par rapport aux services qui ont un budget de fonctionnement propre, notamment les titres papiers, c'est que nous, ici, nous avons un P&L global pour les nouveaux médias. Si on considère par exemple que Lentreprise.com dépense trop, en bonne intelligence, on lui dit qu'il a un problème de structure, car on essaye d'avoir une cohérence entre toutes nos marques et de mutualiser. Le P&L global permet de mieux répartir les bénéfices et de mieux endosser les pertes, toujours dans une optique transverse. »

Lexpress.fr — Editrice — Sophie Gohier

De la gestion du sommet de Roularta Media France jusqu'à la gestion de chaque site en particulier, l'organisation est la même. Deux personnes sont chargées conjointement du modèle économique et éditorial du site, sans que l'une d'elle n'ait d'ascendant hiérarchique sur l'autre. Au sommet, Corinne Pitavy est la directrice générale en charge de la partie business et Christophe Barbier le directeur général en charge de la partie éditoriale de l'activité. Au niveau de Lexpress.fr, l'Editrice Sophie Gohier, qui dépend de Corinne Pitavy, est en charge de la partie business et des fonctions transverses, tandis que le rédacteur en chef Eric Mettout, qui dépend de Christophe Barbier, est en charge de la partie éditoriale.

Par rapport au titre papier, Lexpress.fr fonctionne plus étroitement avec le marketing et les fonctions transverses, mais aussi avec les autres sites du groupe, du point de vue éditorial. Par exemple, un film pourra être traité de deux façons différentes par deux journalistes, l'un de *Studio Ciné Live*, l'autre de la rubrique culture de *L'Express*, tandis que sur le web il n'y aura qu'un seul article publié sur la

rubrique « Culture » de Lexpress.fr (le magazine *Studio Ciné Live* n'a donc pas de site à son nom). Cela est dû à la stratégie de « Coupole » dont nous parlerons au chapitre suivant (cf. 8.4.1). Par ailleurs, selon le rédacteur en chef de Lexpress.fr, il existe un contact plus étroit de la rédaction web avec les fonctions transverses qu'entre ces dernières et la rédaction papier.

Dans un tel contexte organisationnel, étant donnée une gestion mutualisée et des divisions rapprochées, la Direction de l'acquisition d'audience, dont dépend le référenceur, joue un rôle de pivot. Elle s'occupe conjointement des sites de Roularta Media France et mutualise les coûts tout en élaborant des synergies dans le but d'augmenter la visibilité de l'ensemble des sites. La Direction de l'acquisition d'audience comprenait 5 employés au moment de nos entretiens : un directeur, un référenceur, un employé en charge de la mesure d'audience et deux employés en charge de la partie « animation et partenariats ».

7.4.3 Méthode agile et coordination

L'agilité est un concept organisationnel issu des sciences de la gestion, désignant « la capacité d'une entreprise à croître dans un environnement marqué par un changement continu et imprévisible d'un marché global, caractérisé par une demande de qualité supérieure, de haute performance, de faible coût et de produits et services correspondant aux exigences des consommateurs » (Barzi, 2011, p. 30). Les notions de *flexibilité*, de *réactivité* et d'*adaptabilité* en constituent le cœur. Même si le mot « agilité » n'a pas été employé lors de nos entretiens, il nous a semblé que la question du référencement était traitée chez RMF selon une méthode agile, dans la mesure où il ne s'agissait pas d'apporter une réponse unique et valable à une question fixée et stable, mais plutôt une multitude de réponses dans la durée, au cas par cas, dans le but d'ajuster, de préciser et de réviser la stratégie de diffusion. L'organisation est ainsi prête à réagir, et le rôle du référenceur consiste à coordonner les actions des métiers du groupe dans le but d'optimiser constamment la visibilité des sites.

Le concept d'agilité vaut également dans le cas de la relation de l'éditeur au dispositif, considéré alors comme une sorte de client auquel il s'agirait de fournir des contenus adaptés dans le but de recevoir, non pas un paiement, mais du trafic.

Le référenceur lui-même fonctionne sur une méthode agile vis-à-vis de ses partenaires dans l'entreprise. Il détecte quelles peuvent être leurs actions, effectue des recommandations et des formations, répond aux questions formelles et informelles, et participe dès le départ à l'élaboration des nouveaux projets techniques, commerciaux ou éditoriaux.

« Je suis une sorte de chef de projet transversal. Je travaille seul et avec tout le monde. Je fais des recommandations techniques. Je forme les journalistes. Je fais de la veille et du testing. Puis je fais de nouvelles recommandations. Et je réponds aux journalistes. Je les conseille. [...] Le SEO, finalement, a un rôle de manager. Je contrôle que ça a été bien fait... Je contrôle que la rédaction utilise les outils qu'on leur a mis en place. »

L'express.fr — Référenceur — Renaud Perrin

D'après l'Editrice Sophie Gohier, l'audience en provenance de Google Search et Google Actualités est « l'enjeu n°1 » pour *L'express.fr*, au moins sur le court terme, ce qui justifie la coordination de l'ensemble des métiers autour du poste de référenceur. Par ailleurs, elle nous a expliqué que la question du référencement était cruciale au point que certains projets jugés intéressants du point de vue éditorial n'étaient pas réalisés si leur potentiel SEO était jugé trop faible :

« L'enjeu n°1 pour nous, c'est l'audience naturelle, donc ça va être la structure du site au sens technique HTML du terme et la stratégie de référencement, au sens large. Il s'agit de faire en sorte que nos contenus soient bien lus par Google. Ça se passe très en amont des projets, et ensuite tout au long. C'est une stratégie d'évolution de la structure de site, et une stratégie de backlinking. C'est le job de Renaud au quotidien [i.e. du référenceur]. [...] Chaque projet qui doit être pensé en fonction de son potentiel SEO. On prend des décisions parfois de ne pas lancer des projets parce qu'on n'y voit aucun potentiel d'audience naturelle, alors qu'il y aurait éventuellement un intérêt éditorial. »

L'express.fr — Editrice — Sophie Gohier

Pour chaque nouvelle modification du site, le référenceur est intégré au trinôme constitué par un journaliste (en général le rédacteur en chef ou un rédacteur en chef adjoint), un chef de projet fonctionnel et un chef de projet technique. Le prédécesseur du référenceur que nous avons interrogé travaillait isolément de ce trinôme, ce qui, selon les acteurs interrogés, ne servait pas à grand-chose. Cependant, l'arrivée d'un nouveau référenceur en 2009, Renaud Perrin, a changé la manière de procéder. Il s'est

intégré au trinôme dès le départ des projets puis en a suivi le développement, selon une méthode « agile » consistant à fournir des conseils informels plutôt que des ordres stricts. Le chef de projet fonctionnel et le chef de projet technique lui rendaient régulièrement compte de l'état d'avancement des projets. D'autre part, le référentiel interagissait également avec les employés de la division en charge du marketing et de la relation avec les annonceurs auprès desquels, là encore, il est intégré dès le commencement des projets.

7.4.4 Interventions hors ligne

En plus de jouer un rôle de catalyseur du réseau autour de la stratégie de maximisation du trafic, la Direction de l'acquisition d'audience joue un autre rôle, celui-là hors-ligne et plus officieux, concernant directement l'optimisation de la mise en visibilité des contenus sur le dispositif d'infomédiation Google. Ce rôle consiste à maintenir le contact avec les équipes de Google par téléphone ou par mail. Normalement, cela ne peut pas avoir d'impact sur la visibilité des documents. Mais les employés de RMF, comme d'ailleurs d'autres entreprises de notre panel (*20minutes.fr* et *Ouest France Multimédia*), sont pourtant persuadés du contraire.

Lorsqu'un changement d'algorithme est annoncé, le Directeur de l'acquisition d'audience téléphone au siège de Google France dans le but d'obtenir des informations au sujet de ce qu'il faut faire pour ne pas être sanctionné. Les équipes de Google répondent en général qu'elles ne sont pas habilitées à donner des informations à ce sujet. Cependant, selon le Directeur de l'acquisition d'audience Marc Bouguié, un tel contact permet de « se montrer » et de « prévenir » Google. De la même façon, si *Lexpress.fr* est déclassé dans le dispositif, Marc Bouguié prévient les employés de Google et leur demande la raison d'un tel déclassement et ce qu'il faut faire pour y remédier. Il se fait connaître dans le seul but de pouvoir, si jamais un problème survenait, s'appuyer sur l'autorité de la marque pour que les employés de Google fassent en sorte, manuellement, de corriger le tir. Marc Bouguié n'est pas certain que cela a un impact significatif ou direct, mais il juge important d'avoir « aussi avec Google une relation humaine, et pas seulement des contacts machiniques entre nos lignes de code et leurs robots ».

CHAPITRE 8 : LE SITE DE PRESSE COMME HOTE DU CRAWLEUR

« [D]ans l'environnement connecté, il semble qu'une plus grande variété de groupes d'acteurs soient impliqués et qu'ils aient un impact plus direct sur la production et le processus que ce qui avait été observé jusqu'ici dans les salles de rédaction de la presse imprimée et des médias radiodiffusés »

(Boczkowski, 2004, p. 183).

A présent que nous nous sommes intéressés à l'organisation, il s'agit de nous intéresser au site lui-même, aux pages, au texte, au positionnement dans la topographie hypertexte et à l'actualisation par les éditeurs des « pouvoir de faire » et « pouvoir faire » décrits dans le chapitre 4. Nous verrons notamment quels usages sont faits par les éditeurs du texte, de l'hypertexte et de l'architexte, dans le but de performer le dispositif. Et nous verrons comment les SIPI se positionnent dans l'arbre de navigation décrit dans le chapitre 5.

8.1 Devant et derrière la page

Nous nous basons ici sur l'observation de six SIPI comparables : lexpress.fr, lemonde.fr, lefigaro.fr, lenouvelobs.com, leparisien.fr, Lepoint.fr, et en particulier sur notre corpus de 252 articles constitué en octobre 2012. Nous observons à la fois ce que nous voyons sur les pages d'accueil et les pages article, ce qu'il y a derrière ces pages (code, Sitemaps, microdonnées, balises, robots.txt) et intégrons ce que les acteurs nous ont dit à leur sujet.

A plusieurs reprises, nous donnerons l'exemple des articles du 1^{er} octobre, cela car nous avons retrouvé les mêmes architextes dans tous les articles du mois. En effet, certains des éléments observés résultent de procédures systématisées et sont, par conséquent, communs à l'ensemble des pages d'un même site. Le corpus nous a permis de vérifier que c'était bien le cas, tout en révélant certains points qui n'étaient pas systématiques.

8.1.1 Adresses URL

Conformément aux préconisations de Google décrites dans le chapitre 4, les adresse URL des pages des SIPI comportent toutes au moins trois chiffres (cf. 4.5.1.A). On y trouve également le nom de la rubrique où se trouvent l'article et le titre. Sur les sites *lemonde.fr*, *lefigaro.fr*, *leparisien.fr*, *nouvelobs.com* et *lepoint.fr*, la date de publication figure dans l'adresse URL, tandis qu'elle ne figure pas dans les URL de *lexpress.fr*.

8.1.2 Optimisation des titres

Nous avons vu qu'optimiser un titre consistait pour l'éditeur à y placer un mot-clé le plus tôt possible (cf. 4.2). Ainsi, les titres sont en général considérés comme optimisés quand ils sont sous la forme « Mot-clé :... ». Par exemple, « *DSK et l'affaire du Carlton : le parquet ne fait pas appel* » est un titre optimisé. Ceci est d'autant plus intéressant à repérer que les journalistes n'apprécient pas la pratique qui consiste à uniformiser les titres (cf. chapitre 9). Dans notre échantillon, 56 % ont des titres optimisés, répartis comme suit :

Tableau 9. Optimisation des titres

Octobre 2012 (5 articles par jour)	Titre « Mot-clé :... »
<i>Lexpress.fr</i>	69%
<i>Lefigaro.fr</i>	36%
<i>Lemonde.fr</i>	40%
<i>Nouvelobs.com</i>	57%
<i>Leparisien.fr</i>	79%
<i>Lepoint.fr</i>	55%

Tandis que près de 80 % des titres du site du *Parisien* sont optimisés, nous voyons ici que les deux sites des quotidiens *Le Figaro* et *Le Monde*, dont le trafic et la part d'accès directs sont les plus forts, ont sensiblement moins de titres optimisés que les autres.

8.1.3 Disposition de la page

En observant les pages des SIPI de notre panel, on remarque qu'elles ont des structures similaires ou du moins très proches. Or comme l'a fait remarqué le journaliste Vincent Coquaz (Coquaz, 2013), cette structure est semblable à ce que conseille le consultant en SEO Nicolas Plantelin dès lors qu'il s'agit de maximiser l'apport de trafic de Google Actualités (Plantelin, 2013). Pour chaque site, nous trouvons, disposés comme sur la maquette de Nicolas Plantelin, les mêmes éléments : balises <Title>, <H1> et <H2>, date de publication, date de dernière modification, auteur, chapeau, photo en centre de page, liens vers d'autres articles, espace destiné aux commentaires, sigles de partage et de recommandation sur les réseaux sociaux.

Sur chacun des six sites du panel apparaît également un « fil d'Ariane » au-dessus de l'article (Par exemple : Actualités > Monde > Syrie), dont chacun des éléments est cliquable. Cela nous a été présenté, en particulier par le référenceur du *Point*, comme un élément permettant d'optimiser le référencement en rendant plus facile le travail du crawler. Le fil d'Ariane améliore également l'affichage des liens sur le moteur : les rubriques apparaîtront en-dessous du lien sous forme cliquable. Ainsi, il s'agit d'une optimisation qui permet d'améliorer le service que fournit Google à ses utilisateurs, et donc qui est susceptible d'être récompensée par l'algorithme d'appariement.

8.1.4 Méta-tags

Comme nous l'avons vu dans le chapitre 4, les méta-tags permettent d'interrompre la communication avec le crawler, de la fluidifier, de la spécifier ou de la localiser. Ils se situent directement dans le code HTML des pages. Certains sont généralistes et d'autres concernent en particulier les articles liés à l'actualité et référencés sur Google Actualités. Leur usage est systématisé par les sites de presse, dont les concepteurs paramètrent le système de gestion de contenus (*Content Management System*, CMS) afin que les méta-tags appropriés soient attachés à chaque nouvelle page créée par les journalistes. Pour les articles du 1^{er} octobre, nous avons :

Tableau 10. Usage des balises méta-tags (1^{er} octobre 2012)

	Titre de l'article + URL	Balises méta-tags
Lexpress.fr	Mikheïl Saakachvili: "La démocratie géorgienne existe bel et bien" http://www.lexpress.fr/actualite/monde/europe/mikheil-saakachvili-la-democratie-georgienne-existe-bel-et-bien_1167234.html	<title> [...] <meta name="description" content="Le président géorgien Mikheïl Saakachvili, qui a reçu L'Express le 21 septembre, défend son bilan. Et accuse Moscou de basses manoeuvres." /> [...] <meta name="news_keywords" content="Mikheïl, Saakachvili, démocratie, géorgienne, existe, bien" />
Lefigaro.fr	L'Élysée aussi devra se serrer la ceinture http://www.lefigaro.fr/conjoncture/2012/10/01/20002-20121001ARTFIG00409-l-elysee-aussi-devra-se-serrer-la-ceinture.php	<title>L'Élysée aussi devra se serrer la ceinture</title> <meta charset="UTF-8"> <meta name="title" content="L'Élysée aussi devra se serrer la ceinture" /> <meta name="description" content="La présidence de la République verra en 2013 son budget diminuer de près de 4% sur un an, à un peu plus de 100 millions d'euros. Les enveloppes consacrées aux travaux, déplacements et frais de fonctionnement seront les plus réduites." /> <meta name="keywords" content="Actualité économique, entreprises, économie, bourse, emploi, impôts, cac 40, création d'entreprise, chef d'entreprise, grands patrons, consommation, multinationales, privatisation, délocalisations, concurrence, monopole, crise, bourse, licenciements, union européenne, etats-unis, chine, pmi, pme, tpe, salaires, relance, pib, pnb, aides sociales, japon, récession, économie verte, fmi, reprise, croissance, news, actu"> <meta name="news_keywords" content="Elysée, Budget 2013, Conjoncture" /> <meta name="author" content="lefigaro.fr" /> <meta name="copyright" content="lefigaro.fr" /> <meta name="robots" content="index,follow,noarchive" />
Lemonde.fr	Tabac : une hausse pour rien ? http://www.lemonde.fr/societe/article/2012/10/01/tabac-une-hausse-pour-rien_1764475_3224.html	<title>Tabac : une hausse pour rien ?</title> <meta name="description" content="Ce lundi, le prix des paquets de cigarettes augmente de 40 centimes. Beaucoup craignent que cette mesure n'ait qu'un effet marginal sur la consommation."> <meta name="robots" content="index, follow, noarchive">
Nouvelobs.com	GEORGIE. Des législatives sous très haute tension http://tempsreel.nouvelobs.com/monde/20120928.OBS3919/georgie-des-legislatives-sous-tres-haute-tension.html	<meta charset="utf-8"> <meta name="robots" content="index, follow, noarchive"> <meta name="Description" content="Infraction à la loi électorale, censure des députés, scandale dans les prisons et Larry King !> <meta name="Keywords" content="actualités, actus, news, informations, actualité internationale, Nouvel Obs en temps réel, actualité sociale, politique"> <meta name="#ffffff" content="msapplication-tilecolor"> <meta name="http://tempsreel.nouvelobs.com/themes/header/images/logo-nobs.png" content="msapplication-tileimage"> <meta name="news_keywords" content="Arménie, GEORGIE"> <title>GEORGIE. Des législatives sous très haute tension - Le Nouvel Observateur</title>

Leparisien.fr	Florange : ArcelorMittal donne 60 jours au gouvernement pour trouver un repreneur http://www.leparisien.fr/economie/florange-arcelormittal-devrait-annoncer-la-fermeture-des-hauts-fourneaux-01-10-2012-2194353.php	<pre><title>Florange : ArcelorMittal donne 60 jours au gouvernement pour trouver un repreneur</title> <meta name="description" content="La direction d'ArcelorMittal qui a annoncé lundi en fin de matinée lors d'un comité central d'entreprise (CCE) la fermeture définitive des haut" /> <meta name="keywords" content="" /> [...] <meta name="news_keywords" content="" /></pre>
Lepoint.fr	Promettre l'impossible, c'est mentir aux Français ! http://www.lepoint.fr/politique/parti-pris/promettre-l-impossible-c-est-mentir-aux-francais-01-10-2012-1512014_222.php	<pre><title>Promettre l'impossible, c'est mentir aux Français ! - Le Point</title> <meta itemprop="description" content="Hervé Gattegno, rédacteur en chef au "Point", intervient sur les ondes de RMC du lundi au vendredi à 8 h 20 pour sa chronique politique "Le parti pris";" /> <meta name="news_keywords" content="PS, François Hollande, Jérôme Cahuzac, Jean-Marc Ayrault, Pierre Moscovici, Hervé Gattegno, chronique" /> [...] <meta name="author" content="Le Point, magazine" /> <meta name="Identifier-URL" content="http://www.lepoint.fr" /> <meta name="robots" content="index, follow" /> <meta name="publisher" content="Le Point" /> <meta name="copyright" content="Le Point" /></pre>

Nous voyons ici que la balise <Title> est systématiquement implémentée, ainsi que la balise < meta name="description" content=... >, de manière à contrôler ce qui apparaîtra sur le lien hypertexte généré par un service comme Google : le titre et le chapeau de l'article. La balise « mots-clé » généraliste (*keywords*), qui n'est pas lue par le crawler de Google, est implémentée sur les seuls sites du *Figaro* et du *Nouvel Observateur*, tandis que la balise « mots-clés actualités » (*news_keywords*), prise en compte par Google Actualités, est implémentée sur les sites de *L'Express*, du *Figaro*, du *Nouvel Observateur*, du *Parisien* et du *Point*. Seul le site du *Monde* ne l'utilise pas. Le site du *Parisien*, quant à lui, même s'il a implémenté les deux balises « *keywords* » et « *news_keywords* » les laisse systématiquement vides, soit parce que l'interface de publication ne prévoit pas de systèmes de tagging permettant aux journalistes de remplir ce champ, soit parce que les journalistes le peuvent mais ne le font jamais. Notons enfin que *Le Figaro*, *Le Monde* et *Le Nouvel Observateur* sont les seuls à demander de ne pas proposer de version « cache » de leurs articles grâce au méta-tag « noarchive ». Les autres SIPI laissent les crawlers suivre les liens, indexer le contenu et l'archiver sur les serveurs de leurs propriétaires, lequel pourra alors proposer une version « cache » à ses propres clients/utilisateurs.

Dans l'ensemble, l'usage fait des balises méta-tags est très permissif. Google est libre par exemple, en l'absence d'indication contraire, de proposer une traduction et d'afficher des *snippets*²⁷⁶ pour les pages crawlées. Par ailleurs, à part la balise « news_keywords », nous avons remarqué dans l'ensemble de notre corpus l'absence notable de certains méta-tags spécifiques à Google Actualités et pourtant mis en avant avec insistance par les porte-parole de Mountain View. En particulier, aucun des SIPI observés n'utilise ni la balise « syndication-source » ni la balise « standout ». Cela nous a étonnés, notamment pour « standout » dont le but est de permettre à un éditeur d'indiquer au crawleur que c'est lui l'auteur original du document et qu'il considère que ce contenu est important. Selon Renaud Perrin, le référenceur de *Lexpress.fr* au moment de notre étude (maintenant consultant chez *Resoneo*), aucun site français n'utilise cette balise parce qu'elle ne serait reconnue pour l'instant qu'aux Etats-Unis. Par ailleurs, toujours selon Renaud Perrin, sur les 100 sites français les plus présents dans Google Actualités, seuls 14 sites utilisent de manière récurrente la balise « news_keywords », qui est pourtant susceptible d'envoyer « un signal positif à Google News » (Perrin, 2013).

8.1.5 Microdonnées

Les microdonnées, rappelons le, permettent d'enrichir les liens générés par le moteur et ainsi pour un éditeur, en plus d'optimiser potentiellement son référencement, de mieux contrôler l'apparence des liens vers son contenu. Leur usage est systématique dès lors qu'elles sont intégrées au système de gestion de contenus (CMS). Il ressort de nos observations une utilisation relativement différente des microdonnées selon les SIPI auditionnés. Seuls *L'Express*, *Le Monde* et *Le Point* ont implémenté le protocole spécifique aux articles de presse (schema.org/NewsArticle). Cependant, ils n'utilisent pas les fonctions spécifiques aux contenus issus de la presse imprimée que sont « *printEdition* », permettant de désigner le numéro du journal qui contient un article, « *printPage* », permettant de dire à quelle page ce contenu se trouve dans le titre imprimé, ou encore « *printSection* », permettant de désigner la

²⁷⁶ Pour rappel, le « snippet » est la partie textuelle non cliquable du lien qui apparaît sur Google Search ou sur Google Actualités (cf. chap. 4).

rubrique du titre imprimé où se trouve l'article. Ainsi, comme avec les méta-tags décrits plus haut, il semble que les éditeurs ne se saisissent pas des microdonnées concernant leur activité en particulier. En conséquence, l'usage que font *L'Express*, *Le Monde* et *Le Point* de « schema.org/NewsArticle » est exactement le même que celui qu'ils auraient pu faire de « schema.org/Article ».

Le Figaro segmente quant à lui l'usage du protocole en utilisant les microdonnées spécifiques aux articles (<http://schema.org/Article>), aux images (<http://schema.org/ImageObject>) et à l'auteur (schema.org/Person). Les fonctions utilisées sont notamment « *mainContentOfPage* », « *headline* », « *dateModified* », « *datePublished* », « *author* » et « *about* » pour l'article, les fonctions « *primaryImageOfPage* » et la fonction « *caption* » pour l'illustration. Cela permet potentiellement à Google de générer un lien où figurera, en plus des éléments usuels, les dates de publication et de modification, une légende pour l'image et le nom de l'auteur (donnant ainsi la possibilité à Google de respecter le droit à la paternité qui, comme nous l'avons vu dans le chapitre 6, pose problème dans les pays attachés au droit d'auteur). Par ailleurs, certaines microdonnées ne sont pas là pour enrichir les liens générés par le dispositif mais pour permettre à l'éditeur de mieux contrôler la diffusion de ses contenus, en s'assurant notamment que ce sera la bonne image qui apparaîtra près de son texte, et que la description figurant dans le *snippet* sera celle qu'il aura choisie plutôt qu'une portion de texte aléatoirement sélectionnée. *Le Figaro* est l'éditeur qui a le plus recours aux microdonnées. Le site du *Parisien*, quant à lui, les utilise très peu : en fait, elles lui servent simplement à désigner les zones de commentaires. Les autres sites utilisent les fonctions les plus usuelles du schema.org, notamment les indications quant à la date de publication et de modification ainsi que la désignation du titre principal et de l'auteur. Les SIPI de notre échantillon ne se saisissent donc pas d'un grand nombre d'outils qui leur permettraient de prendre la main partiellement sur l'apparence que pourront avoir les liens générés vers leurs contenus.

Enfin, l'observation des microdonnées révèle que les sites de *L'Express*, du *Nouvel Observateur*, du *Parisien* et du *Point* utilisent tous les trois la balise « *rel=author* » décrite dans le chapitre 4. Mais cette balise n'est pas relié à des

comptes « Google + », condition pourtant *sine qua none* pour que le lien puisse être enrichi par la photo de l'auteur. Ainsi, il semble que cela constitue une possibilité d'optimisation et de contrôle dont les éditeurs ne se sont pas saisis, ou ne se sont saisis qu'à moitié. Cela peut être dû au fait que les journalistes n'aient pas voulu créer leur compte « Google Plus », ou bien au fait qu'ils n'en aient simplement pas été avisés. La seule justification qui nous a été donnée à ce sujet l'a été par le référenceur du site du *Point*. Selon elle, la balise « rel=author » sert avant tout à Google à inciter un grand nombre de producteurs de contenus à adhérer à son réseau social dans le but de concurrencer Facebook. La firme se servirait ainsi du désir de l'auteur de voir apparaître sa photo dans les listes de résultat comme d'un mécanisme d'intéressement, afin de mobiliser de nouveaux membres dans son réseau social. Nous retrouvons l'idée que la firme puisse utiliser son quasi-monopole sur le marché des moteurs pour être davantage concurrentielle sur les autres marchés où elle opère. Par ailleurs, le référenceur du *Point* nous a également dit qu'une des causes qui pouvait freiner l'implémentation du « rel=author » était liée au fait que la photo et le sous-titre apparaissant dans la liste de résultats pointent vers « Google + », et non vers le site de presse. Cela risquerait par conséquent de détourner une partie des internautes du contenu original.

Sur les cent sites d'information testés en janvier 2013 par Renaud Perrin grâce à un outil de l'agence *Resoneo* (Perrin, 2013), il se trouve que quatorze sites français seulement ont implémenté la balise « rel=author » :

- <http://tempsreel.nouvelobs.com>
- <http://www.leparisien.fr/>
- <http://www.lepoint.fr/>
- <http://www.lexpress.fr/>
- <http://www.24matins.fr/>
- <http://www.challenges.fr/>
- <http://www.stars-actu.fr/>
- <http://www.huffingtonpost.fr/>
- <http://web-tech.fr/>
- <http://www.newsring.fr/>
- <http://www.terrafemina.com/>
- <http://www.pcinpact.com/>
- <http://www.journaldunet.com/>
- <http://www.gizmodo.fr/>

Nous retrouvons dans cette liste les quatre sites de notre échantillon, qui ne l'ont pourtant implémentée que partiellement. En fait, onze des quatorze sites mentionnés ici n'implémentent la balise que pour désigner la page d'un auteur en ligne sur le site, ce qui permet de relier chaque article à son auteur (tous les articles signés par un auteur apparaîtront sur sa page) mais ne permet pas d'enrichir le lien généré par Google Search ou Google Actualités.

Les trois seuls sites d'information français à relier le dispositif de signature numérique aux comptes Google Plus des journalistes sont web-tech.fr, pcimpact.com et gizmodo.fr, tous trois spécialistes des thématiques ayant trait aux nouvelles technologies. Il semble ainsi que ces sites, étant donné une culture proche de celle de Google, soient davantage impliqués dans le réseau.

8.1.6 Robots.txt

Comme nous l'avons dit dans le chapitre 4, le robots.txt est une ressource implémentée à la racine du site. Il s'agit du premier élément que le crawler visite, donnant des indications à propos du contenu et de ce qu'il peut en faire, et en particulier du contenu qu'il ne faut pas scanner. A moins que cela soit explicitement spécifié (ce qui n'est pas le cas dans notre échantillon), le crawler « Googlebot » n'est pas le seul destinataire des informations contenues dans le robots.txt.

Nous avons visité une à une les pages mentionnées par les différents robots.txt de notre échantillon. En réalité, la plupart d'entre elles n'existent plus. Il est donc en théorie inutile de les mentionner dans le fichier robots.txt. Les articles dont les pages existent encore sont en général relativement datés. Par exemple, l'article de *L'Express* à propos de la mise en examen du Président de Skyrock Pierre Bellanger en 2008, figurant dans le robots.txt, n'est plus d'actualité dans la mesure où le procès, en première puis en seconde instance, est désormais clos. Par ailleurs, pour certains articles, au *Point* par exemple, leur présence dans le robots.txt est liée à l'action des journalistes : dès que l'un d'eux retire un article mis en ligne, l'URL de la page est automatiquement intégrée au protocole d'exclusion. Cela va dans le sens du droit à la paternité et du droit de repentir des auteurs (cf. chapitre 6).

Finalement, l'usage du robots.txt est assez restreint dans le cas des sites auditionnés. Les fonctions spécifiques comme le délai de crawling, ne sont pas utilisées. Le robots.txt du site du Figaro, en particulier, est presque vide. Pour les cinq autres, le protocole est utilisé afin d'exclure les pages qui ne concerneraient pas directement l'actualité, comme celles où figurent les offres d'abonnements, les mentions légales et les pages « contact ». Le site du *Point* se distingue des cinq autres en cela qu'il exclue des formats spécifiques : (.ico ; .exe ; .js ; .css ; .zip ; .xls). Cette exclusion est directement liée au SEO : il s'agit d'écarter le crawler de formats qui risqueraient de surcharger la bande passante, afin que la fréquence de crawling soit maintenue élevée et de maximiser en valeur absolue le nombre de documents indexés.

L'Express, quant à lui, exclue du champ de *crawling* sa chaîne vidéo (<http://www.lexpress.fr/chaine-video-lexpress/>). Là encore, il s'agit de ne pas surcharger la bande passante. Enfin, notons qu'aucun des six SIPI n'a implémenté le protocole ACAP décrit dans le chapitre 6 et les référenceurs, quand nous les avons interrogés, n'en avaient même pas connaissance.

8.1.7 Sitemaps XML

Il est possible pour un éditeur de transmettre les *Sitemaps*²⁷⁷ directement à Google via les *Outils pour les Webmasters*, sans que ceux-ci ne soient accessibles à un observateur extérieur. Il s'agit dans ce cas d'une forme de communication privée entre l'éditeur et Google. Ainsi, même si nous savons que tous les sites de notre panel utilisent les *Sitemaps*, et en particulier le *Sitemap* Actualités, nous n'avons pu accéder qu'à certains d'entre eux. Selon les référenceurs interrogés, l'utilisation des *Sitemaps* est une clé de voûte de leurs stratégies SEO. C'est pourquoi ils ne les rendent pas forcément publics.

Même s'il ne souhaite pas en divulguer la teneur, le référenceur du *Point* nous a donné certaines informations intéressantes à propos de sa stratégie *Sitemap* :

²⁷⁷ Pour rappel, un *Sitemap* est un sommaire permettant au crawler de savoir exactement sur quelles pages il devra se rendre (cf. 4.1.4 et 4.5.2.A).

« Moi j'ai une technique : je fais une Sitemap par thématique. Par exemple, on fait un index de Sitemap pour tous les Sitemaps (une centaine), donc un méta-Sitemap, et un sous Sitemap spécialisé sur les rubriques « Economie », « Bourse », « Société ». Comme ça, on fournit à Google 600 000 URL et lui il en prend en compte environ 550 000. Le fait de segmenter ça permet de voir où t'as des problèmes précisément. On s'est rendu compte par exemple que le Sitemap des archives était mal crawlé ou pas du tout par Google, alors on a fait des sous-Sitemaps dans les Sitemaps archives en fonction des catégories, pareil que pour le reste, par rubriques, et ça marche mieux maintenant.

[...]

Dans les Sitemaps, on a mis aussi des mots-clés. On a demandé aux rédacteurs de rentrer des mots clés. J'ai créé une documentation pour leur expliquer. Par exemple : prénom + nom dès qu'il s'agit de personnalités. S'il y a une géolocalisation, aussi, il faut le mettre, par exemple s'il fait un meeting à Paris, parce que c'est important pour Google. Je préfère que les bonnes pratiques soient libres, donc je n'oblige pas les journalistes, mais je leur montre. Il y a une liste pour Google, mais je préfère que les journalistes fassent comme ils veulent, en fonction de ce qu'ils jugent préférable.

[...]

Il y a un Sitemap pour les images, un pour les vidéos, un pour les news. Après, nous n'hébergeons pas les vidéos sur Youtube, mais sur Ipost, qui est une solution d'hébergement, comme ça la plupart des vidéos restent sur le site, ce qui est une bonne chose. Parce que quand tu les héberges sur Youtube, c'est compliqué pour ce qui est des droits. Donc on les garde de notre côté, ça empêche que Google se fasse de l'argent avec de la pub sur nos vidéos. »

Lepoint.fr — Laure Aubry — Référenceur

Plusieurs enseignements peuvent être tirés de ce verbatim. D'une part, il est possible de mettre en place une stratégie de segmentation des *Sitemaps*, ce qui constitue pour l'éditeur une manière de se saisir de la granularité de son champ de manœuvres, et de performer avec d'autant plus de précision l'action du dispositif. Une telle segmentation lui permet qui plus est de s'affranchir de la limite de 1000 articles par *Sitemap* fixée arbitrairement par Google. L'éditeur se saisit alors d'un « pouvoir faire » grâce auquel il contourne une des dispositions souhaitée par le concepteur. D'autre part, nous constatons que ce sont les journalistes qui renseignent certains champs du *Sitemap*. Des consignes leur sont données, qu'ils sont libres de respecter ou non. Nous voyons ici que l'optimisation du site pour Google Search et Google Actualités dépend des actions coordonnées de plusieurs acteurs et que, dans le cas où les journalistes ne participeraient pas au projet du référenceur, alors la granularité du champ de manœuvres serait sous-utilisée. Enfin, cette citation nous

montre pourquoi la tentative de Google de maximiser le degré d'enrôlement des éditeurs en les incitant à utiliser sa plate-forme Youtube est enrayée, étant donné un partage des revenus entre la firme et les éditeurs non désiré par ces derniers.

Nous avons pu accéder aux *Sitemaps* généraliste et spécifique du *Parisien*, tous deux publics²⁷⁸. Le *Sitemap* généraliste est un simple répertoire de l'URL des articles publiés et de la date de modification pour chacun d'eux. Dans le *Sitemap* Actualités, les balises usuelles sont utilisées, de manière à indiquer le titre, la langue, la date de publication et les mots-clés de chaque article. En revanche, la balise spécifique `<news:genres>` n'est jamais utilisée, c'est-à-dire que le Googlebot comprend systématiquement qu'il s'agit d'un article qui n'est ni une opinion ni une satire.

Le *Sitemap* généraliste du *Nouvel Observateur*, lui aussi accessible²⁷⁹, nous a révélé que le référenceur y procédait à une première segmentation des *Sitemaps* par année. Ainsi, on trouve un *Sitemap* généraliste pour chaque année d'existence du site, de 1999 à 2013. Celui de 2013 révèle une utilisation semblable à celle du *Parisien*. En revanche, pour *Le Nouvel Observateur*, nous n'avons pas eu accès au *Sitemap* Actualités, directement transmis à Google.

Lemonde.fr laisse, quant à lui, accès à l'ensemble de ses *Sitemaps*. On observe dans le *Sitemap* Index²⁸⁰ une segmentation fine qui permet au *Monde*, à l'instar du *Point*, de s'affranchir de la limite des 1000 adresses URL, et de donner des indications différentes à Google. La segmentation est faite à partir de la date, sur un modèle proche de celui du *Nouvel Observateur*, mais aussi à partir du type de contenu : articles, images, vidéos ou contenus destinés aux mobiles. Ainsi, il semble que le titre utilise autant que possible la granularité qui lui est offerte par le champ de manœuvres. *Le Monde* a notamment la particularité d'avoir un *Sitemap* Articles²⁸¹ qui lui permet de préciser à Google que l'actualisation a lieu au moins une fois par jour. Par ailleurs, quand on observe son *Sitemap* Actualités²⁸², il semble que le Monde joue bien plus sur la granularité que ne le font ses concurrents. Notamment, le titre est le seul à utiliser la balise « `news:genres` » permettant d'indiquer à Google s'ils s'agit

²⁷⁸ <http://www.leparisien.fr/sitemap.xml> et http://www.leparisien.fr/sitemap_news.xml (dernière visite le 22 août 2013)

²⁷⁹ <http://tempsreel.nouvelobs.com/sitemap.xml> (dernière visite le 22 août 2013)

²⁸⁰ http://www.lemonde.fr/sitemap_index.xml (dernière visite le 22 août 2013)

²⁸¹ Par exemple : par exemple : <http://www.lemonde.fr/sitemaps/articles/2012-47.xml.gz> (dernière visite le 22 août 2013)

²⁸² http://www.lemonde.fr/sitemap_news.xml (dernière visite le 22 août 2013)

d'un communiqué de presse (*PressRelease*), d'une opinion (OpEd ou Opinion), d'un contenu amateur (*UserGenerated*), d'un contenu satirique (Satire) ou d'un blog (Blog). Ainsi, pour une prise de position datant du 29 mai 2013 à propos de l'armement par l'Union Européenne des rebelles syriens, le *Sitemap* Actualités était²⁸³ :

Figure 33. Capture d'écran : Sitemap Article Lemonde.fr

Armer les rebelles syriens augmenterait le risque d'enlèvement (29 mai 2013)

```
- <urlset>
- <url>
- <loc>
  http://www.lemonde.fr/idees/article/2013/05/29/armer-les-rebelles-syriens-augmente
</loc>
<lastmod>2013-05-29T14:52:56+02:00</lastmod>
<changefreq>hourly</changefreq>
- <news:news>
- <news:publication>
  <news:name>Le Monde</news:name>
  <news:language>fr</news:language>
</news:publication>
<news:publication_date>2013-05-29T14:52:56+02:00</news:publication_date>
- <news:title>
  Armer les rebelles syriens augmenterait le risque d'enlèvement
</news:title>
<news:keywords>Idées</news:keywords>
<news:genres>Opinion</news:genres>
</news:news>
- <image:image>
- <image:loc>
  http://s1.lemde.fr/image/2013/05/29/0x0/3420103_3_7c9d_apres-l-accord-europe
  levee_cb9971073f7e72694778a1b90249c19f.jpg
</image:loc>
- <image:title>
  Après l'accord européen de lundi sur la levée de l'embargo sur les armes en Syrie,
  toujours pas où se trouve leur intérêt.
</image:title>
</image:image>
</url>
```

*

²⁸³ Il est également intéressant de constater ici que le *Sitemap* Actualités du *Monde* indique à Google qu'il faudrait que les articles qui y figurent soient crawlés une fois par heure grâce à la balise <changefreq>.

Pour conclure cette première sous-partie, nous voudrions souligner que les éditeurs ne se saisissent pas de l'ensemble du champ d'action décrit dans le chapitre 4. Ils *délèguent* en grande partie la capacité qu'ils pourraient avoir d'influencer la hiérarchisation et l'apparence des liens hypertextes. Par ailleurs, un des éléments frappant est la ressemblance des quelques manœuvres effectuées, alors que l'optimisation se joue précisément sur la différenciation. Finalement, le seul outil HTML utilisé avec finesse est le *Sitemap*. De celui-ci, contrairement aux autres, Lemonde.fr fait un usage extrêmement soigné et joue ainsi le jeu méta-éditorial prévu par Google tout en se différenciant de ses principaux concurrents.

8.2 PageRank Sculpting : valorisation par le lien

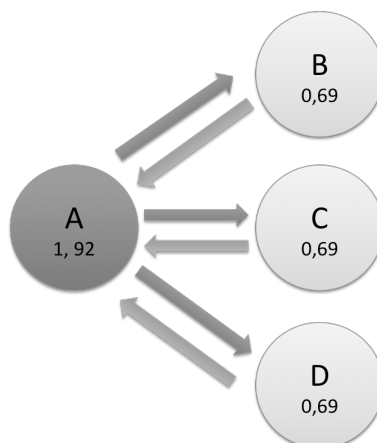
Avant de rendre compte des stratégies de « *PageRank Sculpting* » mises en œuvre, nous présenterons certaines particularités de l'algorithme PageRank, notamment certaines contradictions inhérentes à la formule mathématique, et nous décrirons la manière dont Google essaye d'atténuer les effets de ces contradictions. Puis nous montrerons comment les éditeurs de notre panel actualisent les caractéristiques et les contradictions du PageRank et tentent parfois de contourner le programme d'action des concepteurs dans le but de défendre leurs intérêts en se saisissant de leur capacité d'action. Après quoi nous montrerons avec l'exemple des conjugueurs du *Monde* et du *Nouvel Observateur*, comment de nouvelles formes de valorisation du contenu apparaissent étant donné précisément les stratégies de *PageRank Sculpting* déployées. Enfin, nous situerons nos résultats parmi les études effectuées en SHS à propos de l'usage des liens hypertextes par les journalistes.

8.2.1 Psychanalyse du PageRank

Il est possible de révéler que l'algorithme PageRank, considéré ici comme un *actant* au sens de l'ANT, se *contredit*. En effet, étant donné la manière dont il incite certains acteurs à agir, PageRank risque de ne plus être opérant. Considérons par exemple la page d'accueil A d'un site, pointant vers les pages B, C et D du même

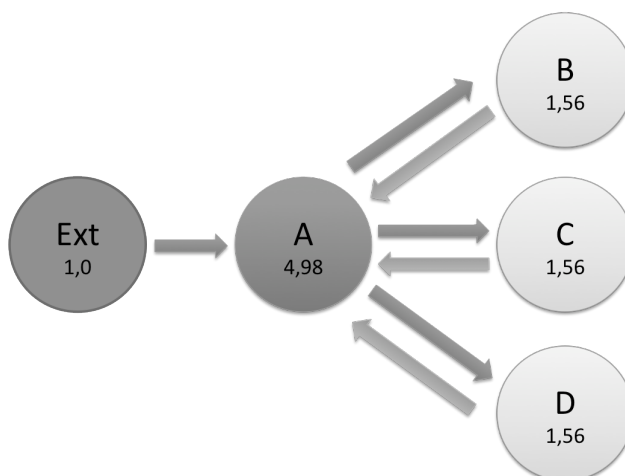
site, elles-mêmes pointant vers la page d'accueil du site en question. Si nous attribuons valeur « 1 » par défaut à chacune des pages, nous obtenons²⁸⁴ :

Figure 34. Calcul du PageRank (1)



Les chiffres affichés représentent le PageRank de chaque page²⁸⁵ tandis que chaque flèche représente un lien. Imaginons maintenant qu'une page « Ext », appartenant à un autre site et elle aussi dotée de la valeur « 1 », pointe vers la page d'accueil. Nous obtenons une augmentation du PageRank de chaque page du site :

Figure 35. Calcul du PageRank (2)

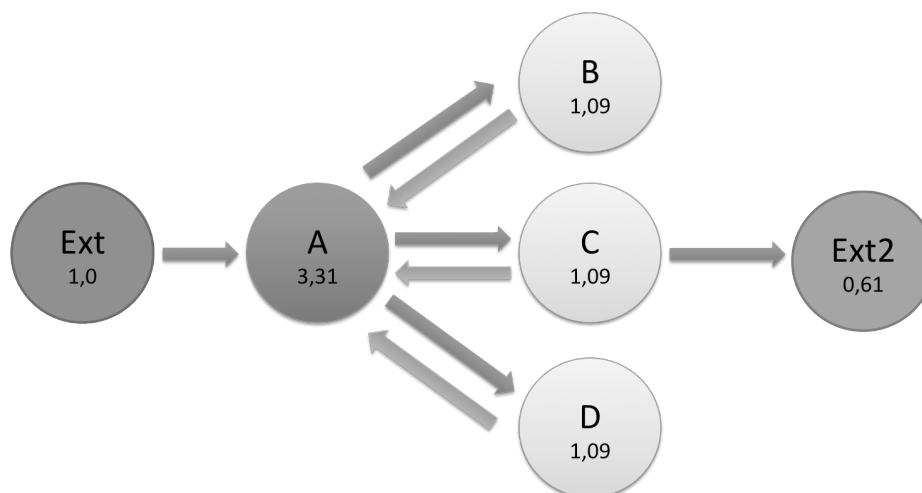


²⁸⁴ Même si nous les avons vérifiés nous-mêmes en effectuant les itérations, les calculs présentés ne sont pas directement de notre fait. Nous les avons trouvés sur le blog de Ian Rogers : <http://www.sirgroane.net/google-page-rank/> (dernière visite le 22 août 2013)

²⁸⁵ La valeur 1 est attribuée par défaut à chaque page au départ. Le PageRank de la page A, $PR(A)$, est calculé selon la formule : $PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$, où $T1, \dots, Tn$ représentent les pages pointant vers la page A, où $C(T1)$ représente le nombre de liens émis par la page $T1$, et où d , appelé « coefficient d'amortissement », a la valeur 0,85 (Rieder, 2012).

Si nous ajoutons un lien de la page C vers une page extérieure « Ext2 », nous obtenons une diminution du PageRank de l'ensemble des pages :

Figure 36. Calcul du PageRank (3)



Un éditeur a intérêt à pointer vers ses propres pages et à recevoir des liens provenant de pages extérieures. En revanche, un lien vers l'extérieur dilue le PageRank de l'ensemble des pages. Le PageRank incite donc les éditeurs à ne pas pointer vers d'autres sites, ce dont pourtant Google a besoin pour que les crawlers et le PageRank puissent fonctionner. La formule porte ainsi en elle sa propre négation. Un psychologue dirait qu'elle est suicidaire. L'énoncé PageRank performe en effet ses propres conditions d'énonciation, et nuit à leur félicité. En disant *à la fois* « les liens entrants font qu'un contenu est pertinent » et « les liens sortants diminuent la pertinence de ce contenu », la formule mathématique incite l'éditeur à maximiser le nombre de liens entrants et à minimiser le nombre de liens sortants.

Si aucun éditeur n'en connaissait l'existence, PageRank pourrait fonctionner sans que cela ait d'influence sur le nombre de liens. Mais dès lors que les éditeurs connaissent la formule, ils sont incités à ne pas faire de lien vers l'extérieur, alors que cela va contre l'intérêt (et le script) de ceux qui ont mis au point la formule. Le « pouvoir faire » des éditeurs menace alors la chaîne d'associations, étant donné précisément la manière dont Google leur a dit que cette chaîne fonctionnait. Et le « pouvoir de faire faire » de la firme ainsi que le caractère performatif de la formule

PageRank sont dans ce cas un frein à sa propre action, son propre projet, ses propres intérêts. Dès lors, Google cherchera à réduire son influence, c'est-à-dire à réduire la performativité du PageRank en produisant d'autres énoncés, notamment d'autres critères algorithmiques et un discours encourageant la production de liens sortants.

Si l'on estime, en revanche, que le moteur remplit une fonction que les éditeurs ne sont pas capables de remplir (ou ne veulent pas remplir), alors les éditeurs ont malgré tout intérêt à effectuer des liens vers l'extérieur, car si aucun d'eux ne le faisait, cela ne servirait plus à rien de ne pas en faire. La performativité du PageRank peut ainsi être atténuée par le fait que, dans le cas d'une performance complète, ceux qui auraient modifiés leurs cours d'action n'atteindraient pas les objectifs qu'ils poursuivaient. Considérés comme un ensemble homogène, les éditeurs ont donc intérêt à faire des liens vers l'extérieur. Le risque pour Google, c'est qu'un éditeur se désolidarise du groupe et agisse dans son seul intérêt. Il fera alors ce que ses homologues ne doivent pas faire pour que, justement, son action puisse être bénéfique. Son action n'aura ainsi l'effet escompté que si les autres agissent de manière inverse et si la performativité du PageRank est incomplète, c'est-à-dire que la performativité de son propre énoncé dépend du caractère unique, ou rare, de l'énoncé en question. Nous voyons très bien apparaître ici certains risques menaçant la chaîne d'associations dont la liste de liens hypertextes est le résultat. Cette « boucle infernale » révèle plusieurs éléments :

- à quel point les actions des uns et des autres sont interdépendantes, susceptibles de se renforcer comme de s'annuler ou se nuire ;
- à quel point les mécanismes qui placent Google en position dominante peuvent être fragiles ;
- de quelle manière la technique peut agir contre les intérêts de ses concepteurs, et comment elle peut être elle-même victime de son propre succès ;
- comment la performativité d'un énoncé peut nuire aux conditions de son énonciation ;
- enfin, pourquoi Google peut avoir intérêt à ne pas faire la lumière sur l'ensemble de ses critères algorithmiques comme il a fait, dès 1998, pour le PageRank. En effet, le fait que le PageRank soit connu de tous dans ses moindre détails, et notamment depuis le dépôt par Google d'un brevet protégeant sa formule, aboutit à des actions de la part de certains éditeurs contraires aux souhaits des concepteurs. Ainsi, le secret dont l'algorithme d'appariement fait l'objet vise à atténuer les effets négatifs de sa performativité.

Nous avons vu dans le chapitre 4 qu'un éditeur pouvait utiliser la balise « nofollow » pour pointer vers le site d'un autre éditeur sans y envoyer le crawler de Google, et donc ne pas redistribuer d'autorité PageRank. C'est ce que fait systématiquement l'encyclopédie participative Wikipédia. L'éditeur se saisit ainsi de la possibilité de faire un lien hypertexte et y mêle un second « pouvoir faire » lui permettant de ne pas être pénalisé. Il ne joue plus alors dans l'intérêt de Google, ni dans celui de l'éditeur vers lequel il pointe, mais dans son seul intérêt. La *traduction* (le dispositif sert mes intérêts en même temps qu'il sert ceux de Google et ceux des autres acteurs enrôlés) devient, au sens de Michel Callon (1986), *trahison* (le dispositif sert mes seuls intérêts).

Le PageRank peut donner lieu à plusieurs autres stratégies de la part des éditeurs. Par exemple, certains d'entre eux pourront définir une politique de partage de liens avec d'autres sites : stratégie théoriquement gagnant/gagnant, mais déconseillée par les porte-parole de Google qui menacent de sanctionner ceux qui la pratiquent. Nous comprenons également en quoi certains éléments échapperont toujours au contrôle de l'éditeur. Notamment, celui-ci ne pourra jamais obliger d'autres éditeurs à faire des liens vers son contenu. C'est pourquoi certains observateurs disent que « la chasse aux *backlinks* est la forme moderne de la "conquête du Graal" »²⁸⁶.

De son côté, Google agit afin de convaincre les éditeurs que les autres paramètres de l'algorithme compensent l'impact négatif des liens vers l'extérieur. La firme tente ainsi de contrebalancer la performativité de l'énoncé PageRank en y ajoutant d'autres énoncés discursifs ou techniques. Si nous filons notre métaphore psychanalytique, nous pouvons dire que les ingénieurs conçoivent alors des « psychotropes » constitués indifféremment de technique et de discours, permettant de soigner les pulsions suicidaires du dispositif.

En 2009, l'ingénieur Matt Cutts annonça que les « nofollow » seraient désormais comptabilisés par l'algorithme, c'est-à-dire qu'une page qui effectuerait trois liens dont deux en « nofollow » ne distribuerait malgré tout qu'un tiers de son autorité

²⁸⁶ <http://www.webmaster-hub.com/publication/imprimer17.html> (dernière visite le 22 août 2013)

PageRank via le seul lien comptabilisé. Cette décision visait clairement à enrayer les pratiques de « PageRank sculpting » et à « contre-performer » la formule. D'autre part, Google indiqua que les liens effectués vers leurs propres contenus par les éditeurs (liens internes) seraient moins valorisés que les liens provenant de sites extérieurs. Mais certains éditeurs créèrent alors des « sous-domaines » identifiés par le crawler comme des sites extérieurs, permettant de faire un lien interne valorisé comme un lien externe. (C'est notamment ce qu'a fait *L'Express* avec la « Coupole » dont nous parlerons ci-après (cf. 8.4.1)). Les éditeurs se sont ainsi saisis de leur « pouvoir faire » pour contourner le « pouvoir de faire faire » que la firme tâchait d'exercer à leur endroit. Par conséquent, Google a procédé en 2011 à un nouveau détour en annonçant que les liens depuis les sous-domaines seraient désormais considérés comme des liens internes²⁸⁷. Cette mesure a été perçue par certains spécialistes du domaine comme un « coup de pied dans la stratégie de référencement »²⁸⁸. Par ailleurs, les sous-domaines permettaient également aux éditeurs de contourner la limite maximum de deux résultats d'un même site affichés dans une même liste hypertexte, et, éventuellement, d'obtenir tous les liens affichés par Google sur la première page²⁸⁹. Cette fois, Google limitait leur « pouvoir faire » en décrétant que les sous-domaines, puisqu'ils étaient utilisés par les éditeurs comme tels, seraient dorénavant assimilés à des répertoires du domaine principal. Le dispositif était ainsi agrémenté d'autres énoncés, techniques et discursifs, dont la performativité visait à contrebalancer les effets négatifs de la performativité du PageRank. Les experts du SEO se sont alors demandés « si Google gérait algorithmiquement la différenciation des sous domaines de plateformes connues »²⁹⁰. Dans la pratique, nous verrons ci-après que les référenceurs continuent à procéder à des liens internes, car même s'ils sont moins efficaces que des liens externes, ils sont tout de même efficaces d'après eux, et ce sont les seuls liens sur lesquels ils exercent un contrôle leur permettant d'agir dans le but de maximiser l'apport de trafic. Nous retrouvons ici les ajustements et les détours chers à l'ANT, les *traductions*, c'est-à-

²⁸⁷ <http://googlewebmastercentral.blogspot.fr/2011/08/reorganizing-internal-vs-external.html> (dernière visite le 22 août 2013)

²⁸⁸ <http://www.arobasenet.com/2011/09/referencement-google-declassifie-les-sous-domaines/> (dernière visite le 22 août 2013)

²⁸⁹ <http://www.abondance.com/actualites/20071210-8612-google-sur-le-point-de-traiter-les-sous-domaines-comme-des-repertoires.html> (dernière visite le 22 août 2013)

²⁹⁰ <http://blog.axe-net.fr/google-et-les-sous-domaines/> (dernière visite le 22 août 2013)

dire les interprétations et les déplacements, par lesquelles la technique, le réseau, les acteurs-réseau, passent pour parvenir à leurs fins, et dont ils négocient directement ou indirectement à la fois les conditions, les modalités et le résultat.

Google stipule que « la quantité, la qualité et la pertinence des liens influencent [le] classement »²⁹¹. La firme indique également que les liens sortants vers des sites pertinents seront également valorisés, contrebalançant ainsi directement la performativité du PageRank en incitant les éditeurs à faire des liens vers l'extérieur. A ce sujet, Matt Cutts, porte-parole de la firme, explique : « Google a moins confiance en des sites qui font des liens vers des sites pleins de spams ou dont le voisinage est mal perçu ; nous encourageons les liens vers de bons sites. »²⁹² De nombreux experts en SEO le confirment : « le lien vers une source qualitative, au sens où elle est jugée qualitative par le moteur de recherche, peut valoriser le site. »²⁹³ En incitant les sites à faire des liens vers l'extérieur, et vers du contenu pertinent, Google représente ses propres intérêts et ceux de ses utilisateurs, car non seulement le défaut du PageRank sera potentiellement corrigé mais, en plus, les liens externes vers des sites pertinents seront encouragés. Cependant, nous verrons ci-après que les liens vers l'extérieur sont encore très peu utilisés par les sites de presse, préférant procéder à des liens vers leurs propres pages pour des raisons essentiellement liées à Google.

Certaines pratiques de « linking » sont fortement déconseillées par Google²⁹⁴ :

- « Achat ou vente de liens. Cela inclut l'achat de liens ou de messages contenant des liens, l'échange de biens ou de services contre des liens, ou encore l'envoi d'un produit "gratuit" en échange d'un commentaire incluant un lien.
- Échange de liens excessif ("Établissez un lien vers mon site et j'établirai un lien vers le vôtre")
- Liens vers des spammeurs Web ou des sites sans rapport dans le but de manipuler PageRank
- Conception de pages partenaires utilisées exclusivement à des fins de liaison transverse
- Utilisation de programmes ou services automatisés pour créer des liens vers votre site »

²⁹¹ <http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=66356> (dernière visite le 22 août 2013)

²⁹² <http://www.mattcutts.com/blog/pagerank-sculpting/> (dernière visite le 22 août 2013)

²⁹³ <http://www.renaud-joly.fr/index.php/post/2011-liens-sortants> (dernière visite le 22 août 2013)

²⁹⁴ <http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=66356> (dernière visite le 22 août 2013)

Google fait largement campagne contre l'achat et la vente de liens, essayant ainsi d'empêcher qu'un marché ne s'organise autour de la performativité du PageRank. Une image très connue par les experts du SEO présente Matt Cutts tenant un chaton dans les mains et annonçant : « A chaque fois que vous achèterez un lien, Matt Cutts tue un chaton. S'il vous plaît, pensez aux chatons. » (cf. ci-dessous). Cette tentative d'intimidation illustre la limite du « pouvoir sur » de Google, borné par le « pouvoir faire » des éditeurs qui peuvent vendre et acheter des liens dans la mesure où la loi ne le leur interdit pas. En revanche, s'ils achètent ou vendent trop de liens, ils prendront le risque que Google les sanctionne par un biais systématique, censurant de fait leurs documents, dans la mesure où la loi ne le leur interdit pas non plus (absence de loi Bichet). La communication mise en place par Google, liant soudain le sort de chatons à celui des liens hypertextes, permet de « charger » autrement l'énoncé, c'est-à-dire de le « déplacer » pour augmenter le « degré d'attachements » de ceux qui collaborent activement au programme d'action des concepteurs et ainsi de faire que l'énoncé « gagne en réalité » (Latour *et al.*, 1991) tout en diminuant la performativité de la seule formule PageRank.



Il est également indiqué par les porte-parole de Google que les « liens cachés » dans le contenu « en vue de manipuler le classement des résultats de recherche Google peuvent être considérés comme une technique trompeuse et constituer une infraction aux Consignes aux webmasters »²⁹⁵. Un éditeur qui effectuerait de telles pratiques s'exposerait à une sanction. Nous verrons pourtant que certains SIPI, et en particulier Lemonde.fr, cachent systématiquement certains liens (cf. 8.2.5).

²⁹⁵ <http://support.google.com/webmasters/bin/answer.py?hl=fr&answer=66353> (dernière visite le 22 août 2013)

Enfin, Google demande aux internautes et aux éditeurs de « dénoncer » les sites qui ne respecteraient pas ses préconisations en matière de politique de liens²⁹⁶. Un éditeur peut ainsi être identifié manuellement par un autre acteur (éditeur ou internaute) qui identifie la *trahison* et dénonce ce qui constitue à ses yeux une entrave au bon fonctionnement du dispositif auquel il a lui-même pris part. Comme elle l'a fait en Allemagne lors du lobbying contre le projet de loi « Lex Google » (cf. 5.1.4), la firme Google s'appuie ainsi sur l'association des utilisateurs à son action pour faire pression et exercer un « pouvoir de faire faire » sur les éditeurs, étant donné les dénonciations dont ils sont susceptibles de faire l'objet. Ce seront les internautes, alors, qui produiront un énoncé dont la performativité contrecarrera celle du PageRank, non désirée par Google.

8.2.2 Egocentrisme hypertexte

Les éditeurs de notre panel effectuent des liens principalement vers leurs propres contenus. En effet, la part des liens figurant dans le corps des 252 articles observés et pointant vers des pages du même site varie entre 86% pour Lefigaro.fr et 96% pour Lemonde.fr et Lepoint.fr. Parmi les liens sortants, la part pointant vers des sites concurrents varie quant à elle entre 1,8% pour Lemonde.fr et 9,6% pour Lefigaro.fr. Au total, 94% des 2117 liens observés pointent vers un contenu du site sur lesquels ils se trouvent.

²⁹⁶ Il suffit de remplir un formulaire en ligne : <https://www.google.com/webmasters/tools/paidlinks> (dernière visite le 22 août 2013)

Tableau 11. Liens entrants et sortants

	Nombre de liens dans le corps de l'article	Nombre liens vers une page du même site	Nombre de liens vers un autre site	Nombre de liens vers un autre site d'information
Lexpress.fr	315	284 (90%)	31 (10%)	21 (6,6%)
Lefigaro.fr	155	133 (86%)	22 (14%)	15 (9,6%)
Lemonde.fr	1258	1202 (96%)	56 (4%)	23 (1,8%)
Nouvelobs.com	128	114 (89%)	14 (11%)	5 (3,9%)
Leparisien.fr	188	177 (94%)	11 (6%)	11 (6%)
Lepoint.fr	73	70 (96%)	3 (4%)	3 (4%)
TOTAL	2117	1980 (94%)	137 (6%)	78 (3,7%)

Tableau 12. Moyennes du nombre de liens entrants et sortants

	Moyenne du nombre de liens par article	Moyenne du nombre de liens internes p.a.	Moyenne du nombre de liens sortant p.a.	Moyenne du nombre de liens vers des sites d'information p.a.
Lexpress.fr	7,5	6,8	0,69	0,50
Lefigaro.fr	3,7	3,2	0,5	0,36
Lemonde.fr	30	29	1,1	0,55
Nouvelobs.com	3,05	2,7	0,33	0,12
Leparisien.fr	4,5	4,2	0,5	0,26
Lepoint.fr	1,7	1,6	0,071	0,071

Selon notre observation, le site du *Monde* est à la fois celui qui fait le plus de liens et celui qui a le plus de liens internes, étant donné la « stratégie du conjugueur » (cf. 8.2.5), mais aussi celui qui effectue le plus de liens vers des sites d'information politique et générale concurrents du sien (23 liens). Hormis le cas particulier du *Monde*, Lexpress.fr est le site qui pointe le plus souvent vers l'extérieur (31 liens), et ce même si le nombre de liens sortants pour chaque article est faible, avec en moyenne 6,8 liens par article dont 0,69 vers l'extérieur et 0,5 vers des sites

concurrents. Le site du *Point*, quant à lui, est de loin celui qui fait le moins de liens, à la fois vers ses propres contenus (1,6 lien par article en moyenne) et vers l'extérieur (0,071 lien sortant par article en moyenne). Le référenceur du *Point* nous a expliqué qu'il avait demandé aux journalistes de privilégier les liens internes, notamment vers les archives, afin de valoriser le référencement de ces dernières sur Google Search. Cependant, il semble que les journalistes fassent malgré tout peu de liens, y compris des liens internes, du moins dans le cas de notre échantillon.

Le site du *Figaro*, quant à lui, cite assez peu les sites extérieurs, avec 0,5 lien sortant en moyenne par article, pour 3,7 liens par article, et ce même si la proportion de liens sortants sur l'ensemble de l'échantillon est de 9,6 %. Là encore, il y a un décalage entre le discours du référenceur et les usages des journalistes. Cette fois, le référenceur encourage à faire des liens vers l'extérieur, et ce même si cela doit nuire au fonctionnement optimal du PageRank. Dans notre entretien, le référenceur du *Figaro* a mis en avant le fait qu'une stratégie de liens sortants pourrait ensuite être récompensée par les auteurs des contenus cités, qui eux-mêmes continueraient de citer le SIPI. En revanche, selon lui, un repli sur soi peut aboutir à une sanction, notamment de la part des communautés de blogueurs, qui ne citeront plus le SIPI, ce qui risque d'altérer l'autorité PageRank.

« A partir du moment où les liens sortants sont pertinents, on a aucun intérêt à ne pas le faire. C'est justement le principe de l'internet : faire des liens, lier des documents entre eux. Quelqu'un comme Le Monde, qui se met à l'écart parce qu'il ne fait pas de lien, il ne faut pas s'étonner après si certains blogueurs vont avoir tendance à moins citer le monde.fr parce que Le Monde ne fait jamais de liens vers eux. »

Lefigaro.fr — Référenceur — Michael Vuillaume

Il semble ainsi que certains référenceurs encouragent la production de liens vers l'extérieur, y compris des liens vers les concurrents, mais que journalistes pointent plus volontiers vers les contenus de leurs propres sites. Ainsi, pour le référenceur du *Parisien*, comme pour le référenceur du *Figaro*, le faible nombre de liens sortants serait lié à une réticence de la part des sites adossés à des médias traditionnels, qui ne « jouent pas le jeu du web », et qui espèrent ainsi augmenter le PageRank alors que cela pourrait au contraire finir par le diminuer.

« Les journalistes font ce qu'ils veulent. Moi, je n'ai pas d'inconvénient pour qu'on cite un concurrent. On cite Le Monde, très bien, faisons un lien vers Le Monde. Mais je sens que les journalistes ont beaucoup de mal à faire ça. Ils disent : "Je vais pas mettre de liens vers mes concurrents, parce que ça va augmenter son trafic et ça va augmenter son PageRank en plus". Alors que pour moi le web, c'est des liens. Il y a des sites plus petits qui mettent des liens. Mais les grands médias ont du mal. Ils ne le font pas, et ils croient qu'ils ont raison de ne pas le faire. C'est problématique ça. »

Leparisien.fr — Référenceur — Guillaume Giraudet

Malgré les détours effectués par Google dans le but d'encourager la production de liens sortants, les techniques de *PageRank Sculpting* sont encore largement utilisées, et ce malgré ce que nous ont dit les référenceurs du *Figaro* et du *Parisien*. La performativité de la formule demeure. Le dilemme est tranché pour le moment du côté du lien interne plutôt que du lien sortant, comme cela a déjà été observé plusieurs fois (Tsui, 2008 ; De Maeyer, 2013). Sur les six sites observés, la présence de nombreux liens internes est due à une automatisation de la production des liens vers des pages d'atterrissage, effectuée pour maximiser l'apport de trafic de Google (cf. 8.2.3). Ainsi, la communication de Google à propos de la diminution de l'autorité distribuée par le maillage interne n'a pas réussi à susciter le changement espéré. A l'inverse de ses homologues du *Parisien* et du *Figaro*, le référenceur de *Lexpress.fr* n'encourage pas la production de liens vers l'extérieur par les journalistes, et notamment vers des sites concurrents. Selon lui, cela avantage les concurrents, qui eux-mêmes ne feront pas de liens en retour. Il y a ici une mise en tension entre les pratiques des journalistes et les recommandations du référenceur, c'est-à-dire entre la volonté pour ce dernier de maximiser sa visibilité et de ne pas aider les concurrents à maximiser la leur, et la volonté de certains journalistes d'utiliser les liens pour citer leur source. Un exemple nous a été donné avec un article concernant un incendie à Pantin, dont la source était le site du *Parisien* :

« D'une part, tu as ton maillage interne qu'on appelle PR sculpting, et les liens qui viennent de l'extérieur. Ceux là, ils vont te permettre d'avoir de la puissance sur tes pages à toi. Ce PageRank, ton maillage va te permettre de le rediffuser sur les pages les plus importantes. C'est-à-dire qu'une page qui reçoit énormément de liens, il va falloir que depuis elle, tu renvoies vers des pages dont tu estimes qu'elles sont importantes pour le SEO, toujours vers toi-même : vers des pages de chez toi.
[...]

Il y a encore un problème que je n'ai pas encore réussi à régler. Je te prends un exemple : ce matin, il y a eu un incendie à Pantin. On fait un papier rapide qui remonte dans Google News. « Selon Le Parisien... » et on met un lien vers Le Parisien. Tu vois, ce lien, selon moi, il ne sert à rien niveau SEO, parce que tu donnes de la puissance au papier d'un concurrent. Ça veut dire que tu estimes que tu n'as pas l'info, et que t'envoies les personnes ailleurs. Je veux bien le comprendre sur une info qui est un peu froide, mais sur une info qui est chaude ça ne sert à rien de faire des liens vers les concurrents. Et les journalistes en font trop je trouve. »

Lexpress.fr — Référenceur — Renaud Perrin

Le référenceur considère ici la pratique journalistique comme « un problème qu'il faut régler ». Il porte un jugement sur le travail des journalistes, qui traitent une actualité déjà traitée ailleurs, à propos de laquelle ils n'ont rien, ou pas grand chose, à ajouter, et dont l'action finit par contribuer à augmenter le PageRank d'un concurrent tout en nourrissant le phénomène de « circulation circulaire » (Rebillard, 2006). De tels liens sont ajoutés manuellement aux articles par la rédaction, et le référenceur n'a pas de contrôle direct à leur endroit. Nous retrouvons l'idée avancée dans le chapitre 5 selon laquelle la stratégie est co-produite par les acteurs, dont les intérêts, ou les projets, les valeurs, peuvent différer, ce qui peut aboutir à des actions qui ne sont pas considérées comme étant optimales par l'ensemble des acteurs impliqués dans (et concernés par) le réseau de production et de diffusion. En outre, nous voyons comment la stratégie de SEO peut conduire des employés dont la fonction pourrait apparaître comme étant essentiellement technique à être concernés par la partie éditoriale du travail, c'est-à-dire par la manière dont l'actualité est traitée et dont les sources sont citées. En fait, il y a une mise en tension à l'intérieur de l'organisation entre deux manières de concevoir la fonction du lien hypertexte : l'une journalistique, l'autre liée au SEO. L'enjeu, dans l'idéal, serait de réussir à ce que chaque lien puisse remplir *à la fois* l'une et l'autre de ces fonctions.

Il a également été intéressant de remarquer dans notre échantillon que la balise méta-tag <nofollow> n'était jamais utilisée pour les liens du corps de l'article, y compris quand un concurrent était cité. Cela pourrait pourtant permettre de ne pas distribuer l'autorité PageRank. Les journalistes citeraient alors leurs sources sans ne jamais envoyer les crawlers sur des sites concurrents (mais en diluant quand même l'autorité redistribuée à leurs propres pages). Par ailleurs, la balise <standout>,

permettant de désigner des travaux originaux à forte valeur ajoutée (cf. 4.5.2.B), n'est elle non plus jamais incorporée aux liens hypertextes. Les éditeurs ne se saisissent pas de ce « pouvoir de faire » qui pourrait pourtant augmenter le poids aussi bien des liens internes que des liens sortants. Les journalistes interrogés ne connaissaient pas ces balises. Il semblerait ainsi qu'ils ne choisissent pas de ne pas les utiliser, mais plutôt qu'ils ne savent pas qu'il est possible pour eux de les utiliser.

8.2.3 Pages d'atterrissage

Certaines pages nous ont été présentées comme des « pages d'atterrissage ». Celles-ci ne sont en général pas accessibles depuis la page d'accueil des sites de presse. Elle sont construites autour d'un mot-clé ou d'une thématique — un nom de lieu (Florange), un nom de personnalité (François Hollande), un nom d'événement (Roland Garros) — et conçues à la fois pour répertorier les archives à propos d'un même sujet et pour positionner la page en question sur Google Search. On y trouve répertoriés sous forme de liste de liens hypertexte (en général titre + accroche + image) l'ensemble des articles concernant un même sujet, classés dans l'ordre *ante-*chronologique. Par exemple, ci-après, une capture d'écran partielle de la page d'atterrissage « Florange » du site Lefigaro.fr.

Figure 37. Capture d'écran : Lefigaro.fr (5 juin 2013)



(Source : <http://plus.lefigaro.fr/tag/florange>)

Il s'agit de faire « atterrir » les crawlers sur des pages dont le contenu est optimisé et régulièrement mis à jour. Dans la pratique, dès qu'un journaliste écrit un article où figure le mot-clé d'une page d'atterrissage, un lien est généré automatiquement sur ce mot-clé vers la page d'atterrissage. Du point de vue du maillage interne, cela permet de déporter systématiquement le PageRank vers la même page du site, pour un même mot-clé, afin d'optimiser le référencement de celle-ci sur Google Search. Les liens vers ces pages sont donc destinés *a priori* aux

crawleurs plutôt qu'aux visiteurs du site. Ce sont des moyens de structurer le site et de gérer le déplacement des crawleurs, davantage que des éléments de navigation au sens où on l'entend en général quand on parle de lien hypertexte. Dans notre échantillon, nous avons :

Tableau 13. Liens vers pages d'atterrissage

	Nombre de liens dans le corps de l'article	Nombre de liens vers pages d'atterrissage (hors conjugueur)	Nombre de liens vers conjugueur
Lexpress.fr	315	131 (42%)	0
Lefigaro.fr	155	54 (35%)	0
Lemonde.fr	1258	441 (35%)	700
Nouvelobs.com	128	84 (66%)	0
Leparisien.fr	188	74 (40%)	0
Lepoint.fr	73	54 (74%)	0

Tableau 14. Moyennes du nombre de liens vers pages d'atterrissage

	Moyenne du nombre de liens par article	Moyenne du nombre de liens vers pages d'atterrissage par article (sans conjugueur)	Moyenne du nombre de liens vers pages d'atterrissage par article (avec conjugueur)
Lexpress.fr	7,5	3,12	3,12
Lefigaro.fr	3,7	1,29	1,29
Lemonde.fr	30	10,50	27,17
Nouvelobs.com	3,05	2,00	2,00
Leparisien.fr	4,5	1,76	1,76
Lepoint.fr	1,7	1,29	1,29

Les sites de *L'Express*, du *Figaro* et du *Parisien* ont tous trois entre 35% et 42% de leurs liens pointant vers des pages d'atterrissage. Par ailleurs, *Lepoint.fr* et *Nouvelobs.com* ont des chiffres significativement plus élevés avec 74% et 66% des liens dans le corps des articles effectués vers des pages d'atterrissage.

Pour ce qui est des moyennes, celle du *Monde* (hors conjugueur) est de 10,5 liens par article pointant vers des pages d'atterrissage, ce qui est sensiblement plus élevé que les autres. *Le Monde* pratique ainsi une stratégie de maillage interne intensive sur laquelle nous reviendrons lorsque nous présenterons la « stratégie du conjugueur » (cf. 8.2.5). Quant aux cinq autres sites, le nombre de liens vers des pages d'atterrissages y varie de 1,29 (lepoint.fr et lefigaro.fr) à 3,12 par article (Lexpress.fr). Ainsi, chacun des sites effectue une stratégie de maillage interne visant expressément à optimiser le référencement sur Google, du moins d'après les porte-parole interrogés (Lexpress.fr, Lefigaro.fr, Leparisien.fr et Lepoint.fr). Puisque ces pages ne sont pas destinées aux lecteurs qui accèdent directement au site, et ce même si les liens dans le corps de l'article peuvent les y conduire, il s'agit d'une couche éditoriale supportée par une technique d'automatisation et destinée aux usagers de Google, pour qui les pages d'atterrissage jouent le rôle de « rubriques thématiques ».

« On fait des liens vers des pages d'atterrissage. On a un sous domaine qui s'appelle actualités.leparisien.fr dédié à ces pages d'atterrissages. Il liste les requêtes, par exemple « nicolas sarkozy ». Ensuite, quand le journaliste écrit son article, un lien se fait automatiquement sur le nom « nicolas sarkozy ». Les gens qui arrivent par le moteur en tapant ça arriveront sur cette page. En revanche, les internautes qui passent par la home page du site, en accès direct, eux, ne la verront jamais. Les pages d'atterrissages sont inaccessibles pour eux. Elles servent uniquement à ceux qui viennent depuis Google. On fait ça sur beaucoup d'itérations. Dès qu'une affaire se crée, par exemple l'affaire Mera, on crée une page Mera. C'est un algorithme sémantique qui tourne derrière. Dès qu'un article est consacré à Mera, il sera ajouté à la landing page. C'est un peu comme une catégorie, du coup ça a quand même un véritable intérêt au niveau de l'édition. »

Leparisien.fr — Réferenceur — Guillaume Giraudet

Comme le référenceur du *Parisien*, le référenceur du *Figaro* voit dans les pages d'atterrissage un intérêt éditorial, parce que des catégories sont créées par sujet, et non plus simplement par rubrique et par date :

« On est sur un type de page qui a une signification éditoriale, parce que du coup l'utilisateur de Google il va arriver sur cette page, et il va avoir un contenu dédié à cette thématique. Donc ça peut faire page d'atterrissage, spéciale pour le SEO, et en même temps c'est le cœur du métier du Figaro de donner des informations chaudes mais aussi froides sur un même sujet. »

Lefigaro.fr — Réferenceur — Michael Vuillaume

Nous voyons ici que des critères de qualité journalistique et la stratégie SEO sont conciliés autour d'un même dispositif de création de pages d'atterrissages. Par ailleurs, ces pages permettent de contourner le problème du moteur de recherche interne. En effet, si des requêtes peuvent être effectuées sur des sujets à l'intérieur du site de presse, sur un moteur de recherche interne, les pages de résultats de ce moteur ne doivent pas, elles, être indexées par Google, du moins selon les recommandations des porte-parole de la firme. Aussi la création de pages d'atterrissage permet-elle de prévoir de répondre à une requête en amont de cette requête. Ce n'est pas l'énoncé qui performe ici, mais la *possibilité* d'un énoncé.

« Google recommande de ne pas indexer les moteurs de recherche internes, parce que, tu comprends, il ne va pas indexer tes résultats dans ses résultats. Donc on met des <noindex> sur les moteurs de recherche internes. Mais avec les landing page, on peut éviter ce problème et quand même fournir une page de résultats dans les résultats de Google. En fait, le truc, c'est qu'on prépare une page pour les requêtes avant que celles-ci arrivent, et du coup on n'a pas besoin de la générer au moment de la requête : elle existe déjà, elle est là, toute prête, et Google n'a plus qu'à pointer dessus. »

Leparisien.fr — Référencier — Guillaume Giraudet

En plus des pages d'atterrissage par sujet, des pages généralement nommées « Toute l'actualité » sont créées au niveau $n+1$ des SIPI, dans lesquelles on retrouve une liste de liens (titres, accroches, parfois photos) vers les articles publiés lors des dernières 48 heures. Ces pages sont elles aussi inaccessibles, ou en tout cas difficiles d'accès, depuis le site de presse, destinées principalement aux crawlers de Google. Les liens vers les articles sont entassés les uns sur les autres dans l'ordre *ante*-chronologique, sans hiérarchie, sans éditorialisation manuelle, automatiquement, cela pour permettre au crawler de rapidement trouver des articles récents qui ne figurent plus sur la page d'accueil.

« D'expérience, en général, pour le crawling, il peut y avoir un problème de profondeur. Plus c'est loin dans la profondeur depuis la page d'accueil, au niveau $n+2$ ou $n+3$, plus ça va être difficile pour Google d'aller chercher la page. En plus, si je fais une mise à jour sur l'article, il va avoir du mal à venir recrawler la page déjà crawlée. Donc pour l'actualité, il faut une profondeur peu profonde, c'est pour ça qu'on voit souvent des sites d'actualités avec une grande Une, et tout à profondeur $+1$ ou $+2$. Naturellement, c'est lié à la chronologie, et c'est tout à fait normal que sur un site d'actualité, les articles publiés à $J+2$, $J+3$ soient plus profonds. Alors

on utilise des moyens pour réduire cette profondeur et permettre des réindexations de la page par Googlebot. Notamment, vous avez sur la majorité des sites de presse une page qui s'appelle « Toute l'actualité », accessible sur le menu du haut où vous aurez tous les articles publiés depuis les deux derniers jours. [...] Google considère que la page d'accueil c'est niveau zéro. Puis il va trouver un lien vers la page « Toute l'actualité », disons niveau $n+1$, et ensuite il va trouver en niveau $n+2$ tous les articles publiés récemment. Il ne faut pas dépasser $n+2$, qui est une profondeur tout à fait acceptable. Il faut éviter d'aller plus loin ou bien vous serez pénalisés. Les rubriques « Toute l'Actu », c'est donc surtout pour le GoogleBot, parce que les gens passent peu par là, voire jamais. »

Reseoneo — Consultant Senior — Julien Crenn

Le référenceur préconise ici à l'éditeur de presse de se saisir d'un « pouvoir faire » qui va venir s'ajouter aux « pouvoirs de faire » accordés par Google, et en particulier à l'utilisation du *Sitemap*, pour aplanir l'arborescence de navigation grâce à une page d'atterrissage destinée, non plus comme les précédentes à recevoir les utilisateurs de Google, mais cette fois à recevoir les crawlers et à leur indiquer les directions à suivre pour prendre connaissance des articles publiés lors des deux ou trois derniers jours. Il s'agit d'un nouveau détour, à la fois technique et éditorial, automatisé, effectué par les éditeurs pour que les crawlers ne fassent pas, eux, de détours, et que la fréquence de leur passage sur la zone « actualités chaudes » soit maximisée.

8.2.4 Liens invisibles

Dans notre échantillon, trois sites cachent des liens (Lexpress.fr, Lemonde.fr et le Nouvelobs.com). Autrement dit, sur ces trois sites, certains liens hypertextes, qui existent pourtant bel et bien, n'apparaissent pas comme tels aux yeux du visiteur qui ne voit pas qu'il peut cliquer sur ces mots en particulier et se rendre sur une autre page par leur biais. Le visiteur ne pourra voir ces liens qu'en se rendant dans le code HTML de la page, ou bien en passant doucement le curseur de sa souris sur les mots de l'article afin de voir lesquels donnent lieu à un clic possible, vers une autre page. La fonction de monstration du lien lui est alors ôtée, et le lien n'est plus un outil navigationnel mais seulement un outil SEO adressé au crawler — et ce même si Google menace de punir de telles pratiques.

Tableau 15. Liens visibles et invisibles

	Nombre de liens invisibles	Ratio liens invisibles / liens visibles
Lexpress.fr	128	0,7
Lefigaro.fr	0	0,0
Lemonde.fr	1149	10,5
Nouvelobs.com	82	1,8
Leparisien.fr	0	0,0
Lepoint.fr	0	0,0

Les ratios effectués révèlent que Lexpress.fr cache moins de liens qu'il n'en montre, mais qu'en revanche le Nouvelobs.com, pour un lien visible, effectue 1,8 lien invisible, tandis que Lemonde.fr, étant donné un maillage « serré », pour un lien montré à l'internaute cache, dans le corps de l'article, pas moins de 10,5 liens. Les liens cachés par les trois sites sont systématiquement des liens conduisant à des pages d'atterrissage (ou au conjugueur dans le cas du *Monde*). Ainsi, il est évident que ces liens ne sont pas destinées aux acteurs déjà présents sur le site, puisqu'ils ne leur sont pas montrés, mais bien aux crawleurs, à l'indexeur et à l'algorithme de Google, ainsi qu'aux utilisateurs de Google Search, dans le but de maximiser le PageRank et le trafic infomédié, et ce en dépit des propos des porte-parole de Google.

Quand ils sont invisibles, les liens sont renseignés comme suit (ici l'exemple d'un article de lepress.fr du 1^{er} octobre 2012 : *Match truqué de handball: les joueurs vont reconnaître avoir parié*, renvoyant vers la page d'atterrissage « Eric Dupond Moretti », où nous avons mis en gras la balise permettant de cacher le lien) :

```
<a href="/infos/soc/rtl.html" target="_self"
class="none_underline_link">RTL</a></em> lundi matin <a
href="/infos/pers/eric-dupond-moretti.html" target="_self"
class="none_underline_link">Eric Dupond-Moretti</a>,
```

Dès lors que la balise « none_underline_link » est implémentée, le lien n'est plus destiné à une quelconque interactivité. Il n'est plus navigationnel, mais promotionnel, adressé directement au dispositif d'infomédiation dans le but que ce dernier traduise le maillage dans le sens des intérêts de l'éditeur qui le pratique. Si nous reprenons l'image du topographe, du territoire et de la carte, chère à l'ANT,

nous pouvons imaginer une situation où un habitant d'une zone cartographiée construirait une nouvelle route qui ne serait pas destinée aux autres habitants, car son existence leur serait dissimulée, mais au seul topographe, dans le but que celui-ci trouve facilement la maison de l'habitant en question et finisse par lui donner une place de choix, bien visible, bien indiquée, sur la carte qu'il produira. C'est un exemple parfait de la manière dont l'énoncé peut performer la réalité qu'il désigne, et de la manière dont l'énoncé et la réalité sont finalement liés en un même tout sans cesse renégocié, mobilisé, transformé, médié par les actions conjointes de ceux qui les produisent, de ceux qui les utilisent et du champ d'action induit par la technique.

Les référenceurs des sites qui, eux, ne cachent pas les liens vers les pages d'atterrissage, se prévalent d'une « éthique » qui consisterait à ne rien cacher au lecteur. La pratique qui consiste à cacher des liens est alors désignée comme « *black hat* ». Il semble ainsi que ceux qui ne cachent pas leurs liens aient rejoint Google sur ce point *concrètement*, en ne faisant pas de liens invisibles, et aussi *idéalement* en désignant cette pratique comme contraire à leur propre éthique.

« Moi j'essaye d'avoir une éthique. Je pourrais faire des trucs sales, j'en ai fait avec d'autres sites, et via mon activité d'auto-entrepreneur. Le SEO, c'est toujours un peu du black hat, du « grey hat » si tu veux. Mais au Parisien, j'ai une éthique, qui est de dire qu'un lien ne sera jamais invisible. Lemonde.fr a choisi de cacher des liens, mais ça, moi, je trouve que ce n'est pas éthique. Voilà pourquoi je ne fais pas, je ne ferai jamais, de liens invisibles. »

Leparisien.fr — Référenceur — Guillaume Giraudet

D'autres référenceurs, même s'ils considèrent que la technique qui consiste à cacher des liens est « sale », ne voient pas ce qu'il y aurait ici comme manquement à une quelconque éthique. C'est le cas par exemple du référenceur de *Lexpress.fr* :

« Le Monde, ils font des trucs très sales, mais moi je trouve ça génial : ils planquent par exemple des liens sur toutes leurs pages. C'est assez hallucinant. Tu pourrais dire « ohlala c'est pas bien », mais moi je trouve ça génial, purement génial, parce que le mec il a vu que ça passait, la technique l'a fait, l'internaute ça ne le dérange pas, les journalistes ne sont pas concernés, ça ne change rien à leur boulot, et voilà... C'est ça le SEO. C'est de la bricole. C'est un jeu. »

Lexpress.fr — Référenceur — Renaud Perrin

Sur Lepoint.fr, les liens vers les pages d'atterrissage ne sont pas cachés, même s'ils sont clairement destinés à améliorer le référencement sur Google Search autour de certains mots-clés, mais il se trouve que les internautes cliquent massivement sur ces liens, ce qui n'était pas prévu au départ par ceux qui ont mis en place le dispositif de *linking* automatisé.

« Pour les pages d'atterrissage, on a mis en œuvre une technique qui permet, dès que tu mets Nicolas Sarkozy, que ce soit souligné et que ça renvoie vers une page où tu auras tous les articles à propos de Nicolas Sarkozy. On a dit qu'on le faisait pour l'internaute, mais en vrai c'était pour le référencement, seulement pour le référencement. Mais il y a eu une chose amusante avec ça, qu'on n'avait pas prévue au départ. En fait, sur ces pages d'atterrissage, on a 75% de visiteurs qui viennent des pages article de notre site, et très peu de Google, donc finalement c'est utile aux internautes alors qu'on l'avait fait pour Google. Et ça augmente le taux de rebond. C'est tout bénéf'. En fait, c'est une manière détournée qui ne correspondait pas à notre objectif initial. On l'avait mis en place pour le positionnement PageRank, et finalement ce sont les articles qui pointaient vers les pages thématiques qui ont été valorisées sur Google plutôt que les pages thématiques elles mêmes, ce que je trouve très bien, parce que je préfère que ce soit le dernier article sur Sarkozy qui sorte sur Google plutôt que la landing page. »

Lepoint.fr — Réferenceur — Laure Aubry

Nous voyons très bien dans la citation précédente comment un dispositif dont l'intérêt éditorial n'était pas perçu au départ du point de vue journalistique, mis en place pour maximiser la visibilité sur Google, a finalement eu un intérêt pour les lecteurs, qui ont utilisé les liens, visibles ceux-ci, pour naviguer, et qui ont consulté les pages d'atterrissages pour accéder à l'historique du traitement d'un sujet en particulier. Cette fois, le détour est venu des internautes eux-mêmes, qui se sont saisis d'une possibilité mise à leur disposition, certes, mais destinée normalement au seul crawler de Google. La performativité a ainsi subi plusieurs déplacements inattendus, qui ont trouvé une justification aussi bien en termes de visibilité sur le dispositif d'infomédiation qu'en termes d'utilité perçue par les lecteurs déjà présents sur le site.

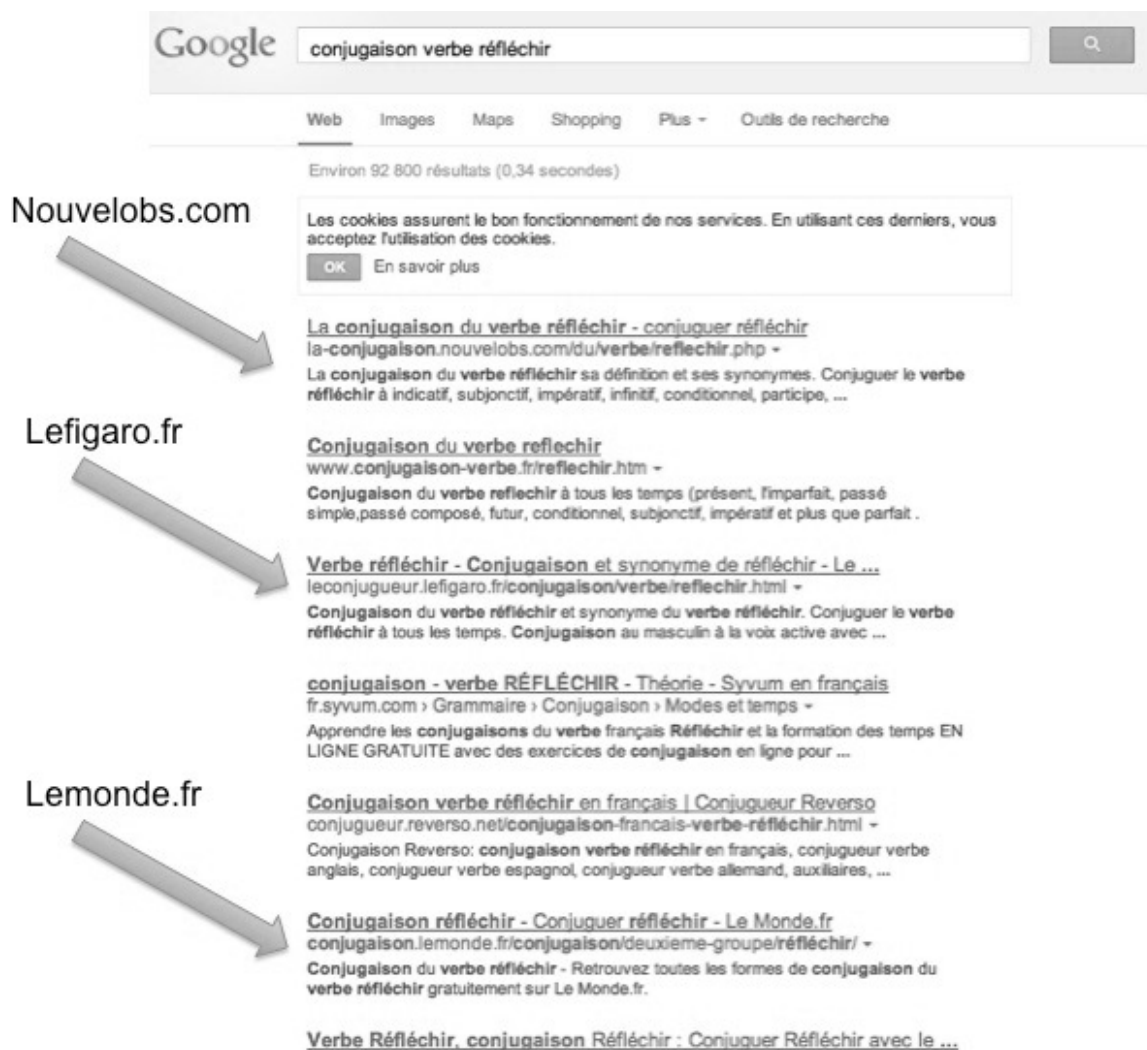
8.2.5 La stratégie du conjugueur

Parmi les entreprises de notre échantillon, trois possèdent un « conjugueur », c'est-à-dire un sous-domaine sur lequel figurent les tables de conjugaison de la langue française.

- Lefigaro.fr : <http://leconjugueur.lefigaro.fr>
- Lemonde.fr : <http://conjugaison.lemonde.fr/conjugaison/>
- Nouvelobs.com : <http://la-conjugaison.nouvelobs.com>

Les conjugueurs permettent d'afficher des publicités ailleurs que sur les seules pages traitant de l'actualité. Ils sont ainsi intégrés à une stratégie de diversification visant à développer certaines synergies autour de la marque. Les pages du conjugueur, qui ne sont pas lourdes à crawler, sont intégrées au sous-domaine. A chaque fois qu'elles seront affichées, des publicités du réseau de partenaires de l'entreprise de presse s'afficheront également, générant des revenus au CPM. La force de la marque est ainsi utilisée pour facturer des publicités figurant pourtant auprès d'un contenu à faible valeur ajoutée, mais malgré tout pertinent. Sur les pages du conjugueur, plus de 95% des visiteurs proviennent de Google Search selon les approximations qui nous ont été données par les acteurs interrogés. En effet, certains internautes utilisent Google pour la vérification orthographique, et peuvent alors être conduits sur un des trois conjugueurs du *Monde*, du *Figaro* ou du *Nouvel Observateur*. Par exemple, si nous formulons sur Google Search la requête « conjugaison verbe réfléchir » (6 juin 2013), nous obtenons :

Figure 38. Capture d'écran : Conjugueurs
 (Requête « conjugaison verbe réfléchir » - 06/06/13)



La stratégie du conjugueur consiste à attirer un trafic vers un contenu qui n'est pas lié à l'actualité et à le facturer au même réseau d'annonceurs que celui qui a été tissé autour de la production journalistique. Cette stratégie est d'autant plus intéressante qu'elle ne demande pas d'importants moyens financiers ou humains, et qu'elle maximise les effets de réseaux croisés provoqués par la présence des internautes sur le versant des annonceurs. Une fois le conjugueur intégré au site, l'éditeur pourra monétiser des pages vues sans avoir à en remettre à jour le contenu ou la structure.

Etant donné que presque tout le trafic des conjugueurs provient de Google et qu'il s'agit d'un secteur concurrentiel, les propriétaires d'un conjugueur ont intérêt à optimiser leur référencement. Pour ce qui est des trois sites de notre panel, tandis que Lefigaro.fr se contente du conjugueur tel quel, *Nouvelobs.com* et *Lemonde.fr* utilisent leur autorité PageRank pour faire monter leurs conjugueurs dans les classements. Ils pratiquent cette stratégie malgré les dispositions que Google dit avoir prises pour ne pas prendre en compte le maillage interne et considérer les sous-domaines comme des sous-parties d'un même domaine.

Le Nouvel Observateur n'effectue pas les liens vers son conjugueur dans le corps de l'article mais en bas de chaque page dans un encart « Verbes thématiques » répertoriant certains verbes à l'infinitif comme s'il s'agissait de mots-clés, dont chacun pointe vers la page qui lui est dédiée sur le conjugueur. Cela se présente ainsi :

Figure 39. Capture d'écran : Verbes thématiques (Nouvelobs.com)



Le Monde a sans aucun doute la stratégie de référencement la plus intensive en matière de conjugueur. En effet, contrairement au *Nouvel Observateur*, *Lemonde.fr* intègre les liens vers le conjugueur dans le corps de l'article sous forme de liens invisibles. En fait, chaque verbe mentionné à l'infinitif dans un article du site *Lemonde.fr* est un lien invisible vers le conjugueur. Dans notre échantillon, sur les 42 articles observés sur *Lemonde.fr*, il y avait au total 1258 liens, parmi lesquels 700 pointaient vers le conjugueur, soit 56%. En cumulant cette part à la part représentée par les liens vers des pages d'atterrissage, nous avons 1141 liens effectués automatiquement vers des pages n'étant pas des pages articles, soit 91% du total des liens. Cette stratégie est uniquement imputable au *PageRank Sculpting*. Elle vise à profiter de l'autorité très forte des articles produits et mis en ligne par les journalistes pour augmenter, par synergie, la visibilité et le trafic du conjugueur. Ainsi, *Le Monde* ignore les mises en garde de Google et les propos des porte-parole

d'après lesquels les stratégies de maillage interne ne seraient pas opérantes. C'est à cause de cela que lorsque les référenceurs travaillant pour des sites de presse concurrents du *Monde* évoquent des pratiques « *black hat* », « non éthique » ou « sales », c'est toujours les pratiques du *Monde* qui sont mises en avant.

Figure 40. Code HTML : Lemonde.fr (17/10/12)

```
<p>J'ai voulu, depuis mon élection, que l'Europe se donne
comme priorité la croissance sans <a target="_blank"
onclick="return false;" class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/troisieme-
groupe/remettre">remettre</a> en question le sérieux
budgétaire, rendu indispensable par la crise des dettes
souveraines. Si nous ne donnons pas un nouveau souffle à
l'économie européenne, les mesures de discipline ne
pourront <a target="_blank" onclick="return false;"
class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/premier-
groupe/trouver">trouver</a> de traduction effective.</p>
<p>Le retour de la croissance suppose de <a target="_blank"
onclick="return false;" class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/premier-
groupe/mobiliser">mobiliser</a> des financements à
l'échelle de l'Europe, c'est le pacte que nous avons adopté
en juin, mais aussi d'améliorer notre compétitivité, et
enfin de <a target="_blank" onclick="return false;"
class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/premier-
groupe/coordonner">coordonner</a> nos politiques
économiques. Les pays qui sont en excédent doivent <a
target="_blank" onclick="return false;"
class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/premier-
groupe/stimuler">stimuler</a> leur demande intérieure par
une augmentation des salaires et une baisse des
prélèvements, c'est la meilleure expression de leur
solidarité. On ne peut pas <a target="_blank"
onclick="return false;" class="lien_interne"
href="http://conjugaison.lemonde.fr/conjugaison/premier-
groupe/infliger">infliger</a> une peine à perpétuité à des
nations qui ont déjà fait des sacrifices considérables, si
les peuples ne constatent pas, à un moment, les résultats
de leurs efforts.
```

Source : Extrait du code HTML de l'article publié sur Lemonde.fr le 17 octobre 2012

« François Hollande : "L'Europe ne peut plus être en retard" », par Sylvie Kauffmann

L'autorité redistribuée au conjugueur dilue l'autorité redistribuée aux sources ou même aux autres sites et sous-domaines du *Monde*. En effet, plus les liens sont nombreux, et moins chaque lien apporte de l'autorité PageRank au contenu vers lequel il pointe. Ainsi, considéré du point de vue de Google, automatiser la

production de liens vers le conjugueur, c'est systématiquement retirer du crédit aux contenus pointés manuellement par les journalistes, et donc, en un sens, aux journalistes eux-mêmes à l'intérieur du dispositif. Les liens effectués par ces derniers ne feront presque pas, en tout cas beaucoup moins, profiter les documents pointés de leur autorité, et ce quand bien même il s'agirait de contenus produits par des journalistes du site *Lemonde.fr*. Ainsi, ce qu'il y a d'étonnant, c'est que cette pratique peut contribuer à diminuer le trafic reçu par les articles, dont le PageRank est dilué, même si elle augmente le trafic reçu par le site (en fait, par un sous-domaine).

Les référenceurs interrogés (hélas, personne au *Monde* n'a pas accepté de nous répondre), trouvent ces pratiques « non-éthiques ». Selon eux, si *Lemonde.fr* n'est pas puni par Google, c'est parce que c'est un titre trop important dans le paysage médiatique français, dont Google a besoin pour offrir un service de qualité aux internautes et profiter des effets de réseaux croisés afférents à leur présence. Ainsi, ce serait parce que *Le Monde* a déjà une marque forte, dont Google a *besoin*, qu'il pourrait se permettre de se livrer à des pratiques théoriquement interdites par la firme sans que celle-ci ne rompe l'association. D'après les acteurs interrogés, n'importe quel « petit éditeur » se livrant à une telle pratique serait sanctionné, ce qui signifierait que Google est plus tolérant vis-à-vis du Monde qu'il ne l'est avec les autres éditeurs, et que ce dernier en profite pour maximiser ses revenus d'une manière que seuls quelques sites pourraient imiter. Dès lors, un sentiment d'injustice se crée dans la mesure où certains acteurs y voient une impunité accordée à certains seulement, lesquels peuvent biaiser les classements non pas seulement grâce au SEO classique, mais en pratiquant des techniques *black hat* que leurs concurrents ne peuvent pas utiliser sans risquer d'être punis.

« Lemonde.fr, ils font ça sur le conjugueur. C'est du pur black hat. C'est fait automatiquement : des liens invisibles. Ils les rendent invisibles parce qu'il y en a tellement, et c'est tellement cochon, qu'il fallait bien le cacher. Ce qui est marrant sur lemonde.fr c'est que ce qu'ils font est complètement interdit par Google [...]. Les SEO, les annonceurs, les sites de presse ont dénoncé ces pratiques. Tout le monde les a dénoncées. C'est le truc dégueulasse. Mais voilà : c'est Le Monde, et Le Monde, ils peuvent tout se permettre, Google va pas les punir. Personne ne le croit. Eux non plus, ils ne le croient pas. Du coup, ils y vont à fond, ils en profitent. »

Reseoneo — Consultant Senior — Julien Crenn

« Chez nous, on ne ferait jamais ça. Je trouve ça horrible. T'es dans un article, tu cliques sur les verbes conjugués. Ça fait un gros scandale sur les forums. Tout le monde s'est excité contre Lemonde.fr. Ils ont été dénoncés des milliards de fois mais ils n'ont pas reçu de pénalités. Sûrement parce que c'est Le Monde. C'est dégueulasse. Le moindre petit site fait ça et il se fait dégligner directement. C'est pour ça qu'il y a vraiment une impunité sur les marques. Google ne va pas blacklister Le Monde. Ça n'aurait pas de sens. Ce serait anti-pertinent. Le Monde fait ça. Moi je trouve ça bien sur les landing page, mais là c'est poussé à l'extrême. Ça fait de la page vue. Mais c'est que pour la pub et moi les trucs que pour la pub je trouve ça débile. C'est mercantile. On est plus sur le site, on est sur autre chose. C'est de l'achat. Pour moi c'est du SEO, mais pas intéressant. »

Lepoint.fr — Référencier — Laure Aubry

Nous avons vu ici comment les liens pouvaient être « au cœur de la stratégie commerciale » (De Maeyer, 2012, p. 2). Le contenu journalistique et l'autorité de la marque sont utilisés non pas pour mettre en avant le contenu journalistique sur Google, mais un contenu froid, à faible valeur ajoutée, qui n'a rien à voir ni avec le métier du *Monde* ni avec l'actualité ni même avec les campagnes publicitaires des annonceurs. Diminué de leurs fonctions de monstration, les liens ne sont plus des « signes passeurs » (Davallon, Jeanneret, 2004) mais des outils de valorisation. Peu importe que l'internaute y clique, car ce n'est pas le trafic interne qui est visé par ces liens, mais le trafic externe en provenance de Google. La stratégie du conjugueur consiste à intégrer un contenu froid au contenu d'actualité, et à tisser un réseau de liens entre les différents documents afin que le PageRank de toutes les pages du site augmente. Ce n'est pas de la « crédibilité » qui est distribuée (Halavais, 2008, p. 43), mais de la visibilité. L'usage des liens n'est donc pas journalistique ici (et dilue même l'autorité des liens effectués par les journalistes) mais purement, simplement, automatiquement, économique.

8.2.6 Mise en perspective

Nos observations sont intéressantes à rapprocher des travaux de Juliette de Maeyer effectués à propos de l'usage des liens hypertextes par les sites de presse (De Mayer, 2012, 2013). De l'état de l'art extrêmement détaillé par notre collègue, il ressort que les liens sont peu utilisés par les sites d'information. Par exemple, Tankard et Ban ont observé 296 articles de presse en 1998 et trouvé que

94% ne comportaient aucun lien (Tankard et Ban, 1998). Plus tard, Kenney *et al.* ont observé 100 sites de presse en ligne et constaté que seulement 32% utilisaient des liens hypertextes (Kenney *et al.*, 2000). Puis des études ont été faites, montrant que lorsque les liens étaient utilisés, ils servaient essentiellement à pointer des contenus du même site. Ainsi, Dimitrova *et al.* (2003) ont étudié 473 sites d'information dans lesquels seuls 4,1% des liens pointaient vers l'extérieur. Puis Dimitrova et Neznanski ont étudié un corpus de 26 sites d'information de 17 pays différents et conclu que seulement 8 % des articles pointaient vers l'extérieur (Dimitrova et Neznanski, 2006). Enfin, Jonathan Stray a étudié un corpus de 262 articles en 2010 dans lesquels le nombre moyen de liens sortants était de 0,65 par articles (Stray, 2010).

Selon Coddington (2012 *cit. in.* De Maeyer, 2013), les hyperliens sont utilisés dans les sites issus des médias traditionnels « comme un outil de mise en contexte et de référence, en pointant principalement vers des contenus intemporels, qui rassemblent des informations factuelles sur un sujet et font autorité » (*op. cit.*, p. 16). Selon Weber (2012, *cit. in.* De Maeyer, 2013) les liens sont utilisés par les éditeurs issus des médias traditionnels pour se positionner dans « l'écologie du réseau ». Mais finalement, Juliette de Maeyer pointe qu'étant donné certains défauts de ces études, notamment leur caractère instantané, la non-uniformité et l'absence du contexte de production, « la question de ce qui explique ou éclaire la création des liens n'est pas résolue » (De Maeyer, 2013, p. 20). Notre propre étude présente le même défaut d'instantanéité, et notre échantillon de 252 articles peut paraître faible, cependant il suffit à montrer que de nombreux liens sont générés de manière automatique. Dans notre corpus, 35% des liens au minimum ont été générés automatiquement vers des pages d'atterrissage créées d'abord pour le SEO. Pour les sites du *Nouvel Observateur* et du *Point*, ce chiffre dépasse les 65%. Et pour le site du *Monde*, lorsque nous additionnons les liens effectués automatiquement (atterrissage + conjugeur), la proportion dépasse 90%. La stratégie de *PageRank Sculpting*, encore largement pratiquée malgré les efforts de Google pour contrer ce que Bruno Latour nomme des « antiprogrammes » (Latour *et al.*, 1991), explique donc en partie l'« attitude protectionniste » des sites de presse en ligne (Steensen, 2011, p. 315),

précisément car la plupart des experts en SEO sont persuadés, et ce quoi qu'en dise Google, que le maillage interne est encore une stratégie payante en matière de SEO.

Toutefois, d'après Juliette de Maeyer, le faible nombre de liens vers l'extérieur relèverait d'une composante culturelle chez les journalistes plutôt que véritablement d'une stratégie de PageRank Sculpting :

« Le lien semble dès lors représenter une matérialisation trop incarnée du mimétisme et de la circulation de l'information. À première vue, les réticences sont d'ordre économique : c'est d'ailleurs de cette manière que les journalistes les plus ouvertement opposés aux liens vers des concurrents le formulent — on ne va quand même pas leur envoyer du trafic, leur faire gagner des lecteurs. Mais à ce calcul simple s'ajoutent des couches d'enjeux. Ainsi, le poids d'infomédiaires (Rebillard et Smyrnaio, 2010) comme Google et le mystère autour de leur mode opératoire exact créent des injonctions contradictoires : au sein d'une même rédaction, on voit certains affirmer que plus de liens externes, y compris vers des concurrents, sont une stratégie gagnante en termes d'optimisation pour les moteurs de recherche ; et d'autres défendent tout le contraire. S'il y a un effet économique des liens externes vers des concurrents, l'ampleur de celui-ci ou même son sens (s'agit-il d'un effet positif ou négatif ?) ne sont pas connus — en tout cas pas des journalistes — et, dans les deux rédactions étudiées, il ne semble pas exister de pressions directes de l'entreprise sur les journalistes pour produire ou ne pas produire de liens. Les réticences à ouvrir des routes directes vers la concurrence seraient donc plutôt d'ordre culturel, ou le fruit d'une culture d'entreprise bien intégrée par les journalistes. » (De Mayer, 2013, p. 248)

Les propos de Juliette de Maeyer vont clairement dans le sens de ce que nous ont dit certains des référenceurs interrogés lorsqu'ils accusaient les journalistes d'être la cause de l'absence de liens externes. Cependant, ses propos ne concernent que les liens effectués par les journalistes directement. Pour ce qui est des liens effectués automatiquement vers des pages d'atterrissage, il nous semble évident que lorsqu'ils sont invisibles, ils ne sont pas destinés à l'internaute, mais utilisés comme un outil de valorisation destiné au seul crawler. En revanche, nous avons vu dans le cas du *Point* que les liens automatiques, lorsqu'ils ne sont pas invisibles, peuvent être massivement utilisés par les lecteurs, et ce même si ce n'est pas leur objectif.

Il est intéressant de souligner ici que le maillage interne qui résulte de l'automatisation du « linking » n'est pas lié directement au fonctionnement *réel* des crawlers, de l'indexeur et de l'algorithme, et encore moins au fonctionnement *annoncé* par Google, mais à la manière dont le fonctionnement des crawlers, de

l'indexeur et de l'algorithme est *compris* et *perçu* par l'éditeur. En effet, étant donné le flou dont le fonctionnement exact du dispositif fait l'objet, c'est ce que les acteurs comprennent au sujet de ce qu'ils ont intérêt à y faire qui induit leurs actions, et non ce qu'ils ont vraiment, possiblement ou hypothétiquement intérêt à faire. La performativité est indirecte. Quant à nous, chercheurs, c'est en comprenant ce que comprendraient les acteurs que nous avons pu rendre compte du sillon tracé par leurs cours d'action, leurs intérêts et leurs projets. C'est en comprenant leur vision de l'énoncé, et non l'énoncé lui-même, que nous en avons sondé la performativité.

8.3 Dans les ramifications de l'arbre de navigation

Il nous reste à voir un point important à partir de notre analyse descriptive de la Partie 2. En effet, maintenant que nous avons vu de quel arbre de navigation la liste hypertexte était la racine, il s'agit de positionner les sites des éditeurs de notre échantillon dans cet arbre, afin de voir, d'une part à quel endroit ils se situent dans le réseau d'incitations, d'autre part ce qu'ils peuvent en attendre et, enfin, à quelles incitations ils sont eux-mêmes soumis.

8.3.1 Du bon côté du réseau d'incitations

Sur les six sites observés, tous sont partenaires du réseau de syndication Doubleclick, et deux sont également partenaires du programme AdSense (Nouvelobs.com et Leparisien.fr). Autrement dit, les pages de chacun des SIPI auditionnés sont potentiellement génératrices de revenus pour Google. Par ailleurs, il a été particulièrement intéressant de faire notre observation en octobre 2012, car Google menait alors une campagne de publicité pour ses tablettes numériques *Nexus*. Sur certaines des pages de notre échantillon, il y avait donc des publicités *de* Google, mais aussi des publicités *pour* Google. Par exemple, ci-après, sur le site *Leparisien.fr*, une publicité « Doubleclick » est faite en haut de la page pour la tablette Nexus.

Figure 41. Capture d'écran : Leparisien.fr (11/10/12)



Ainsi la firme, en plus d'être un fournisseur de trafic, est un concurrent, un partenaire et un annonceur, donc un client des éditeurs, sur le marché de la publicité. Cela rend plus complexe encore la situation de coo-pétition décrite dans le chapitre 5. Finalement, si nous résumons nos observations dans un tableau, nous obtenons :

Tableau 16. Positions dans l'arbre de navigation

	Publicités DoubleClick	Publicités AdSense	Nombre publicités Google	de pour
Lexpress.fr	X		2	
Lefigaro.fr	X		1	
Lemonde.fr	X		2	
Nouvelobs.com	X	X	3	
Leparisien.fr	X	X	12	
Lepoint.fr	X		0	

Google a donc *intérêt* à générer des liens pointant vers les SIPI, dont l'audience est pour lui profitable au coût par affichage. Les éditeurs qui adhèrent à DoubleClick et AdSense ne le font pas en pensant que Google manipule les résultats au profit de ses partenaires, mais ceux que nous avons interrogés avaient à l'esprit que cela pouvait être possible et qu'étant donné les incitations entremêlées, comme le

disaient Sergey Brin et Larry Page eux-mêmes en 1998²⁹⁷, l'ajout « discret » d'un paramètre à l'algorithme pourrait permettre d'avantager systématiquement les partenaires de Google sur le marché de la publicité sans que cela soit remarqué.

Par ailleurs, en termes de SEO, adhérer au réseau de syndication de Google permet que d'autres crawlers que le Googlebot scannent les pages, et en particulier le *Mediapartners* (également appelé *AdSensebot* ou *Mediabot*) exclusivement dédié à prendre connaissance du contenu des sites adhérant au réseau AdSense. Dans le robots.txt du site Leparisien.fr, on retrouve le *Mediapartners*-Google explicitement mentionné et juxtaposé à la fonction « allow », comme s'il s'agissait de « l'inviter » à prendre connaissance des pages. Si cela n'a théoriquement aucun effet (comme nous l'avons expliqué dans le chapitre 6, la fonction « allow » n'est pas reconnue par le protocole d'exclusion), c'est tout de même significatif d'une certaine position de la part de l'éditeur, qui fait ainsi en sorte que ses pages soient scannées à la fois par *Googlebot* et par *Mediapartners* (sans pour autant en être assuré).

Lors de notre enquête, plusieurs acteurs nous ont donné des informations intéressantes à propos de la manière dont ils concevaient le réseau d'incitations et la possibilité que Google avantage ses partenaires dans les résultats naturels.

« Est-ce que tu penses que ça aide au bon référencement des pages d'avoir ces modules AdSense ?

Eh bien... non, mais oui ! (rire) non mais oui, oui, oui, oui. En fait, ce qui est assez drôle, c'est que chez Google, les départements sont très compartimentés mais... en théorie ! En vrai, c'est toujours mieux d'être à la fois annonceur et acheteur. Après, quelle est la part d'aide ? J'en sais rien, je ne le saurai jamais, il n'y en a sûrement pas d'ailleurs, enfin je sais pas, personne ne sait, mais c'est exactement la même chose que le relationnel que tu peux entretenir avec tes contacts : tu ne sais jamais à quel point le fait de connaître quelqu'un t'aide ou pas à avoir un meilleur rapport commercial avec lui, mais quand même : tu te dis que c'est mieux de le connaître et de bien s'entendre avec lui. Ça ne peut qu'arranger les choses ! »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bouguié

« [Les porte-parole de Google] disent que AdSense n'influence pas le positionnement. Mais après, ça, c'est ce qu'ils disent... On ne peut pas en être sûr. »

Leparisien.fr — Référencier — Guillaume Giraudet

²⁹⁷ « However, there will always be money from advertisers who want a customer to switch products, or have something that is genuinely new. But we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm. » (Brin et Page, 1998, Annexe, A).

« Sur AdSense, il y a des tests à faire. L'idée que ça puisse augmenter le référencement naturel d'avoir du AdSense sur mes pages, moi je trouve ça très pertinent. Il faudrait voir pour un template Article Actualité, si on active AdSense ou DoubleClick, quand le crawler AdSense est dédié, si ça augmente le trafic. Dans les guidelines de Google, il y a marqué que AdSensebot sert à explorer les pages, mais qu'il y a aussi une analyse des pages pour afficher des pubs en rapport à l'article, qui est une analyse assez fine. Par exemple, si je suis sur un site de fromage, il va m'afficher des pubs qui parlent de raclette ou d'instruments. Donc il y a une analyse sémantique qui est faite. AdSensebot fait venir sur le site un crawler différent qui prend des informations qui a priori peuvent aussi être employées pour le SEO.

Après, il faudrait prendre des pincettes énormes pour dire ça. Il faut faire des tests, parce que la question se pose vraiment. En attendant, même si c'est faux, c'est possible que pleins de gens fassent du AdSense juste pour gonfler leur référencement naturel, sans se rendre compte que ça ne marche pas. La notion de ranking de la source est prédominante à cette question d'AdSense, parce que de toute façon, AdSense ne suffira pas. Ce serait comme dire que si j'ai une campagne AdWords, ça va forcément améliorer mon ranking naturel. Ça c'est faux et archi faux. Il y a aussi un bot pour le SEM. Sauf qu'AdSense il y a une analyse sémantique qui est faite de la page pour ressortir une pub pertinente, et c'est possible que ça récupère de la data pour le SEO. »

Resoneo — Consultant Senior — Julien Crenn

Nous voyons comment certaines actions peuvent être effectuées non pas étant donné le fonctionnement du moteur de recherche, mais étant donné la multiplicité des métiers de Google et le réseau d'incitations qui enserré le processus de communication. Sur le plan méthodologique, nous avons ici un exemple de ce que la démarche entreprise pour cette Partie 3 permet d'ajouter au déploiement du dispositif effectué dans la Partie 2. Notre enquête ne permet pas de dire si la majorité des éditeurs qui adhèrent à DoubleClick et/ou à AdSense le font parce qu'ils estiment que ces services sont meilleurs que les autres, ou dans l'espoir de recevoir davantage de trafic en provenance des moteurs. Cependant, elle permet de révéler que certains acteurs, parmi les SIPI de notre échantillon, pensent qu'il est préférable d'être du côté des éditeurs que Google aurait intérêt à avantager dans le cas d'un biais, et agissent en fonction. Ainsi, sur le marché de la publicité, la firme pourrait être d'autant plus compétitive qu'elle augmente les incitations à s'enrôler pour les éditeurs, étant donné

non pas sa performance sur ce marché mais le monopole dont elle jouit sur le marché des moteurs et le secret dont l'algorithme d'appariement fait l'objet.

8.3.2 Ligatus, marqueur d'indépendance

Tous les sites de notre échantillon sont partenaires du réseau de syndication Ligatus, concurrent direct de la régie AdSense de Google. Parmi eux, Nouvelobs.com et Leparisien.fr ont la particularité d'être clients des deux régies en même temps. « Lancé fin 2009, Ligatus travaille pour des acteurs « premium » issus de la presse imprimée tels que LeFigaro.fr, Lequipe.fr, LesEchos.fr, Lepoint.fr, LeMonde.fr, NouvelObs.fr, Lexpress.fr, Lexpansion.com. L'entreprise aurait commercialisé 500 millions d'impressions publicitaires en liens sponsorisés depuis son lancement jusqu'à février 2010. [...] Il s'agit là d'une façon pour les acteurs traditionnels de l'industrie de la presse et de la publicité de ne pas alimenter le succès de Google lorsqu'ils désirent proposer cette forme de publicité sur leur site, même si cela reste une part négligeable en valeur des revenus issus de l'activité publicitaire de la régie web. » (Ouakrat, 2011, p.252)

La volonté mise en avant par Alan Ouakrat de « ne pas alimenter le succès de Google » a été clairement mentionnée par les interviewés de notre enquête, selon lesquels l'adhésion à Ligatus était justifiée étant donné la qualité du service proposé et, d'autre part, étant donné une émancipation relative vis-à-vis de Google.

« Maintenant, on a Ligatus. On a essayé de faire les choses de manière à être moins "Google-dépendant" qu'on était. Nous étions dans le réseau AdSense avant, mais nous n'y sommes plus, et c'est très bien comme ça. »

Lefigaro.fr — Rédacteur en chef — Luc de Barochez

Le référenceur du site Leparisien.fr, client à la fois d'AdSense et de Ligatus, nous a dit qu'il trouvait que Ligatus était plus qualitatif, mais qu'il ne comptait pas se priver pour autant des opportunités offertes par AdSense :

« Je trouve que Ligatus est plus pertinent. Tous les acteurs médias ont du Ligatus, c'est pas pour rien. Mais ma recommandation est simple : moi je pense que dès que je gagne de l'argent avec un truc, je le mets. Si ça ne me faisait rien gagner, je ne mettrais pas AdSense. »

Leparisien.fr — Référenceur — Guillaume Giraudet

Alan Ouakrat a montré que, dans le cas des éditeurs issus de la presse imprimée, la publicité « search », type AdSense et Ligatus, était utilisée pour monétiser les espaces publicitaires invendus (Ouakrat *et al.*, 2010 ; Ouakrat, 2011). Toutefois, il a également mis en évidence le dilemme auquel faisaient face les SIPI quant à ce procédé de monétisation : « D'un côté, il est tentant de confier cet inventaire à des *ad-networks*, qui pourront l'intégrer à leurs offres à bas prix ou à la performance, et faire en sorte qu'"il rapporte quelque chose plutôt que rien". De l'autre, accepter de vendre une partie de leurs inventaires à la performance revient à accepter une opération de déqualification, le transfert de leur produit d'un dispositif de qualification à un autre, qui contredit l'ensemble de leurs efforts pour défendre la qualité de leurs espaces. La présence d'espaces publicitaires "de marque" au sein des *ad-networks* crée un risque de dévalorisation de la marque : auprès des acheteurs, qui sont susceptibles de s'en rendre compte et d'être plus réticents à payer 5 € les mille affichages l'espace que d'autres ont acheté à 1 € le clic ; et auprès des annonceurs, qui peuvent estimer que le contrat de lecture de la page web d'un site de qualité est rompu lorsqu'y apparaissent des publicités émanant d'acteurs très peu prestigieux. » (Ouakrat *et al.*, 2010., p.156).

A la lumière de ces réflexions et de nos propres observations, il semble que la publicité « search » soit confiée essentiellement à Ligatus dans le but de s'émanciper de Google, à part pour le cas de Nouvelobs.com et Leparisien.fr. D'autre part, il est important de souligner qu'une telle stratégie peut aboutir à une « déqualification » et tirer les prix des publicités « display » vers le bas. Il sera intéressant de suivre dans les prochains mois les résultats de l'accord noué entre la presse française et Google, dont le deuxième volet prévoit un partenariat privilégié sur le marché de la publicité (cf. 6.2.4). En effet, ce deuxième volet aboutira-t-il à ce que les éditeurs quittent Ligatus pour se tourner à nouveau vers AdSense ? Dans ce cas, une stratégie commune DoubleClick/AdSense sera-t-elle mise en place conjointement par Google et les éditeurs afin d'éviter la déqualification ? *Quid*, alors, de la dépendance des éditeurs vis-à-vis de Google ?

8.3.3 Référencement payant

Il peut arriver que les SIPI effectuent des enchères sur des mots-clés, via le dispositif AdWords, afin d'apparaître dans la liste de liens sponsorisés, à droite de la liste de résultats « naturels » de Google Search. Dans le modèle économique multiversant des éditeurs, la subvention accordée aux lecteurs dans le but de les faire embarquer à bord de la plate-forme — lecteurs dont l'attention est monnayée sur le versant des annonceurs — n'est plus basée alors sur la distribution gratuite, mais sur la captation payante des lecteurs. Le versant non-rémunérateur devient source de dépenses supplémentaires, ce qui oblige l'autre versant, c'est-à-dire l'activité publicitaire des éditeurs, à être encore plus rémunérateur pour recouvrir les coûts de captation. En conséquence, un arbitrage doit être fait pour paramétrer l'enchère AdWords : à combien l'éditeur évalue-t-il la présence d'un internaute supplémentaire sur son site ? combien cette présence générera-t-elle ?

Aucun chiffre exact ne nous a été donné à propos du nombre de visiteurs arrivés par ce biais, mais nous avons tout de même obtenu certaines informations intéressantes. D'une part, la présence sur AdWords nous a été présentée, à *L'Express* et au *Parisien*, comme un levier d'action permettant d'augmenter rapidement l'audience totale du site et ainsi de se maintenir au-dessus d'un certain nombre de visiteurs uniques et de pages vues. Ce levier est en général confié à un prestataire.

Les visiteurs uniques et les pages vues sont mobilisés lors des négociations avec les annonceurs, et constituent par conséquent des variables explicatives des revenus totaux générés par la publicité facturée au coût par affichage. Ainsi, même si un internaute arrivé par un lien sponsorisé coûte plus cher qu'il ne rapporte à titre individuel (coût marginal supérieur au profit marginal), sa présence sur la plate-forme s'additionne à la présence des autres internautes, ce qui permet, au total, d'atteindre un nombre de pages vues suffisamment valorisé par les annonceurs pour que les prix des publicités soient maintenus élevés et qu'ainsi l'ensemble des coûts, y compris ceux de l'acquisition d'audience par AdWords, soient recouverts. Par ailleurs, un lecteur arrivé par le biais d'AdWords peut devenir rentable s'il est fidélisé :

« Un internaute provenant d'AdWords te coûte-t-il plus cher qu'il ne te rapporte ?

Oui, sauf si je réussis à capter son adresse email. Sauf si je le renvoie vers un dossier qui va générer beaucoup de pages vues [i.e. un fort taux de rebond]. Sauf s'il me génère un partage sur Facebook. Je peux donc m'y retrouver. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bouguié

Les mots-clés choisis sur AdWords peuvent être liés au nom du titre — *Monde, Figaro, Parisien* — mais peuvent aussi être liés à des actualités : DSK, Présidentielles, Syrie, Printemps arabe. Ainsi, le lien généré par AdWords est une forme de publicité pour le titre de presse mais aussi pour le traitement d'un fait d'actualité en particulier, c'est-à-dire pour un article ou un dossier que l'organisation souhaite mettre en avant en payant les clics des internautes. En plus de faire de la publicité au sens classique du terme, l'éditeur ouvre ainsi une nouvelle voie d'accès à son contenu en positionnant ses liens à droite de la liste naturelle. Nous verrons ci-après dans le cas de *L'Express* comment AdWords peut être utilisé pour se positionner sur certains sujets d'actualité, certains thèmes, pour lesquels le titre estime qu'il n'a pas la reconnaissance qu'il mérite étant donné le travail effectué par les journalistes. Il ne s'agit plus alors de rentabiliser directement les internautes qui arriveront par ce biais, mais de positionner la marque en haut des pages de résultats, sur des requêtes où elle n'aurait peut-être pas figuré dans les résultats naturels. C'est alors directement la légitimité des journalistes qui est mise en avant (et pour laquelle l'éditeur paye).

Dans les entreprises de presse qui n'adhèrent pas à AdWords, comme c'est le cas au *Point*, cela peut créer une frustration du côté du référenceur. Cependant, celui du *Point* nous a finalement dit qu'après avoir regretté l'absence de stratégies de *Search Engine Marketing* (SEM, et non SEO), il avait fini par trouver que la situation était mieux ainsi, car elle permettait de ne pas verser d'argent à Google :

« Ils ne veulent absolument pas faire d'achat de mots-clés. Au début, je le regrettais. Mais finalement AdWords, c'est très artificiel. Tu donnes de l'argent à Google en achetant le trafic. Donc je n'ai plus de regret. »

Lepoint.fr — Référenceur — Laure Aubry

Pour la stratégie SEO, et donc pour le référencement dit « naturel », les stratégies de référencement payant et de SEM peuvent pourtant s'avérer intéressantes. Les économistes Sha Yang et Anindya Ghose ont montré en effet qu'un cocktail judicieux entre les liens sponsorisés et le référencement naturel était susceptible d'aboutir à un meilleur classement naturel, autrement dit qu'une stratégie SEM pouvait participer à la bonne marche d'une stratégie SEO. Selon eux, le taux de clics des liens sponsorisés aurait une influence positive sur le taux de clics des liens naturels, et vice et versa (Yang et Ghose, 2010). Par ailleurs, le fait d'être un client d'AdWords permet à un éditeur de recevoir sur ses pages le crawler dédié à la plateforme : AdsBot. Ainsi, l'éditeur qui serait à la fois client d'AdWords et partenaire d'AdSense pourrait potentiellement recevoir la visite des trois crawlers *AdsBot*, *Mediapartners* et *Googlebot*, et renforcer ainsi son implication dans le dispositif.

Il existe deux autres arguments en faveur d'une stratégie de SEM incorporée à la stratégie SEO. D'une part, le SEM permet de tester à très court terme la réactivité de certains mots-clés. Le SEO, en revanche, est une stratégie de longue haleine, qui consiste à se positionner à moyen et long terme et dont on ne peut voir les résultats que plusieurs mois après le début des mises en œuvre. Ainsi, le SEM pourrait servir à celui qui met en place une stratégie de SEO à « ne pas se tromper » lorsqu'il choisit sur quels mots-clés, quels thèmes, quels sujet, il se positionne.

« Le SEM permet de se positionner en direct. Adwords me permet de savoir si mon site est bien fait pour convertir sur tel ou tel mot. Après, je vais éventuellement lancer une campagne SEO, dont je ne verrai les résultats que beaucoup plus tard. Donc le SEM, ça me sert de test. Il me permet de savoir si c'est une bonne idée de se positionner là dessus. Je dis à mon client : Mets toi au moins un peu en SEM, débloque un petit budget en SEM pour savoir quoi faire, et après on fera du SEO efficace. Imagine que tu dépenses 10 000 euros sur un mot-clé dont tu pensais qu'il allait cartonner, et tu te rends compte que finalement c'était l'autre sur lequel il aurait fallu parier... Ce serait quand même dommage, non ? »

Resoneo — Consultant — Jean-Christophe Baudrier

Enfin, il existe un dernier argument en faveur du SEM. Il se trouve que la firme Google donne davantage d'informations aux éditeurs engagés sur la plateforme AdWords et partenaires des réseaux de syndication DoubleClick et AdSense — via notamment son « Outil pour les webmasters » et « Google Analytics » — qu'elle n'en

donne à ceux qui ne sont ni client ni partenaire. En fait, les informations des internautes identifiés par Google (parce qu'ils ont par exemple un compte Gmail ou Google+) sont cachées aux éditeurs qui ne sont ni clients d'AdWords ni partenaires d'AdSense. Ainsi, faire du SEM ou adhérer à AdSense permet à un éditeur d'obtenir d'avantage d'informations aidant à définir une stratégie de SEO. Cela nous a été présenté comme une stratégie de la part de Google visant à rendre plus attractif son service AdWords (rémunérateur) et les partenariats AdSense et DoubleClick (rémunérateurs). Les éditeurs sont ainsi invités à s'associer un maximum aux différents métiers de Google s'ils veulent maximiser leur visibilité sur les listes de résultats du moteur. De son côté, la firme présente cette mesure comme un moyen de protéger les données des usagers de ses services. Ainsi, les internautes sont eux aussi invités à s'inscrire un maximum, en ouvrant des comptes sur Gmail ou sur Google+, afin que leurs données soient protégées (tout en permettant à Google d'effectuer (et de monétiser) un meilleur ciblage).

« Google en ce moment, sur Analytics, ils ne donnent plus toutes les données. Par exemple, je n'ai plus accès aux requêtes de ceux qui sont connectés à leur compte Gmail. Quand quelqu'un qui a un compte Google, et qui est connecté, tape « voyage », sur Analytics, ils me mettent : « not provided ». Du coup, je ne saurai jamais qu'est-ce que le gars a tapé pour arriver sur mon site. En tant que SEO, ça me gonfle à mort, parce que c'est clairement contre nous : Google a fait ça pour nous mettre des bâtons dans les roues. Alors qu'en SEM, en revanche, les éditeurs payent, et du coup ils ont accès à toutes les données. Ceux qui payent n'ont pas ce problème du not provided. Comme ils sont partenaires, Google leur donne tout. Par contre, nous en tant que SEO, ils s'arrêtent de tout nous donner. Raison de plus pour faire du SEM, y compris pour le SEO. En faisant ça, Google nous a dit qu'il voulait protéger l'utilisateur, sauf qu'en tant qu'éditeur, ou en tant qu'agence, je m'en fiche, d'un tel ou un tel, parce que moi ce que je veux ce sont les données agrégées. Mais si je suis en SEM, je paye, donc j'ai tout ! Google invite donc fortement les éditeurs à faire du AdWords, et les internautes à avoir un compte Gmail. »

Resoneo — Consultant — Jean-Christophe Baudrier

Nous retrouvons ici la stratégie de diversification concentrique opérée par Google (chapitre 5). Les différents métiers de la firme se renforcent les uns les autres. Par ailleurs, il est intéressant de noter que Google, apparemment, ne vend pas seulement du trafic aux clients d'AdWords, mais également des données, dans la

mesure où l'accès à certaines données est conditionné au fait que l'éditeur soit un client d'AdWords. Nous retrouvons ce que Yann Moullier-Boutang nomme la « pollinisation » (Moullier-Boutang, 2009, 2010) : en adhérant à Gmail ou à Google+, les internautes génèrent des données que Google ne divulguera qu'à condition qu'un éditeur paye en utilisant AdWords, ou ne partage avec elle ses profits en devenant un partenaire du réseau AdSense ou de la régie DoubleClick. L'apiculteur accepte que ses abeilles pollinisent le champ de l'agriculteur à la seule condition que ce dernier achète le miel qu'il produit grâce à elles.

8.3.4 Vente de liens

Nous n'avons trouvé aucun « lien vendu » par les SIPI lors de nos observations. Vendre un lien visible ou invisible, notamment depuis la page d'accueil, permettrait pourtant à celui dont le contenu est pointé d'augmenter, grâce au « *PageRank Sculpting* », ses chances d'être bien référencé. En outre, cela lui permettrait de se positionner à un niveau n+2 de l'arbre de navigation et de bénéficier ainsi possiblement d'une part du trafic reçu par un site bien positionné. Il semble ainsi que Google ait réussi à empêcher qu'un marché du lien ne se crée en France autour des marques de presse imprimée. Non pas que les éditeurs aient peur que Matt Cutts ne tue un chaton à chaque fois qu'un lien est vendu (cf 8.2.1), mais plutôt que, pour les éditeurs, la vente de liens soit perçue comme une mauvaise pratique, voire une « tromperie » dans le cas d'un lien invisible, et ce quand bien même la vente ou la location d'un lien hypertexte pourrait générer des revenus conséquents. Ce ne serait plus alors seulement le trafic qui serait vendu, mais également l'autorité PageRank. Pour le moment, dans le cas des SIPI français, celle-ci est encore distribuée gratuitement à ceux dont le contenu est pointé. En revanche, étant donné les stratégies intensives de maillage interne, elle demeure relativement faible.

La vente de liens est pourtant pratiquée par certains sites, qui monétisent ainsi leur autorité PageRank. C'est le cas en Angleterre de *Express Group* qui vend des liens à des partenaires depuis les sites d'actualité du *Daily Express* (<http://express.co.uk/>) du *Daily Star* (<http://www.dailystar.co.uk>) et de

OK (<http://ok.co.uk/>)²⁹⁸, et ce malgré les menaces de sanction de Google. Le lien est mis en ligne en tête de rubrique, pour une durée variant entre deux semaines et un mois, puis archivé sur une page article pour une période de 15 mois. La facture, pour le partenaire, s'élève à 1000 Livres Sterling et peut être négociée. Les partenaires adhérant au programme peuvent préciser sur quelles ancres les liens doivent être positionnés (i.e. sur quels mots), et vers quelles URL ils doivent pointer.

Nous pourrions tout à fait imaginer, dans le cas des SIPI aux marques fortes, parce que Google a besoin de référencer leurs contenus pour bénéficier des effets de réseaux afférents à leur présence, qu'ils ne respectent pas les recommandations de la firme et vendent des liens à des partenaires pour monétiser leurs contenus autrement qu'avec de la seule publicité « display » ou « search ». Cela est d'autant plus tentant quand on considère les difficultés des éditeurs à équilibrer leurs modèles d'affaires. Nous avons déjà vu d'ailleurs dans le cas du maillage interne du *Monde* comment certains sites pouvaient ne pas respecter les recommandations de Google, parce que justement ils estimaient que la firme n'aurait malgré tout pas intérêt à les sanctionner.

« Google dit qu'il ne faut pas acheter des liens, mais finalement ça se fait beaucoup notamment quand les sites sont importants pour Google, parce que quand le site est important, Google n'a pas intérêt à le déclasser. »

Lepoint.fr — Réferenceur — Laure Aubry

« Metro France, ils vendent des liens. Et ça c'est très bon. Et ça peut coûter très cher. Tu as un lien, sur la home du Monde, un lien qui peut très bien être un footer à peine visible, c'est génial. Après, ça reste un revenu complémentaire pour un site. Ça dépend combien il vend de liens, mais il ne peut pas vendre 100 000 liens sur la home page. Mais ça peut quand même te faire un revenu. Si j'étais un site de presse, je vendrais des liens au moins sur mes articles. Après, c'est pour tromper Google. Ça peut être un jeu dangereux. Tu as différents types de vente de liens. Tu as le revenu en location d'un lien, c'est-à-dire que je vais louer un lien sur une page pour un temps déterminé, genre six mois ou un an, et je vais te payer 300 euros par mois. Ça peut être même plus cher. Ou alors, je vais acheter un lien dans un article, qui ne sera annoncé en home page qu'un moment, et qui va redescendre ensuite dans l'arborescence du site. Dans ce cas, je vais acheter un lien à 100 euros, à vie. Il y a des contrats comme ça. Sur Le Monde, dans « voyage », ou « immobilier », ils pourraient faire ça. Ou si je vends des chaussures de sport, je veux acheter un lien sur L'Equipe.

²⁹⁸ <http://econsultancy.com/fr/blog/5798-express-group-turns-to-selling-links-big-companies-caught-buying> (dernière visite le 23 août 2013)

C'est un peu caché tu vois. Ça passe derrière. Sous le manteau, quoi. Les gens ne disent pas ça sur les forums. Google a déjà sanctionné des gens pour ça. Après, est-ce qu'ils sanctionneront Le Monde ? Je ne pense pas. »

Resoneo — Consultant — Jean-Christophe Baudrier

Au sujet de la vente de liens, étant donné le peu d'éléments que nous avons récoltés à ce sujet, nous n'irons pas plus loin que le seul questionnaire. Nous voudrions simplement souligner que ne pas vendre de lien est une manière pour les éditeurs d'adhérer au script de Google et de contribuer, sur ce point, à ce que la hiérarchisation de l'espace web soit conforme à ce que les concepteurs du moteur ont défini comme des critères de pertinence. Au contraire, la vente massive de liens « tromperait » le PageRank en aboutissant à un système où, en quelque sorte, les votes pourraient être vendus au plus offrant plutôt qu'orientés librement vers le plus pertinent. Ainsi, les éditeurs qui ne vendent pas de liens ne *trahissent* pas la chaîne d'associations, tandis que ceux qui en vendent mettent en péril le fonctionnement de l'ensemble du réseau au profit de leurs seuls intérêts : les internautes risqueraient de trouver que le moteur génère des résultats non-pertinents, les éditeurs seraient encouragés à investir dans l'achat de liens plutôt que dans la qualité des contenus, les éditeurs respectant les recommandations de Google pourraient en pâtir, et Google voir ses revenus diminuer.

De la même manière qu'il existe dans le cas de Google des incitations pour biaiser les résultats en faveur des partenaires d'AdSense et de DoubleClick, il existe donc des incitations dans le cas des éditeurs pour acheter des liens. Si ceux-ci ne le font pas, ou peu, rien ne dit que cela durera, car les incitations, elles, demeurent. En effet, exactement comme les incitations révélées dans le chapitre 5, les incitations à la vente de liens « pèsent » sur le processus de communication dans la mesure où elles sont susceptibles d'influencer son résultat. Parce que les difficultés financières des éditeurs de presse contribuent à renforcer ces incitations, la traduction est menacée par une possibilité de trahison, de la même manière que dans le célèbre exemple de Michel Callon, les pêcheurs, trop intéressés par les profits à court terme, finissent par trahir les scientifiques en ramassant les coquilles Saint-Jacques de la baie de Saint-Brieuc (Callon, 1986, cf. 1.3.4). C'est pourquoi la firme Google menace, sanctionne et communique, désireuse d'éviter autant que possible qu'un marché de l'autorité

PageRank ne s'organise et ne nourrisse la performativité de la formule, et c'est pourquoi elle va jusqu'à lier le sort de chatons innocents à celui de la chaîne d'associations articulée par l'algorithme.

8.4 Sous la Coupole de L'Express

Nous mentionnerons ici certaines particularités concernant le site Lexpress.fr. D'abord, nous exposerons la gestion des sous-domaines qui y a été mise en place. Puis nous décrirons la stratégie « *Sitemaps* ». Après quoi nous nous intéresserons au maillage interne et en particulier aux pages d'atterrissage. Enfin, nous parlerons de la stratégie AdWords. Nous pourrions ainsi préciser certaines des observations effectuées ci-avant au niveau mezzo, pour le pan monographique de notre terrain.

8.4.1 Un seul domaine

Les quatre titres de Roularta Media France traitant l'information d'actualité politique, générale, économique, financière et industrielle ont été rassemblés en 2012 sous un domaine commun lors d'un projet nommé « la Coupole ». Ce projet a été réalisé dans le but de créer des synergies de plusieurs ordres, parmi lesquelles nous nous concentrerons ici sur les éléments concernant le SEO.

Tableau 17. Les 4 titres et la Coupole

Titre papier	Périodicité	Spécialité	Adresse du sous-domaine
<i>L'Express</i>	Hebdomadaire	Politique et générale	http://www.lexpress.fr
<i>L'Expansion</i>	Mensuel	Economie	http://lexpansion.lexpress.fr
<i>L'Entreprise</i>	Mensuel	Industrie	http://lentreprise.lexpress.fr
<i>Votre Argent</i>	Mensuel	Finance	http://votreargent.lexpress.fr

Pour les trois titres mensuels *L'Expansion*, *L'Entreprise* et *Votre Argent*, trois sous-domaines ont été créés. Du point de vue SEO, autant que du point de vue de la stratégie commerciale, cela permet d'avoir davantage de force : mutualiser le

PageRank et l'AuthorRank, rassembler le maillage interne à l'intérieur d'un seul domaine, peser plus lourd en termes de nombre de documents mis en ligne, de pages vues et de visiteurs uniques. Tout cela est susceptible de contribuer à maximiser la visibilité des différents titres sur les dispositifs Google Search et Google Actualités. Comme nous l'avons déjà dit, les porte-parole de Google ont annoncé à la fin des années 2000 que les liens effectués depuis des sous-domaines seraient désormais considérés comme des liens internes, c'est-à-dire qu'il serait porteur de moins d'autorité PageRank que s'ils étaient sur d'autres sites web. Cependant, ces liens portent malgré tout de l'autorité et la mutualisation sous un même nom de domaine des différentes marques de RMF permet de simplifier les rapports avec Google et donc, possiblement, l'apport de trafic.

« Le projet Coupole, c'est la migration de toutes les adresses URL sur leexpress.fr. J'ai fait un an de travail pour ça avec un consultant et des experts SEO. Si on a fait ce choix, c'est qu'on considérait qu'on était gagnant, et franchement ça s'est passé beaucoup mieux qu'on aurait cru. Sur ce type de migration, on n'a pas connu de problème majeur, parce que si le risque avait été trop grand on ne l'aurait pas pris. »

Lexpress.fr — Editrice — Sophie Gohier

Etant donné cette stratégie de Coupole, sur la liste hypertexte affichée par Google Actualités, les trois titres de presse imprimée apparaissent comme étant entièrement intégrés à *L'Express*. Tant qu'il n'a pas cliqué sur le lien, l'internaute croit que l'article a été produit par un journaliste de *L'Express* et publié sur le site adossé au newsmagazine, alors que ce n'est pas le cas. Ainsi, les liens vers *Votre Argent* et *L'Entreprise* seront affichés comme des liens vers *L'Express*, c'est-à-dire que la marque du titre papier n'apparaîtra pas sur Google Actualités. *L'Expansion*, *Votre Argent* et *L'Entreprise* sont privés d'existence patronymique. Voici par exemple comment apparaissait sur Google Actualités le lien vers l'article publié le 11 juin 2013 par François Kermaoal, journaliste à *L'Entreprise*, dont l'article a été publié sur le sous-domaine <http://lentreprise.lexpress.fr>²⁹⁹ :

²⁹⁹ http://lentreprise.lexpress.fr/publicite-et-communication/2-35-millions-d-euros-de-stickers-cache-misere-pour-le-prochain-g8_41407.html (dernière visite le 23 août 2013)

Figure 42. Capture d'écran : *L'Entreprise* sur Google Actualités

2,35 millions d'euros de stickers cache-misère pour le prochain G8
 L'Express - il y a 43 minutes
 Les autocollants peuvent faire des miracles **pour** dissimuler la misère. Le gouvernement britannique en a couvert quelques boutiques locales ...

Et voici un exemple d'un lien vers un article de *L'Expansion*, publié à la même date par Raphaële Karayan, journaliste à *L'Expansion*, sur le site <http://lexpansion.lexpress.fr> :

Figure 43. Capture d'écran : *L'Expansion* sur Google Actualités

PS4 vs Xbox One: Sony gagne le coeur des joueurs
 L'Express - il y a 43 minutes
 Entre les deux nouvelles consoles de **Sony** et Microsoft présentées à l'E3, **Sony** prend l'avantage en étant plus proche des besoins des ...

Nous voyons dans les captures d'écran ci-dessus que le lien est associé à l'hebdomadaire *L'Express* et non aux mensuels *L'Entreprise* ou *L'Expansion*. D'après nos entretiens, il peut arriver très rarement que Google associe un lien au magazine *L'Expansion*, cependant dans nos observations les liens du mensuel étaient toujours associés à *L'Express*³⁰⁰.

Les titres traitant de sujets culturels et de mode, comme *Studio Ciné Live* pour le cinéma, *Lire* pour la littérature, ou les suppléments *L'Express Style* (féminin) et *L'Expansion Tendances* (masculin), ont été rassemblés non pas dans des sous-domaines mais dans des rubriques ou des sous-rubriques de lexpress.fr :

Tableau 18. Les 4 sous-rubriques de la Coupole

Titre papier	Périodicité	Spécialité	Adresse URL
<i>L'Expansion Tendances</i>	Trimestriel	Masculin	http://www.lexpress.fr/tendances/
<i>L'Express Style</i>	Hebdomadaire	Féminin	http://www.lexpress.fr/styles/
<i>Lire</i>	Mensuel	Littérature	http://www.lexpress.fr/culture/livre/
<i>Studio Ciné Live</i>	Mensuel	Cinéma	http://www.lexpress.fr/culture/cinema/

³⁰⁰ Nous expliquerons dans le chapitre suivant en quoi l'avalement de la marque *L'Expansion* par la marque *L'Express* peut poser problème aux journalistes.

Ces sites bénéficient ainsi du référencement sur Google Actualités, ce qui n'aurait pas forcément été évident pour chacun d'entre eux étant donné leurs positionnements respectifs. Cette stratégie a ainsi le mérite de faire bénéficier les articles rédigés par les rédactions des magazines mensuels des mêmes conditions de référencement que les articles publiés par la rédaction de *L'Express*. En revanche, ils seront eux aussi privés dans Google Actualité de leurs patronymes. De surcroît, contrairement aux sites évoqués ci-avant, les patronymes des magazines *Lire* et *Studio Ciné Live* ne figureront même pas dans l'adresse URL. Par exemple, voici un lien de Google Actualités vers un article intitulé « *Jan Ole Gerster: "J'ai vite compris que les mecs cools du cinéma, c'était les réalisateurs"* » publié en juin 2013 dans le magazine *Studio Ciné Live* :

Figure 44. Capture d'écran : *Studio Ciné Live* sur Google Actualités



L'exemple de *Studio Ciné Live* est particulièrement intéressant à mentionner, car son intégration à l'URL de *Lexpress.fr* a fait l'objet de négociations compliquées. La citation ci-après nous permet de bien comprendre comment des enjeux liés au SEO peuvent travailler le dispositif de publicisation du contenu, en conduisant certains acteurs à accepter de faire des concessions, et, ce, malgré une culture héritée de la presse imprimée qui peut entrer en conflit avec la culture web.

« *Studio Ciné Live est un bon exemple des problèmes qu'on a pu avoir. Le patron, en 2003, voulait lancer son site. On lui a dit c'était une erreur. Il y avait déjà Allociné en face, il n'y avait plus de place pour un nouvel acteur dans les classements de Google. Il n'y avait aucun enjeux business. Il n'a pas voulu nous écouter. Le choix a été tranché et on a lancé le projet. Deux ans plus tard, on a tout fermé. Impossible de se positionner en SEO, parce que tout était trusté par Allocine.com et Evene.fr. Du coup, on intégré Studio Ciné Live à L'Express et aujourd'hui ils ont une audience même s'ils n'ont plus de marque. [...] Pour des gens du digital, il n'y a pas de problème avec la Coupole. C'est naturel que les gens arrivent par Google. Mais pour des gens qui ont des cultures print fortes, c'est très difficile à comprendre. »*

Lexpress.fr — Editrice — Sophie Gohier

Cette citation montre comment la notion de marque peut être renégociée sur le web étant donné des stratégies de mutualisation liées au SEO. Nous voyons que l'Editrice essaye de faire « passer la pilule » de la perte de patronyme en avançant l'argument d'un trafic démultiplié. Le patron de *Studio Ciné Live* a quant à lui voulu développer son propre site non pas pour des enjeux économiques mais pour préserver la pérennité de sa marque. Finalement, après avoir obtenu gain de cause, il a dû faire machine arrière étant donné que les utilisateurs de Google étaient essentiellement dirigés vers ses concurrents. En intégrant une rubrique de *L'Express*, et en acceptant donc que le nom de sa marque ne figure ni dans l'adresse URL ni dans les liens générés par le dispositif d'infomédiation, *Studio Ciné Live* a pu avoir une présence sur Google que le titre n'aurait pas eu en nom propre. Par ailleurs, une fois qu'un utilisateur de Google aura cliqué sur un lien vers un sujet « cinéma », il verra apparaître le logo de Studio Ciné Live auprès de celui de « L'Express Culture » :

Figure 45. Capture d'écran : logo de *Studio Ciné Live* sur Lexpress.fr



8.4.2 Sitemaps à tiroirs

L'Express a une stratégie de valorisation des archives que nous nommons « Sitemap à tiroirs », également mise en œuvre par *Le Figaro*³⁰¹. Elle permet d'aplanir, par un chemin de navigation extrêmement simple, non hiérarchisé, depuis la racine du site, l'arborescence, et ainsi de faciliter le travail des crawlers. Ce Sitemap³⁰² n'est pas destiné aux internautes, c'est pourquoi il est volontairement caché dans l'arborescence du site. Le premier niveau du Sitemap présente, les uns à la suite des autres, les mois et les années de publication classés par ordre *ante-*chronologique jusqu'en 1953, date à laquelle *L'Express* fut fondé par Jean-Jacques Servan-Schreiber et Françoise Giroud (cf. 3.3). Chaque mois de chaque année est un

³⁰¹ Adresse du Sitemap à tiroirs du Figaro : <http://articles.lefigaro.fr> (dernière visite le 23 août 2013)

³⁰² <http://www.lexpress.fr/outils/sitemap.asp> (dernière visite le 23 août 2013)

lien pointant vers une page présentant la déclinaison des jours de ce mois classés, cette fois, par ordre chronologique. Chacun des jours du mois est à son tour cliquable et pointe vers un page où figure une liste de titres, eux mêmes classés par ordre chronologique, cliquables vers les pages, les articles, les photos et les vidéos publiés ce jour-là, sans autre hiérarchie que celle de l'heure de publication. Cela permet au crawler d'avoir facilement accès aux documents depuis la racine du site, et par ordre chronologique depuis le 16 mai 1953, et l'article de Françoise Giroud intitulé *L'homme qui nous lira*³⁰³, jusqu'au jour et à l'heure de son crawling.

Figure 46. Capture d'écran : le Sitemap à tiroirs de Lexpress.fr



The screenshot displays the L'Express.fr Sitemap interface, which is organized into a hierarchical structure of dates and topics. The main navigation bar at the top includes links for various categories: À la une, Politique, Monde, Economie, Société, Education, Emploi, Médias, Culture, High-Tech, and Sport. The main content area is divided into two columns. The left column lists dates from 06.2013 down to 11.2011, with a sub-column for the month of October 2012 (01.10 to 19.10). The right column lists topics related to the selected date, such as 'Photos des coulisses du défilé Costume National Printemps-Eté 2013' and 'Ryder Cup 2012: l'Europe s'impose après une remontée historique'. A large article titled 'Les pistolets de Bonnie & Clyde vendus 500.000 dollars' is featured prominently on the right side of the page. The article text describes the discovery of two pistols on the bodies of Bonnie Parker and Clyde Barrow, and their subsequent sale for 500,000 dollars. A sidebar on the right side of the page lists 'Toutes les dépêches' (All dispatches) with various headlines, including 'Mandela "dans un état grave mais stable" après quatre jours d'hôpital' and 'Procédure de dissolution de groupes d'extrême droite'.

³⁰³ http://www.lexpress.fr/informations/l-homme-qui-nous-lira_590619.html (dernière visite le 23 août 2013)

Nous apercevons sur notre capture d'écran (Fig. 46), à gauche, le premier tiroir : les dates rangées par ordre *ante*-chronologique. Puis, à droite, dans le deuxième tiroir : les jours du mois d'octobre 2012. Dans le troisième tiroir : les documents publiés le 1^{er} octobre 2012. Et dans le quatrième et dernier tiroir : l'article lui-même. Une telle stratégie permet à Lexpress.fr de ranger ses archives et de s'assurer que le crawler en prenne connaissance sans avoir à trop s'éloigner du niveau zéro de l'arborescence. Il s'agit clairement d'une stratégie de valorisation visant à se positionner sur la « long tail » (Anderson, 2006), c'est-à-dire à recevoir du trafic provenant de requêtes effectuées sur Google Search par des internautes intéressés par des sujets très précis, ayant fait l'actualité à un moment « t », et ayant été traités par les journalistes de l'hebdomadaire ou du site web, mais aussi par les journalistes des autres titres du groupe. En effet, étant donné la stratégie de Coupole, les articles de *L'Expansion*, de *L'Entreprise*, de *Studio Ciné Live*, du magazine *Lire* et des suppléments *L'Expansion Tendances* et *L'Express Style* figurent auprès des archives de *L'Express*, mélangés avec eux selon la chronologie des dates et des heures de publication, et donc rangés exactement au même niveau pour le crawler. Les articles du dispositif de publication destiné aux rédacteurs extérieurs (spécialistes et amateurs), « *Express Yourself* », sont eux aussi rangés au même niveau et valorisés en tant qu'archives au même titre que les articles des journalistes.

8.4.3 Pages d'atterrissage mutualisées

Comme nous l'avons vu plus tôt dans ce chapitre, Lexpress.fr effectue certains liens invisibles vers des pages d'atterrissages. Or il se trouve que la stratégie de la Coupole a permis de mutualiser les pages d'atterrissage et ainsi d'intensifier la stratégie de maillage interne. Que ce soit les articles de la rédaction de *Lire*, les articles de *Studio Ciné Live*, de *L'Expansion*, ou des suppléments masculin et féminin de *L'Express*, tous pointent désormais vers des pages d'atterrissages hébergées sous l'adresse URL : <http://fiches.lexpress.fr/>.

Ainsi, la stratégie de Coupole a permis à Roularta Media France de renforcer le *PageRank Sculpting* autour de la marque *L'Express*. L'internaute qui accèdera ensuite à une des pages d'atterrissage par Google (les liens vers celles-ci effectués

depuis le site n'étant pas visibles, ces sont principalement destinées aux utilisateurs de Google), accèdera à une fiche hébergée par *L'Express* tandis que les autres marques, qui pourtant auront distribué leur PageRank à la page en question, ne seront pas mentionnées.

8.4.4 AdWords

Etant donné la stratégie de Coupole, certaines thématiques sont traitées sur le site *Lexpress.fr* qui ne sont pas, ou peu, associées en général au titre papier *L'Express* : la finance, l'industrie, les PME, le cinéma, la littérature, la mode masculine et la mode féminine. Sur ces thématiques, la part que représentent les accès directs dans le trafic global peut être plus faible que pour les sujets d'information politique et générale. Par ailleurs, nous avons vu avec le cas de *Studio Ciné Live* que sur une thématique comme le cinéma, les « slots » des listes générées par Google Search étaient envahis par les liens vers des services présents depuis plus de dix ans sur la toile comme *Allocine.com* et *Evene.fr*. C'est donc pour positionner la marque *Lexpress.fr* sur les sujets qui ne sont pas associés au magazine *L'Express*, que Roularta Media France a recours aux liens sponsorisés. :

« On va travailler avec la mise en avant de la marque sur des sujets qu'on ne travaillait pas avant ou sur lesquels on n'est pas perçus comme étant légitimes. Par exemple, mettons l'économie : alors qu'on attend plus de la politique comme cœur de métier de L'Express, nous allons utiliser AdWords pour dire : "Regardez L'Express, c'est aussi l'économie". C'est L'Expansion, mais en fait c'est aussi L'Express. Adwords va me rapporter de la visibilité et me faire du trafic, qu'ensuite je vais essayer de fidéliser sur ces thématiques. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bougué

Roularta Media France se saisit de cette marge de manœuvres AdWords dans le but, à moyen et long termes, de pouvoir élargir sur le web l'image de marque héritée du magazine papier *L'Express* en la « couplant » avec les images des marques *L'Expansion*, *Studio Ciné Live*, *Lire*, *Styles* et *Tendances*.

Lorsque nous avons demandé au Directeur de l'acquisition d'audience s'il était déjà arrivé qu'un journaliste lui demande, parce qu'il était particulièrement content

d'un article en particulier, de pousser ce dernier dans les résultats de Google grâce aux liens sponsorisés, nous avons obtenu la réponse suivante :

« Oui, mais à ce moment je demande au journaliste : « Quelle est la valeur générée par ton article ? » C'est parfois dur à entendre pour les journalistes. Ce n'est pas dur à comprendre, mais c'est dur à entendre. Un papier intéressant qui me fait une page vue, je vais avoir du mal à trouver de la rentabilité dessus, donc je ne vais pas y aller en AdWords. Par contre si tu me fais un dossier sur le sujet, qui génère 5 pages vues et de la fidélisation parce que tu as prouvé que tu étais un expert sur le sujet, là on y va sur du Adwords, pourquoi pas. On essaye d'être transparents avec les journalistes. Tout ça c'est mesuré, c'est pas du doigt mouillé : il y a tel volume de recherches sur la thématique, okay on y va ou on n'y va pas. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bougué

Cette dernière citation nous permet d'introduire le chapitre suivant. Celui-ci traitera de l'impact des stratégies de SEO sur le journaliste, c'est-à-dire sur les pratiques de mise en récit de l'actualité mais aussi sur les représentations et le quotidien de ceux dont c'est le métier, leurs valeurs, leurs frustrations, leurs satisfactions, les détours que leurs scripts peuvent subir, les limites qu'ils ne veulent pas franchir et la manière dont ces éléments peuvent peser sur le processus d'infomédiation et influencer son résultat.

CHAPITRE 9 – LES JOURNALISTES ET LE JOURNALISME CHEZ RMF

« [Les journalistes] demeurent encore de véritables référents informationnels pour nombre d'usagers de l'internet. Ce, en raison de leur savoir-faire en matière de traitement de l'information, mais aussi de leur légitimité sociale, construite par une histoire multiséculaire. D'autres boussoles plus récentes, telles que les moteurs de recherche, ne bénéficient pas d'un tel avantage acquis : leur représentation sociale dominante renvoie davantage à un outil performant agissant dans le présent, plutôt qu'à une communauté humaine façonnée par une histoire longue. Et pourtant, ces deux types de boussoles sont appelées à converger : d'un côté, on se trouve en présence d'un groupe social et professionnel auquel les théoriciens critiques reprochent un fonctionnement de plus en plus machinique et machinal. De l'autre, nous avons affaire à un robot que nombre d'acteurs issus des sphères marchandes et savantes tentent de sophistiquer pour le rendre de plus en plus "intelligent". »

(Pélissier et Diallo, 2009, p. 71-72)

Notre but ici sera de voir comment les actions des journalistes se coordonnent avec l'action des référenceurs dans un *rapport dialectique*, c'est-à-dire un « rapport reposant sur le principe de tension-opposition entre deux termes, deux situations, et dépassant cette opposition » (CNRTL). Nous tâcherons de décrire à la fois la manière dont les journalistes agissent au quotidien mais aussi comment ils se représentent leurs actions vis-à-vis du dispositif et ce que cela implique.

Il y a dix ans, dix-sept chercheurs français spécialistes de l'étude du journalisme et des entreprises de presse cosignaient un texte intitulé « *L'information en-ligne : un nouveau paradigme pour le journalisme ?* ». Ils y expliquaient qu'une réorientation stratégique des groupes de presse avait conduit les journalistes « soit à voir le pouvoir dans les organisations leur échapper au profit des "*news comers*" issus de service commerciaux, soit à s'investir eux-mêmes dans des tâches à visées publicitaires et relationnelles, se transformant alors en "*hommes-orchestre*" exerçant

simultanément des activités d'information *et* de communication au profit du groupe de presse » (Pélissier *et al.*, 2003³⁰⁴). Les questions évoquées ici du pouvoir des journalistes dans l'organisation et de l'exercice simultané d'activités d'information et de communication nous intéressent particulièrement pour ce chapitre. Nous sonderons le « pouvoir faire » des journalistes : quel est-il ? pourquoi et comment pourrait-il leur échapper ? que font les journalistes pour conserver leur pouvoir ? Nous nous intéresserons également à l'arbitrage effectué entre les activités d'information et de communication, et à la manière dont il est justifié. Nous nous plaçons pour cela dans la filiation d'une pensée développée depuis longtemps dans les sciences sociales, selon laquelle la routine des journalistes, leur cadre de travail, leur environnement culturel et technologique et leurs valeurs jouent un rôle crucial dans leurs décisions, leurs choix et, finalement, dans la construction et la circulation du fait d'actualité (Breed, 1955 ; Soloski, 1989 ; Shoemaker et Reese, 1996). Nous porterons une attention particulière à ce que devrait être selon les journalistes un traitement idéal de l'actualité, dans le but de comprendre comment une telle représentation peut être conciliée, et à quel point, avec la stratégie de référencement. Ainsi, nous pourrions décrire la mise en tension entre ce que les impératifs d'audience peuvent dicter aux journalistes et ce qui, selon les journalistes, devrait être le traitement de l'actualité.

Analysant les stratégies de SEO mises en place au Royaume-Uni par *BBC News*, *The Guardian* et *Northcliffe*, Murray Dick a effectué un état de l'art des travaux concernant l'impact de la « course aux clics » sur la ligne éditoriale des éditeurs de presse en ligne (Dick, 2011). Il en ressort que les motivations d'ordre économique peuvent être contrebalancées par les valeurs héritées du journalisme traditionnel et que les webjournalistes peuvent se montrer prudents à l'égard des actions que les outils de mesure et de qualification sont susceptibles d'induire (MacGregor, 2007). Murray Dick cite toutefois les travaux de Peter Lee-Wright (2010), qui suggèrent, dans le cas de *BBC News Online*, que la course aux clics puisse exercer une pression sur la priorité accordée aux événements. Il semble en effet, selon Lee-Wright, qu'il y ait une différence notable entre les sujets qui intéressent

³⁰⁴ Nous n'avons eu accès qu'à une version numérique de ce texte. C'est pourquoi le numéro de page ne figure pas.

l'audience et les sujets importants selon les journalistes. Andrew Currah, également cité par Murray Dick, rend lui aussi compte d'une « tension grandissante entre les valeurs éditoriales et la connaissance de ce qui génère effectivement des revenus » (Currah, 2009, p. 48). D'après ses observations, certains éditeurs pensent que les outils de mesure et de qualification du trafic permettent d'améliorer la qualité du contenu journalistique dans la mesure où cela permet de répondre au besoin exprimé par les lecteurs, tandis que d'autres éditeurs craignent que les sujets « populistes » et « *click-friendly* » soient systématiquement avantagés. D'après Currah, cela peut conduire à étioler la réputation de la marque et la spécificité de son audience. Cependant, si ces travaux parlent des effets de la popularité de certains sujets et des pressions concurrentielles sur la ligne éditoriale, ils ne considèrent pas en particulier les effets du SEO. C'est pourquoi Dick Murray a souhaité s'intéresser aux stratégies de référencement, et a montré lors d'une étude exploratoire qu'elles pouvaient avoir une influence sur les lignes éditoriales, sans pour autant que cette influence ne soit « dogmatique » ni ne remette en cause les valeurs des journalistes (Murray, 2011). Mais Murray évoque finalement très peu les journalistes, leurs pratiques, leurs négociations avec le référenceur et les tensions auxquelles la stratégie de SEO peut aboutir. Il n'interroge qu'un seul journaliste sous couvert d'anonymat, dont on ignore pour quel site il travaille. Dès lors, nos observations auront pour but de compléter les siennes, parfois de les nuancer étant donné des différences entre notre terrain et le sien, et d'y apporter des éléments qui n'ont pas été révélés.

Le terrain d'investigation de chapitre est composé exclusivement des entretiens conduits auprès des employés et des dirigeants de Roularta Media France, et de nos observations du site Lexpress.fr, car il nous a semblé que pour suivre les acteurs et pour analyser la coordination de leurs scripts, la juxtaposition de leurs représentations et de leurs intérêts respectifs, ainsi que les frictions et les négociations entre eux, l'approche monographique serait préférable. Le niveau *micro* nous a permis en effet de comprendre de manière approfondie comment la stratégie de référencement était perçue, pratiquée et négociée au sein d'une entreprise en particulier. Le niveau *mezzo* adopté dans les deux chapitres précédents ne nous aurait pas permis d'aller suffisamment loin dans ce chapitre en particulier.

Notre raisonnement ne devra donc être considéré comme valable que pour le cas des journalistes œuvrant pour le site et les sous-domaines du pôle Information de Roularta Media France. Si nous voulions en tirer des conclusions plus générales, d'autres études monographiques devraient être menées et rapprochées de la nôtre et de celle du Murray Dick. C'est pourquoi nous abordons la question de la pratique et des représentations journalistiques dans ce tout dernier chapitre, qui a l'originalité de constituer à la fois un apport empirique central de notre thèse et une base pour des travaux futurs.

9.1 Le pouvoir faire des journalistes

Dans cette sous-partie, nous expliquons quel est le « pouvoir faire » des journalistes de RMF, lié à un « savoir faire », c'est-à-dire à une connaissance des pratiques en ce qui concerne aussi bien l'écriture journalistique que le SEO. Il s'agit de comprendre à la fois ce que les journalistes savent faire et font pour optimiser le référencement de leurs articles, et ce qu'ils font au quotidien pour conjuguer la question du référencement avec leur vision du métier et de ce qu'est à leur avis un traitement de l'actualité *de qualité*. Nous verrons que certaines bonnes pratiques SEO sont conciliables avec les règles de l'écriture journalistique. Mais nous verrons également que les journalistes peuvent être conduits à faire des concessions, et qu'ils peuvent parfois refuser d'agir comme le leur recommande le référenceur.

9.1.1 Pyramide inversée

Les journalistes, quand ils parlent de l'écriture web, ou de « l'écriture pour Google », évoquent toujours les mêmes points : le choix des sujets, le choix des mots du titre, le choix des mots du chapeau, le positionnement des liens hypertextes, la mise à jour des articles. Parfois, et nous y reviendrons, ils peuvent évoquer le choix des angles de traitement, mais c'est plus rare. Finalement, ces points recouvrent l'ensemble de leur champ de manœuvres textuel et hypertextuel, mais jamais le champ architextuel, dont ils ne se saisissent quasiment pas, ou pas du tout.

Avant de parler du choix des sujets traités, du choix des mots, des angles et de la stratégie de maillage, nous évoquerons ici ce qui venait généralement en premier

quand nous interrogeons les journalistes : le caractère informatif, factuel et exhaustif de leur production, et cela dès les premiers mots de l'article.

D'après les journalistes interrogés, écrire pour Google consiste en premier lieu à rédiger des titres et des chapeaux informatifs où les mots-clés apparaîtront le plus tôt possible. Il s'agit d'éviter les jeux de mots ou les ambiguïtés, de manière à ce que le crawler identifie tout de suite le sujet d'une page. Cela n'est pas inconciliable avec les règles enseignées dans les écoles de journalisme, au contraire (Niles, 2010). De telles exigences rejoignent notamment le procédé de la « pyramide inversée », qui consiste à donner l'essentiel de l'information dès le début du texte.

« On essaye de respecter systématiquement la pyramide inversée. L'essentiel de l'info est dans le titre et le chapeau, mais aussi dans le papier. Il faut que le papier soit un tout en lui-même, et le titre aussi. Ça c'est une bonne pratique Google, mais c'est aussi, et surtout, une bonne pratique journalistique. »

Lexpress.fr — Rédacteur en chef adjoint — Thomas Bronnec

Dans un ouvrage intitulé *L'écriture journalistique*, le journaliste Jacques Mouriquand explique qu'« un titre perçu dans une phase de lecture rapide, voire distraite doit être parfaitement clair. L'idée très répandue selon laquelle un effet d'intrigue encouragerait à la lecture est totalement fausse [...] Le souci de clarté interdit aussi de recourir à des mots polysémiques (plus...) qui peuvent donner un sens contradictoire au titre. De même pour les mots pédants, conceptuels ou pour les signes peu répandus » (Mouriquand, 2011, p. 106). Jacques Mouriquand destine son livre autant aux journalistes de presse écrite qu'aux journalistes du web, de radio ou de télévision. Il ne mentionne pas le SEO, et ce même si ses conseils vont dans le sens de ce que Google recommande. A propos du chapeau d'un article, il explique qu'il doit être « à la fois un résumé de l'article et une incitation à la lecture. Il donne l'information de façon sèche et comporte quelques formules heureuses qui suggèrent tout l'intérêt de l'article qui s'annonce. L'écriture doit être dynamique ; les mots forts, vivants et concrets, privilégiés » (*op. cit.*, p. 112). Nous avons clairement retrouvé ces éléments lors de nos entretiens :

« C'est un peu similaire à n'importe quel travail journalistique. Il faut que le lecteur arrive sur la page article, et qu'il ait toutes les informations dès le titre et le chapeau. Donc je ne fais pas des titres informatifs tant pour Google, mais après, bien sûr, je le fais aussi pour les moteurs de recherche. Parce qu'on sait très bien que les moteurs c'est hyper important dans notre trafic, et j'ai envie que mes papiers soient lu ; donc quand même je fais gaffe à ça, mais c'est plus par souci de bien faire mon boulot que par souci de Google [...] Pour le chapeau, on fait comme les journalistes de presse écrite : on essaye de mettre les 5 w. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

« On nous a montré l'importance des mots-clés dans les titres et des chapeaux, presque autant les chapeaux que les titres. On nous a dit de faire attention, si on met « ivoirien » dans le titre par exemple, à mettre « côte d'ivoire » en entier dans le chapeau. Ce sont de petits détails comme ça. Ça devient une habitude assez rapidement. Ça ne me pèse pas. Pour le reste de l'article, dès que les premières lignes sont rédigées, je fais ce que je veux. »

Lexpress.fr — Journaliste (Monde) — Catherine Goueset

Certains journalistes ont pu nous dire qu'ils se sentaient contraints dans leur créativité, toutefois selon eux, comme pour les autres, il est impensable de rédiger autrement. Ils perçoivent cela comme une règle semblable à celle des « 5W » (*Who, What, When, Where, Why*), conformément à ce qu'a préconisé dans un article de la *British Journalism Review* l'Editeur Shane Richmond (Richmond, 2008, p.53). Il ne s'agit donc pas du souhait de pouvoir jouir de davantage de créativité, mais d'une constatation de certaines différences notables entre le web et le papier :

« Pour l'accroche, je me sens contrainte parfois. Chez nous, ça va être une phrase ou deux mots. Alors que j'adore lire les longues phrases, parfois un paragraphe entier de mise en contexte dans les magazines. J'adore le lire, mais chez nous ça n'aurait aucun sens ! Au premier paragraphe, dans ce cas, tu perds le lecteur. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Au sujet de la mise en contexte évoquée ci-avant, étant donné la part de lecteurs non fidèles, les journalistes nous ont expliqué qu'ils effectuaient un effort constant de contextualisation. Cela n'est pas lié à la maximisation du trafic provenant de Google, mais à un souci purement journalistique d'informer au mieux les internautes, nombreux, arrivés par le biais d'un infomédiaire, quel qu'il soit :

« Nous touchons à un public très large. Les gens arrivent sur notre site par des voies différentes, et notamment par Google mais aussi Facebook, Twitter, etc. Donc sur le web tu as cette contrainte de recontextualiser, de répéter, parce que les gens qui arrivent sur ton site n'ont peut-être jamais entendu parler de ce sujet-là ni même peut-être venus une seule fois sur ton site. Il faut leur redonner toutes les balles à chaque fois. C'est pour ça aussi qu'on est très informatifs, très factuels. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Ainsi, le caractère informatif, et cela dès les premières lignes, est un point de convergence dans la rencontre des scripts des concepteurs, du référenceur et des journalistes. Cependant, nous allons voir qu'il existe d'autres zones sujettes à certaines frictions, obligeant les scripts à être révisés.

9.1.2 Choix des sujets

La ligne éditoriale de Lexpress.fr est négociée entre plusieurs acteurs. Le rédacteur en chef du site propose des sujets aux journalistes et valident les sujets que ceux-ci lui proposent. Les rédacteurs en chef adjoints l'assistent dans cette tâche. Le rédacteur en chef du titre papier a quant à lui un rôle plus flou. Chez RMF, Christophe Barbier est théoriquement le supérieur hiérarchique d'Eric Mettout, même si ce dernier jouit d'une indépendance assez marquée vis-à-vis de la rédaction papier, ce qui nous a été confirmé lors de nos entretiens aussi bien par les journalistes du magazine que par ceux du site web. Ainsi, la ligne éditoriale est négociée entre plusieurs personnes, selon les intérêts et les représentations de chacun, sans pour autant que chacun exerce le même *pouvoir* à son endroit. Les journalistes utilisent également des outils de mesures d'audience afin de voir quels sont les sujets les plus prisés par les internautes et, éventuellement, afin de traiter davantage, ou de traiter tout court, ces sujets en particulier. Ainsi, la médiation permettant aux acteurs de se coordonner n'est pas seulement sociale, mais sociotechnique : actants humains et non humains — dont les capacités, les valeurs et l'autorité sont diverses — s'associent en un même réseau d'actions dont le résultat sera la ligne éditoriale, non pas comme elle a été pensée initialement par un seul d'entre eux, mais comme elle a été *négociée* par et parmi des éléments hétérogènes, lors d'une interaction qui n'est ni isotropique, ni synchronique, ni synoptique, ni homogène, ni isobarique (Latour, 2007). Finalement,

la ligne éditoriale est performée par l'énoncé du dispositif, autant qu'elle cherche à performer le dispositif.

La firme Google peut jouer un rôle dans le choix des sujets, *via* notamment le service corollaire au moteur « Google Tendances de recherche ». Cet outil donne, comme nous l'avons expliqué (cf. 4.3.8), des informations à propos des requêtes effectuées par les internautes sur Google Search. Il peut donc être utilisé dans le but de savoir quels sujets intéressent les utilisateurs du moteur. Le rédacteur en chef et le rédacteur en chef adjoint de *lexpress.fr* nous ont tous deux expliqué qu'un tel outil pouvait être prescripteur, et qu'ils ne voyaient pas là une pratique forcément contraire à l'idée d'un journalisme de qualité :

« Qu'un sujet apparaisse dans Google Insight [i.e. Google Tendances de Recherche], c'est déjà une information : il y a des trucs qu'on a vu passer en se disant que ce n'était pas très intéressant, et en fait si, ça devient intéressant parce que les gens s'y intéressent. [...] Ce n'est donc pas un problème seulement marketing. Ça veut dire aussi que nos sources d'infos elles sont là. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Ça peut arriver que les Insights soient prescripteurs de sujets. On ne le fait pas assez je trouve. On devrait le faire davantage. Se renseigner sur ce que cherchent les internautes, pour savoir quelles sont leurs attentes, mais on a le nez sur le guidon. Il ne faut pas que [le référenceur] hésite à venir nous voir. »

Lexpress.fr — Rédacteur en chef adjoint — Thomas Bronnec

Ce dernier verbatim évoque le rôle du référenceur dans la définition de la ligne éditoriale. Celui-ci peut en effet se faire le porte-parole des utilisateurs de Google auprès de la rédaction, en utilisant lui-même « Tendances de Recherche », mobilisé comme une *inscription* visant à *performer*.

*« Utilisez-vous « Tendances de Recherche » pour prévenir les journalistes qu'il faut traiter un sujet en particulier ?
[Le référenceur] le fait beaucoup, et moi aussi. On prévient qu'un sujet est beaucoup cherché quand on ne l'a pas traité. Je fais de la veille là-dessus. Mais à un moment, on ne peut pas se multiplier, c'est un peu chaud vu le nombre de sites qu'on a pour les ressources. C'est pour ça qu'on sensibilise régulièrement les journalistes, on les forme, pour qu'ils sachent utiliser Insights et que ça devienne un réflexe. On essaye de l'intégrer de plus en plus dans le CMS, et de pousser ces données en amont dans leur travail. Parce que ça fait partie de leurs sources. Mécaniquement, je pense qu'il n'y*

a rien de différent avec le boulot que pouvait faire le journaliste il y a trente ans. C'est juste que la multiplication des sources a compliqué leur tâche. Le bruit fait par ces sources fait qu'il devra y avoir plus d'analyse pour déterminer le bon truc à prendre, et l'immédiateté de la réponse, que ce soit à travers le trafic que tu vas générer sur ton article ou les commentaires positifs ou négatifs, et ça les fait tous un peu flipper. Mais objectivement les mécaniques sont les mêmes : il y a un journaliste, des sources, et des lecteurs. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bouguié

Nous apprenons plusieurs éléments ici. D'une part, le directeur de l'acquisition d'audience et le référenceur sont clairement force de proposition quant aux sujets traités. Cela diffère des observations de Murray Dick, d'après lesquelles le choix des sujets continuerait à être indépendant du succès de certains sujets sur les moteurs (Dick, 2011, p. 467). Ainsi, par l'entremise de la Direction de l'acquisition d'audience et de « Tendances de Recherche », le dispositif au sens large — c'est-à-dire ceux qui le conçoivent, ceux qui l'utilisent et *ce* qui fait qu'il fonctionne d'une certaine façon — peut orienter, c'est-à-dire *performer*, la ligne éditoriale de *Lexpress.fr*. Le but affiché par Marc Bouguié est de réussir à faire en sorte que le recours à « Tendances de Recherche » par les journalistes se systématiser, de façon à ne plus avoir à passer par l'entremise de ses services. La citation nous apprend également que des innovations sont intégrées dans le CMS de publication, de manière à ce que les journalistes n'aient pas à se rendre sur le site dédié à « Tendances de Recherche » mais disposent du service à l'intérieur de leur propre outil, de manière à provoquer leur intérêt, leur adhésion. Ainsi, « Tendances de Recherche » est injecté dans l'outil de publication des journalistes, où l'énoncé qu'il produira sera à même de performer d'autant mieux. Enfin, il est intéressant de remarquer que le Directeur de l'acquisition d'audience donne son avis sur le travail des journalistes, en indiquant que celui-ci n'a pas changé avec le web, et qu'il s'agit pour les journalistes aujourd'hui, comme pour celui d'hier, de répondre aux attentes des lecteurs. Dès lors, « Tendances de Recherche » est présenté comme une source d'information, non sans une certaine confusion entre les sources des journalistes au sens classique du terme, d'où ils tirent des informations concernant l'actualité, et les sources du marketing, d'où sont tirées des informations concernant ce qui intéresse le lectorat et son profil.

Le référenceur nous a expliqué quant à lui qu'il assistait autant que possible aux conférences de rédaction dans le but de proposer des sujets, soit à partir de « Tendances de Recherche », soit à partir de Google Actualités directement. Dans ce cas, il regarde quels sujets remontent dans les grappes hypertextes afin de voir si, oui ou non, ceux-ci ont été traités, et de suggérer le cas échéant à la rédaction de le traiter, au moins sous forme de brève, dans le but d'attirer un trafic supplémentaire.

« J'essaye d'aller aux conférences de rédaction hebdomadaires. Pourquoi ? Pour leur donner des pistes, des sujets. J'ai vu que tel ou tel mot-clé était passé dans l'actu, est-ce que vous l'avez vu ? J'essaye de donner des sujets sans l'imposer, mais de dire aux journalistes : « Tiens, il y a tel sujet qui sort. Il y a ça qui va être chaud dans deux semaines. » Des petites choses qui sont recherchées par les internautes et que les journalistes ne vont pas forcément avoir en tête. [...] Je regarde Tendances de Recherche, pour qu'on se positionne sur le moteur, ou bien je regarde directement Google Actu, pour voir quels sont les clusters qui buzzent, et si je vois qu'on n'y est pas, alors je préviens les journalistes.[...] Par exemple, j'écris un calendrier de la coupe du monde de rugby, je le montre au journaliste et je lui dis "ça m'a pris dix minutes à faire et ça fait énormément de pages vues", alors que ces pages vues il les fera en écrivant dix articles classiques. Et là en dix minutes, il a son truc. Il n'y a aucune valeur éditoriale, okay, mais ça vaut tout de même le coût de le faire. Alors voilà, j'essaye de leur montrer ça. Si tu fais un papier comme ça, qui va te faire le gros de ton trafic sur la semaine, après le journaliste il peut s'éclater, et qu'il me fasse des titres optimisés moi je m'en fous à la rigueur, car son quota est rempli : "T'es visible sur ta rubrique avec tel article". »

Lexpress.fr — Référenceur — Renaud Perrin

Pour intéresser le journaliste, le référenceur lui explique que traiter certains sujets ne lui coûtera pas beaucoup de temps, et que cela pourra même à la rigueur n'avoir « aucune valeur éditoriale », mais qu'il pourra ainsi « remplir son quota » et être libéré des exigences liées au SEO. Quand le référenceur dit que « ça vaut tout de même le coût de le faire », nous avons affaire à la question du « prix à payer » pour le journaliste. Il lui est suggéré de traiter certains sujets qui ne seront peut-être pas considérés comme intéressants, notamment dans la mesure où le journaliste n'aura rien à ajouter à ce qui est déjà écrit sur d'autres sites, mais d'accepter cette concession dans le but précisément de ne plus avoir à en faire ensuite. En assistant aux conférences de rédaction, et cela même s'il nous a dit y rester en retrait, le référenceur pénètre un lieu d'habitude exclusivement réservé aux journalistes, où la

ligne éditoriale — les sujets, les angles, la répartition du travail de collecte et de traitement — est déterminée. Cela rejoint les observations de Murray Dick, qui explique que le référenceur du *Guardian* assiste lui-aussi aux conférences de rédaction, où il peut proposer des sujets (Dick, 2011, p. 472). Cependant, le référenceur du *Guardian* explique qu'il ne propose jamais un sujet qui n'a pas de valeur éditoriale (*ibid.*), ce qui, nous le voyons, diffère des propos que nous avons nous-mêmes recueillis.

Les journalistes sont réticents à l'idée que la ligne éditoriale puisse être dictée par quelqu'un qui n'est pas journaliste. Ainsi, en se posant comme un intermédiaire entre les utilisateurs de Google, dont l'intérêt est médié par « Tendances de Recherche », et la rédaction, le référenceur peut desservir sa cause, s'il ne réussit pas à traduire sa demande dans un langage qui parlera aux journalistes³⁰⁵.

« [Le référenceur] demande parfois si on a traité tel ou tel sujet. On lui dit que oui, on a fait un encadré dans la dépêche précédente, mais il nous dit que ce serait bien de le faire à part, et on le refait à part. On avait fait une ligne, et on va faire une dépêche. Faire une actu parce que tout le monde l'a faite, moi je trouve ça débile. Et c'est un peu ça, la logique Google. C'est étrange que ce soit un non-journaliste qui nous dise que ce serait bien de faire tel ou tel traitement sur tel ou tel sujet. Là, par exemple, je fais un papier sur le meurtre d'Agnès et la présomption d'innocence qui n'a pas été respectée. S'il vient et qu'il me dit : "ce serait bien de faire un papier sur le fait que Guéant ait eu raison d'intervenir au JT de TF1", je ne vais pas lui dire : "Ouais, d'accord, je vais faire un papier à part pour expliquer ça". Quand c'est pour une dépêche AFP, okay, mais s'il intervenait sur le sens de mon travail, là je lui dirais non. »

L'express.fr — Journaliste (Société) — Julie Saulnier

Dans ce verbatim, nous voyons qu'il y a bien négociation quant aux sujets traités, et que les journalistes ne sont pas prêts à céder sur tous les points. Par ailleurs, la journaliste explique qu'elle trouve étrange le fait qu'un non-journaliste se mêle du choix des sujets. Ainsi, elle marque une frontière entre ce qui relève de son rôle et ce qui relève du rôle de ceux qui ne sont pas journalistes, en particulier le référenceur, et ce même si le SEO comprend une composante éditoriale. Finalement, en l'absence d'ascendant hiérarchique du référenceur, les journalistes

³⁰⁵ Cela a également été remarqué par Murray Dick au *Guardian* : « [Le référenceur] est conscient qu'en proposant des sujets jugés non pertinents, il pourrait cesser d'intéresser ses partenaires et donc perdre du pouvoir au sein de l'organisation » (Dick, 2011, p. 472).

sont libres d'écouter ou non ses conseils et donc d'adhérer ou non à la stratégie qu'il aura définie. Dès lors qu'il s'agit de prise de position éditoriale, ils ont le dernier mot dans la négociation. Il n'y a pas *performance* sans leur *adhésion*. Interrogé à ce sujet, l'un d'entre eux nous a expliqué qu'il s'agissait de se faire « sa propre religion » :

« Si tu ne fais pas attention, tu te retrouves vite à traiter des sujets dont Google veut parler, parce que Google, quelque part, correspond à ce que les internautes veulent, donc il va pousser des papiers qui sont recherchés par les internautes, et donc toi tu vas suivre ça mais finalement c'est une espèce de cercle vicieux où tu ne fais que parler des sujets qui intéressent les internautes. [...] Quand un sujet est en home de Google News et que tu ne l'as pas traité, après c'est à toi aussi de te faire ta religion là-dessus, et de te dire : "Je ne l'ai pas traité, mais est-ce que ça mérite que je le traite ?". Ça arrive de tomber sur un sujet qu'on avait complètement zappé, et alors on se dit merde, c'est con que je l'ai zappé, et on le fait. Donc je le regarde plus pour un côté veille, mais encore une fois ce n'est pas systématique...»

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Il est intéressant de constater dans ce verbatim que les souhaits des « utilisateurs de Google », dont l'outil « Tendances de recherche » rend compte, se confondent dans les propos du journalistes avec les souhaits des internautes en général. Une montée en généralité est opérée, par laquelle les requêtes effectués sur Google, puis interprétées et partiellement communiquées par « Tendances de recherche », deviennent : « les sujets qui intéressent les internautes ».

Il arrive que les journalistes se sentent « obligés » d'être présents sur certains sujets, et ce même s'ils n'apparaîtront pas dans Google. Le traitement de ces sujets peut ainsi être décidé indépendamment de toute logique de référencement, ce qui s'accorde avec les observations de Murray Dick. Dans ce cas, le programme d'action des journalistes échappe entièrement à Google. Le dispositif pourra référencer et mettre en avant ces contenus en particulier, mais le sujet de ceux-là n'en aura pas moins été choisi indépendamment de leur potentiel SEO.

« J'ai l'impression qu'il y a plus de contraintes qu'avant, mais ce n'est pas Google qui nous dit quoi faire. On a quand même une ligne éditoriale assez claire. On ne se dit pas : "cet article, il ne marchera pas sur Google, donc on ne va pas le faire". Il y a des trucs sur lesquels on est obligés d'être présents, donc que ce soit sexy ou non, que ça plaise à Google ou pas, on les fait, simplement parce que c'est important pour nous.[...] Il y a des moments de la journée où tu sais que si tu fais le bon papier sur le bon sujet tu vas

remonter dans Google Actu. Heureusement, on ne fait pas ça au quotidien. Mais on l'a fait une fois, pour des tests. C'est un peu opportuniste : tu profites des robots pour placer tes papiers. Quelque part, ça change profondément le boulot, parce que le fait qu'on te dise : "regarde mieux l'actu et fais un papier là dessus", c'est inquiétant. Et à terme, c'est clair, on va dans ce sens là. »

Lentreprise.lexpress.fr — Journaliste (Entreprise) — Emmanuel Colombié

Le journaliste exprime ici une vraie inquiétude quant à son indépendance en matière de choix des sujets. Il pense que les visées stratégiques liées au référencement pourraient finir par l'obliger à ne plus traiter que les sujets susceptibles d'apparaître en haut des classements de Google, et qui ne sont pas forcément « dignes de devenir événement » (Mercier, 2009, p. 20). Cependant, et tous les journalistes interrogés ont été clairs là-dessus, ce n'est pas le cas pour l'instant. En revanche, le fait que certains sujets soient traités indépendamment des considérations liées au référencement, et que d'autres soient traités exclusivement étant donné ces considérations risquerait selon certains de rendre plus floue la ligne éditoriale en étioquant la cohérence d'ensemble des choix effectués :

« Avec un système comme on a aujourd'hui, où on est dépendant de Google et où il n'y a que ça qui nous intéresse, tu ne fidélises pas parce que tu n'as pas assez de qualité. Je pense que L'Express a ce problème de positionnement. Ils veulent qu'on puisse avoir tout. Ils veulent faire de l'info décalée, du buzz, et aussi de la qualité. Du coup tu as un amas d'information qui n'est pas très logique et tu ne sais pas, tu ne sais plus, pourquoi tu vas sur Lexpress.fr. C'est pour ça à mon avis qu'aujourd'hui, sur Lexpress.fr, on y va surtout par Google. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Il serait possible ainsi, selon certains journalistes, que la stratégie qui consiste à « écarteler » la ligne éditoriale entre les sujets perçus comme nécessaires et les sujets « qui n'auraient pas été traités si Google n'existait pas », nourrisse la dépendance du site au référencement. Les lecteurs fidèles pourraient alors ne plus percevoir la ligne éditoriale du titre et finir par se lasser.

Les rédacteurs en chef sont sans aucun doute ceux qui exercent le plus de « pouvoir » en matière de ligne éditoriale. Il s'agit pour eux de déterminer 1) ce qui sera fait exclusivement pour se positionner à l'intérieur du dispositif d'infomédiation, 2) ce qui sera fait de toute façon sans penser au référencement et 3) ce qui sera fait autrement que cela ne l'aurait été si la question du SEO n'était pas entré en jeu.

« Il y a des choix éditoriaux qui ne posent pas problème, car il s'agit de sujets dont on sait qu'ils vont marcher et dont on sait que c'est important, pour nos lecteurs, pour nous, pour Google, pour tout le monde. Puis il y a des sujets dont on sait qu'ils ne vont pas forcément marcher sur Google mais qui sont importants pour nous, donc on les traite quand même. Et puis, on a aussi des sujets où on se dit : "Bon, c'est vrai que c'est pas très important, mais ça va être repris par Google", et donc on va les traiter parce que ça va être apporteur d'audience.[...] Ça nous gêne oui et non. Il y a quand même des limites. On ne va pas faire n'importe quoi. Mais c'est vrai, il y a certains sujets qu'on ne traiterait pas sans Google. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

Le rédacteur en chef de Lexpress.fr Eric Mettout nous a quant à lui expliqué qu'à la différence de la rédaction du titre papier, le rapport était plus étroit entre la rédaction du site web et les équipes du marketing et de la Direction de l'acquisition d'audience. Cependant, le but reste selon lui de réussir à traiter des sujets même s'ils ne « fonctionnent » pas, c'est-à-dire indépendamment de toute logique de référencement et d'audience en général.

« Une grosse différence entre une rédaction traditionnelle et une rédaction web, c'est que nous sommes en contact permanent avec l'équipe marketing, parce qu'ils nous donnent des outils, nous expliquent comment ça marche, et comment nous en servir au mieux, pour être lus. Les limites de l'exercice, et on doit être attentif, c'est de bien se dire qu'on ne va pas traiter un sujet parce qu'on sait qu'il va faire 2000 VU. Non. C'est pas parce que la famine au Soudan n'intéresse pas les Français qu'on ne va pas le traiter. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

Le fait de regarder les réactions de l'audience pour définir ou redéfinir une ligne éditoriale n'est pas neuf en matière de journalisme³⁰⁶, toutefois sur le web comme ailleurs, l'idée selon laquelle la demande crée l'offre doit être mesurée. En effet, nos observations nous permettent sans ambiguïté de conclure que si les

³⁰⁶ « ...souvent suivistes par rapport à ce qui circule déjà dans la société, les journalistes ne sont que rarement en situation de faire advenir dans l'espace public un fait dont personne n'aurait déjà pris conscience » (Mercier, 2009, p. 16).

journalistes RMF acceptent certaines concessions quant à la ligne éditoriale, il n'en demeure pas moins qu'ils continuent de traiter les sujets qui leur semblent importants, que cela soit « rentable » ou non. On assiste non pas à un effacement de la ligne éditoriale — qui ne serait plus dictée alors que par les « Tendances de recherche » intégrées au système de gestion de contenu — mais à une certaine dysharmonie entre d'une part les sujets traités à des fins de maximisation de trafic, pouvant ou non être perçus comme dignes d'intérêt par les journalistes, et, d'autre part, les sujets traités de toute façon, indépendamment de toute considération liée au trafic qu'ils rapporteront. En somme, la performance est *partielle* et *localisée*.

Il est intéressant de noter enfin un point qui contredit les propos des porte-parole de Google. Nous avons vu en effet dans le chapitre 4 (cf. 4.5.4.D) que ceux-là encourageaient les journalistes désireux de maximiser le trafic de leur production à écrire à propos de sujets sur lesquels ils auraient une expertise. Cependant, il ressort de nos entretiens que les journalistes estiment que les articles qu'ils écrivent « principalement pour Google » font l'objet d'un traitement rapide, souvent à partir d'une dépêche AFP, sans forcément y ajouter une quelconque analyse, et qui même peuvent n'avoir « aucun intérêt éditorial ».

Cadre 1. Google, le pluralisme des sources et la diversité des sujets *Retour sur 4 études scientifiques*

Sur le web, le pluralisme des sources et la diversité des sujets existent bel et bien, cependant Google en rend très peu compte. C'est une des raisons pour laquelle la concentration du marché autour de l'activité d'un seul infomédiaire a suscité de nombreuses critiques tout au long des années 2000 (Kawaguchi et Mowshowitz, 2002 ; Diaz, 2008³⁰⁷).

Si tous les éditeurs de presse en ligne choisissaient les sujets selon leur seul potentiel de succès sur Google, il nous paraît évident que nous assisterions à un appauvrissement considérable de la diversité. En effet, l'absence de diversité chez l'infomédiaire pourrait provoquer une absence de diversité chez ceux qui ont besoin de passer par cet infomédiaire pour accéder à leur lectorat. Quant au pluralisme des sources, dans le cas extrême où tous les éditeurs se soucieraient exclusivement de leur référencement sur Google, il n'aurait plus d'intérêt véritable dans la mesure où les sujets traités seraient les mêmes.

Le retour sur 4 travaux scientifiques nous permettra de mieux comprendre quelle peut être le traitement effectué par Google du pluralisme des sources et de la diversité des sujets. Trois de ces quatre études datent quelque peu, car l'utilisation par les concepteurs du dispositif d'algorithmes de personnalisation (Rafiei *et al.*, 2010 ; Goldman, 2010 ; Pariser, 2011) rend de plus en plus compliquée l'observation du pluralisme et de la diversité sur Google. En effet, étant donné la personnalisation, les résultats diffèrent selon les utilisateurs (ainsi, le fait de montrer qu'il y a pluralisme et diversité dans les résultats d'un utilisateur ne permettrait pas de prouver que la pareille est vraie pour les autres utilisateurs).

Etude n°1

Hindman *et al.* (2003) ont montré qu'un petit nombre de sites étaient mis en avant sur Google. Selon ce qu'ils ont nommé la « *Googlearchy* », la diversité des sources présentes sur le web est contrebalancée par le moteur de recherche. « Plutôt que de démocratiser la dissémination de l'information, l'étude de la *googlearchy* suggère que les citoyens pourraient continuer à obtenir des informations concernant la politique provenant de seulement quelques sources, et ce même s'il semble

³⁰⁷ « Comme avec tous les intermédiaires de ce types, nous espérons que les moteurs de recherche rendront accessible l'information de manière impartiale et diversifiée ; nous espérons, en d'autres mots, qu'ils soient « démocratiques ». Nous devrions poser aux moteurs de recherche comme Google la même question que celle que les chercheurs ont posé aux médias traditionnels : les voix sous-représentées et les points de vue différents peuvent-ils être entendus étant donné le filtre des moteurs de recherche ? Quel rôle la publicité joue-t-elle dans les résultats ? Un petit nombre d'acteurs dominent-ils l'industrie ? A la seule condition de répondre à ces questions, nous pourrions juger de la réelle capacité « délibérative » du Web. » (Diaz, 2008, p. 15)

apparemment que les points de vue soient en nombre illimité dans le cyberspace » (*op. cit.*, p. 4). Pour tester cela, les chercheurs ont considéré plus de 3 millions de documents abordant six sujets : l'avortement, le contrôle des armes à feu, la peine de mort, le Président, le Congrès et la Politique. Les trois premiers sujets ont volontairement été choisis parce qu'ils faisaient l'objet de débats houleux, mêlant lobbyistes, associations et partis politiques. Les trois derniers étaient mobilisés dans de nombreux débats sans en être directement les objets. Hindman *et al.* ont expliqué leurs résultats par la dynamique du PageRank, selon laquelle les plus visibles deviennent encore plus visibles. Dans toutes les catégories, la distribution des liens entrants suivait une loi du type « longue traîne » (*power law*). Ils ont montré que « chaque site qui est plus loin qu'à trois clics de distance d'un site du top 200 des résultats de Google ou Yahoo sur un sujet donné est définitivement hors jeu, et peu susceptible d'avoir un impact substantiel sur la politique » (*ibid.*, p. 28). Hindman *et al.* en concluent que « l'architecture ouverte n'est pas une garantie en termes de démocratie. Le nouveau média, comme les anciens, concentre l'attention sur un petit nombre de sources divulguant des informations politiques » (*ibid.*, p. 30). Ces résultats montrent à la fois que le dispositif d'infomédiation ne faisait pas état en 2003 de la diversité des opinions et des sources présentes sur le web, mais également que cela n'était pas seulement imputable à la firme. En effet, c'est parce qu'une page est bien référencée sur Google qu'elle reçoit un trafic important, ce qui conduit de nombreux éditeurs à faire des liens vers cette page, ce qui la rendra plus visible encore. Ce « cercle vicieux » (Diaz, 2008, p. 17) dépend à la fois de l'algorithme paramétré par les concepteurs du moteur et de la manière dont les éditeurs et les internautes auront agi étant donné sa performativité.

Etude n°2

En 2004, Susan Gerhart a montré comment les moteurs de recherche avaient tendance à atténuer les controverses en privilégiant le bon côté des débats. « Les moteurs de recherche ne conspirent pas dans le but de supprimer les controverses, mais leurs stratégies conduisent à la prédominance de sites organisationnels, ce qui prive les chercheurs d'une expérience plus riche et, parfois, d'informations essentielles à la prise de décision » (Gerhart, 2004). Gerhart a pris cinq sujets à propos desquels il existait des controverses. Ainsi, pour le nom de pays « Belize », il était possible d'obtenir des réponses liées aux caractéristiques géologiques, démographiques et touristiques du pays, qui ne faisaient l'objet d'aucun débat, ou bien des réponses mentionnant le conflit frontalier du Belize avec le Guatemala. La plupart des sites identifiés lors de l'expérience ne faisaient pas état de la controverse et provenaient de sources anglaises ou américaines et non du Belize, alors que des sources locales existaient. Gerhart a souligné le rôle joué par les éditeurs, faisant des liens vers les sites les plus reconnus et les plus institutionnalisés,

et le rôle joué par les internautes à la recherche d'informations factuelles. Aussi, selon elle, ce n'est pas « la technologie qui supprime les controverses, mais le social » (ibid.). Susan Gerhart souligne enfin que des éléments à propos des controverses existent bien sur le web, mais qu'ils sont relayés après les 50 premiers résultats de la liste. La solution, selon elle, ne devrait pas venir des concepteurs du moteur, mais des internautes qui devraient prendre conscience du biais produit par les moteurs et de leur prédisposition à présenter le bon côté des choses (*sunny side*). Les utilisateurs de Google devraient par conséquent chercher activement à diversifier les résultats qu'ils obtiennent, pour trouver le « *dark side* », en multipliant leurs requêtes et en utilisant plusieurs moteurs.

Etude n°3

En 2005, Eric Ulken a analysé les réponses données par Google Actualités concernant Georges W. Bush et John Kerry, respectivement candidats du Parti Républicain et du Parti Démocrate lors de la campagne pour les élections présidentielles américaines de 2005. Ses conclusions montrent que Google Actualités pourrait avoir un biais politique. Il est parti des constatations de Lasica (2004), lequel s'étonnait de trouver des résultats systématiquement en faveur de Georges W. Bush et se demandait si « Google n'avait pas été piraté par *FoxNews* ». Ulken a comparé les résultats de Google Actualités à ceux de Yahoo Actualités. Selon ses conclusions, Google Actualités était davantage susceptible de présenter un biais politique que Yahoo Actualités, sans qu'on puisse dire à l'avance quel candidat serait avantagé par ce biais. Il a également montré que le moteur Google Search avait tendance à privilégier les traitements négatifs de la campagne présidentielle, tout en précisant que cela ne pouvait être imputé seulement à la firme, mais également aux éditeurs et aux internautes.

Etude n°4

Au début des années 2010, des chercheurs français rassemblés au sein du programme de recherche *Internet, pluralisme et redondance de l'information* (IPRI) se sont posés la question suivante : « l'arrivée de nouveaux acteurs sur le terrain de l'information [...] entraîne-t-elle une originalité accrue ou au contraire une certaine redondance des nouvelles ? Autrement dit, en matière d'information en ligne, la quantité est-elle synonyme de qualité ? La question du pluralisme de l'information, et de ses enjeux fondamentaux pour la vie démocratique, se trouve ainsi posée à nouveaux frais avec l'internet » (Marty *et al.*, 2012, p. 31). Ces chercheurs ont montré que, même si des sujets et des traitements originaux existaient sur le web, l'agenda médiatique était caractérisé par la « surexposition d'une minorité de sujets d'actualité et [la] confidentialité de la majorité d'entre eux » (ibid., p. 45).

Google Actualités procède ainsi à un « matraquage éditorial » en multipliant les articles autour d'un même sujet, qui ne sont des « sujets rares » que dans 3% des cas. Cela rejoint les observations de Bernhard Rieder (2006, p. 144-146), lequel voit dans Google Actualités une « réalisation mécanique » de la théorie de l'agenda médiatique (McCombs et Shaw, 1972). Ainsi, cette théorie aurait *performé* le dispositif.

A la lumière de nos propres observations et de ces quatre études, nous pouvons nous réjouir que les sujets ne soient pas seulement choisis par les journalistes en fonction de l'apport de trafic qu'ils pourraient générer via Google. En effet, le fait que certains choix éditoriaux puissent être faits indépendamment de la question du référencement nous semble fondamental, car ainsi l'internaute qui le souhaite pourra avoir accès à des sujets et des traitements diversifiés, ainsi qu'au « *dark side* » de sujets portant à controverse — cela s'il se rend directement sur le site *Lexpress.fr*. D'autre part, ces études nous apprennent que pour diversifier ses sources, l'internaute devra également diversifier les voies d'accès à ses sources. La balle est donc dans le camp du citoyen qui, pour se forger une véritable opinion au sens de Stuart Mill³⁰⁸ (1859/1978), devra développer certaines capacités lui permettant de trouver un contenu original auprès de sources différentes.

9.1.3 Choix des mots

Nous avons vu que le choix des mots-clés et leur position étaient pris en compte par le dispositif (cf. 4.2). Cela permet à l'indexeur d'identifier clairement le sujet auquel une page web se réfère et, d'autre part, à l'algorithme d'apparier la requête formulée par un internaute à un résultat susceptible d'être pertinent. Dès lors, le formateur en écriture web et le référenceur de RMF conseillent tous deux aux journalistes de positionner les mots-clés les plus pertinents le plus tôt possible dans la page, c'est-à-dire dans le titre et le chapeau de l'article. C'est la raison pour laquelle on voit fleurir des titres écrits sous la forme « Mot-clé :... », comme cela a déjà été remarqué par Murray Dick (2011, p.469). Nous avons vu en effet dans le chapitre 7 que 69% des articles de *Lexpress.fr* avaient un titre construit ainsi (cf. 7.1.2).

³⁰⁸ « Celui qui ne connaît que sa vision d'un cas connaît peu le cas en question. Ses raisons peuvent être bonnes... mais s'il est incapable de réfuter les raisons du parti adverse, et s'il ne sait pas qui le compose, il n'a pas de quoi préférer une autre opinion... Il devrait écouter les arguments... provenant des personnes qui croient en ces arguments, qui les défendent jusqu'au bout et font tout ce qu'ils peuvent pour eux. Il devrait les connaître dans leur forme la plus plausible et la plus persuasive. » (Mill, 1859/1978, p. 35)

Il s'agit pour les journalistes d'écrire les mots qu'écrivent les utilisateurs des services de Google. Pour cela, ils peuvent consulter le module « Tendances de recherche » injecté dans leur CMS, ou bien agir selon leur jugement personnel, étant donné ce qu'ils pensent à propos des requêtes que formuleront les internautes. C'est une supposition quant à un énoncé potentiel (la requête des internautes) qui permet à celui-ci de performer effectivement l'action des journalistes. Quand nous avons mené notre enquête, les journalistes étaient sensibilisés depuis peu à la question du choix des mots, et au fait de les mettre les plus tôt possible dans le titre et le chapeau.

« Il y a six mois on n'en parlait pas, mais là ils nous on dit : "Maintenant il faut que vous mettiez les mots-clés en début de titre et en début de chapeaux", donc ça c'est un peu contraignant quand même. Mais ça fait partie du boulot maintenant. »

Lentreprise.lexpress.fr — Journaliste (Entreprise) — Emmanuel Colombié

Les journalistes nous ont expliqué qu'ils n'avaient pas le temps de regarder systématiquement « Tendances de recherche » et, ce, même si l'outil était injecté dans le CMS. Ainsi, ils choisissent les mots-clés et la forme du titre selon leur appréciation personnelle. L'un d'eux, en revanche, nous a expliqué qu'il lui arrivait de téléphoner au référenceur pour que celui-ci lui dise sur quel mot-clé se positionner :

*« En fait, moi, je vais appeler Renaud et je vais lui demander : "J'ai un sujet sur les femmes chefs d'entreprise, donne moi un titre stp". Je lui demande carrément. Lui il va me dire : "Bah écoute, on n'est pas très bien positionnés sur "femme chef d'entreprise", donc mets ça". Je ne demande pas systématiquement, mais quand je sens que mon papier a un potentiel, ou bien qu'il a une vraie plus value, je le fais. Ça me prend trois minutes. Je préfère faire ça que de pas avoir de lecteurs. »*³⁰⁹

Lentreprise.lexpress.fr — Journaliste (Entreprise) — Sébastien Pommier

Nous avons identifié un certain flou chez les journalistes quant aux règles concernant le choix des mots. Ce flou est clairement lié au manque de transparence de la part des concepteurs du dispositif, visant précisément à ce que les producteurs ne contenus ne sachent pas comment « tromper » l'algorithme, et cherchant donc à réduire la performativité de ce dernier. En plus de nous renseigner à ce sujet, le

³⁰⁹ Il s'agit en fait ici du titre SEO (voir sous-section suivante). Ce journaliste en particulier, en effet, dédouble son titre et ne se sert des indications du référenceur que pour la partie du « titre SEO », que nous présentons ci-après. Cependant, cela n'enlève rien ici à l'utilisation de ce verbatim comme un argument.

verbatim ci-dessous met en évidence qu'une répétition des mêmes mots-clés entre le titre et le chapeau à des fins de référencement peut altérer ce qu'un journaliste considère comme étant une information de qualité.

« J'ai appris que dans les chapeaux et dans les liens, il faut mettre des mots clé. Mais j'ai l'impression que ça a changé. Au début, il ne fallait pas faire de répétition entre les titres et le chapeau, ce qui correspondait à la règle journalistique : employer un synonyme, par exemple DSK dans le titre, et dans le chapeau « l'ancien patron du FMI », pour varier. Mais maintenant, j'ai l'impression que les chefs veulent qu'on fasse des répétitions à gogo. On met DSK trois fois dans le titre, puis on le remet trois fois dans le chapeau. Ça me dérange vraiment, parce qu'à la lecture, c'est désagréable. Moi, en tant que lectrice, j'adore quand le titre et le chapeau se mettent en couple. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Le Directeur de l'acquisition d'audience, quant à lui, confirme que la manière dont les titres et les chapeaux sont rédigés est fortement imposée par la stratégie de référencement. Et il regrette, donnant à nouveau son avis sur la qualité du travail des journalistes, que cela contribue à formater les titres et à les rendre top informatif plutôt qu'incitatif, originaux, drôles.

« La logique de titraille et de chapeau, par exemple, elle est fortement imposée par Google. Faut du factuel pour Google, et le factuel, c'est chouette, parce que c'est hyper clair, mais c'est pas drôle et c'est pas vendeur pour l'internaute. Si tu traites l'affaire DSK de manière marrante ou d'une manière très politicienne, tu peux le faire sentir dans ton papier à la rigueur, mais pas dans ton titre. L'esprit dans les moteurs Google, il ne ressortira pas. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bougué

La pratique qui consiste à mettre des mots-clés dans le titre n'est cependant pas seulement propre à Google, mais peut constituer une règle que les journalistes s'imposent dans un souci de factualité de la ligne éditoriale, plutôt que dans un souci de marketing : « Les titres informatifs résument l'information sans la moindre fantaisie. Les agences de presse ainsi que certains journaux en ont fait une règle absolue, notamment dans un souci d'image. » (Mouriquand, 2011, p. 108).

Dans la rédaction, les *Front Page Editor* (FPE) repasse sur la plupart des titres et des chapeaux rédigés par les journalistes, dans le but de les harmoniser tant pour Google que par souci de cohérence. Les titres peuvent également être remis à jour.

Le rôle des FPE, qui sont eux aussi journalistes et entièrement intégrés à la rédaction (où nous avons pu constater d'ailleurs qu'ils occupaient une position centrale), consiste ainsi à formuler ou à reformuler les premiers mots qu'un internaute lira. Celle que nous avons interrogée nous a dit que la particularité de son métier résidait dans le fait que les titres qu'elle formulait s'adressaient aussi bien à des humains qu'à des non-humains, c'est-à-dire aux internautes autant qu'aux crawlers de Google.

« Je m'occupe de tout ce qu'on va voir en premier, que je sois un lecteur, une machine, un lecteur de Google News, etc. C'est comme un SR qui va faire la Une d'un journal. Je ne suis plus juste face à quelques lecteurs, je suis aussi face à des robots. Et je suis face à un objet qui est beaucoup plus mobile, c'est-à-dire que mon titre peut bouger dans la journée. Par exemple, quand on fait un suivi en live sur une actu, mon titre va bouger dix ou quinze fois dans la même journée. »

Lexpress.fr — FPE — Camille Caldini

Beaucoup de journalistes nous ont dit regretter l'uniformisation des titres sur le modèle « Mot-clé : ... », ce qui rejoint les regrets exprimés par le journaliste interrogé à ce sujet par Murray Dick (2011, p. 473). Certains se sentent bridés dans leur créativité, et déplorent le fait que leurs titres ressemblent à ceux de leurs concurrents.

« Le titre, le choix des mots, c'est un vrai problème. C'est une contrainte de plus qu'il faut gérer. Mais en même temps, on sait qu'on ne peut pas faire l'impasse sur Google. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

Etant donné sa propre marge de manœuvre, un journaliste, s'il estime être bridé dans sa créativité, peut théoriquement décider de ne plus collaborer avec le référenceur. C'est pourquoi le référenceur de Lexpress.fr a modifié le CMS de publication de manière à mettre toutes chances de son côté pour que puisse s'opérer la *composition* de son propre script avec le script des journalistes. Nous présenterons ci-après cette stratégie de « double-titre ».

Cadre 2. Quand Google couronne la faute journalistique

Dans un article du *Poynter*, la journaliste Kelly McBride a révélé quelles pouvaient être les dérives de l'influence de Google sur le choix des mots (McBride, 2010). Elle a montré comment le traitement journalistique du projet de construction d'un centre musulman à deux pâtés de maison et demi du *Ground Zero* à New York avait pu être erroné par l'usage de « Tendances de Recherche ». Puisque les internautes formulaient la requête « *ground zero mosque* » sur le moteur, de nombreux journalistes utilisaient ce terme à des fins de captation du trafic. Mais l'information donnée par les titres sous-entendait qu'une mosquée avait été construite sur le *Ground Zero*, ce qui était faux puisque la mosquée devait être construite beaucoup plus loin (les deux pâtés de maison étant de tailles considérables) et que le projet, porté par des musulmans modérés, n'était pas seulement de construire une mosquée, mais également un centre ouvert à tous comprenant une piscine, un terrain de basket, une salle de gym, un auditorium de 500 places, une librairie, un studio d'art, un restaurant et une école de cuisine séparés du lieu de culte.

Ce point indique la limite de la stratégie qui consiste à reprendre dans le titre d'un article les termes utilisés par les internautes lors de leurs requêtes, et qui est pourtant plébiscitée par certains journalistes (Richmond, 2008 ; Niles, 2010). En effet, l'objectif de la requête est d'exprimer un besoin d'information, tandis que le but d'un article est de traiter l'information. Le premier est une question, le deuxième une réponse. Dès lors, la stratégie qui consiste à reprendre dans la réponse les termes utilisés dans la question, parfaitement valable du point de vue du SEO, peut conduire à certaines erreurs inacceptables du point de vue journalistique. La question, en *performant* la réponse, la rend inexacte. C'est ce que montre cet exemple très largement commenté en 2010 par les spécialistes de la presse en ligne américaine.

9.1.4 Dédoublage du titre

Le référenceur de Lexpress.fr a mis au point un système de double-titre : les journalistes écrivent un titre pour les lecteurs, qui apparaît effectivement sur la page, encadré dans le code HTML par des balises « H1 », et un titre pour Google, qui apparaît seulement dans les résultats du moteurs et sur l'onglet de navigation, encadré par des balises « Title ». Il s'agit d'un système similaire à celui qui a été observé par Murray Dick sur le site de *BBC News* (Dick, 2011, p. 468). Grâce à ce double-titre, les journalistes peuvent écrire un titre incitatif pour le lecteur, tout en écrivant un titre informatif et optimisé pour le moteur. Ce qui est écrit « pour Google » est ainsi dissocié de ce qui est écrit « pour l'internaute », de manière à ce que, théoriquement,

la stratégie de maximisation du trafic n'empiète pas sur la créativité dont voudraient faire preuve les journalistes. L'effet performatif est *circonscrit*.

« J'ai fait ça pour leur permettre de se lâcher sur le titre édito, tout en faisant un titre hyper optimisé. Okay, ils ont plus de boulot, et pas beaucoup de temps pour le faire, mais en même temps ils ont ce qu'ils veulent. Ils ne peuvent plus se plaindre du fait que l'écriture SEO les contraint. »

Lexpress.fr — Réferenceur — Renaud Perrin

Malgré la présence du titre SEO, nous avons vu à *L'Express* que la plupart des titres des articles étaient tout de même informatifs et conçu pareillement : « Mot-clé :... ». Ainsi, les journalistes ne se saisissent pas, ou peu, de leur pouvoir de faire. Cependant, tous nous ont dit regretter de ne pas utiliser systématiquement le titre SEO, accusant le manque de temps et la surcharge de travail. En revanche, les journalistes utilisent le titre SEO pour les articles dont ils sont particulièrement satisfaits et dont ils veulent par conséquent à la fois soigner le titre pour l'internaute, tout en maximisant les chances pour leur article de circuler et d'être commenté.

« J'utilise le titre SEO quand ce sont des papiers très travaillés, pour lesquels j'ai enquêté plusieurs jours. D'une part, on a envie de signer le titre, et c'est aussi un signal lancé à l'internaute : l'internaute qui va voir un titre avec un jeu de mots, va se dire : ce n'est pas une dépêche, ça va plus loin, c'est autre chose... Et l'on peut maintenant faire cela sans nuire au référencement. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Une des attributions du *Front Page Editor* est de corriger les titres lorsqu'il ne les juge pas appropriés. Le titre SEO permet au FPE à la fois de corriger les titres de manière à les optimiser pour Google, et, d'autre part, de manière à respecter au mieux les critères de ce qu'il juge comme étant un titre de qualité. Dans la citation ci-après, la FPE dit clairement qu'il n'apprécie pas les titres « Mots-clé :... », qu'il trouve trop standards et contraires à ses propres critères de qualité. Ainsi, grâce au titre SEO, il peut contourner le point d'achoppement entre son propre script et le script du référenceur, et adhérer pleinement au projet de ce dernier sans nuire au sien propre.

« J'ai un titre SEO qui par défaut est celui est le titre interne. A partir de là, je peux dédoubler mon titre interne et mon titre SEO. J'ai intérêt à faire d'un côté des titres en "Libye deux points...", "Mort de Kadhafi deux points...", "Ben Laden deux points...", parce que les mots-clés arrivent au début.

C'est clair que ce n'est pas pour le lecteur, parce que ce sont des trucs que je ne supporte pas de voir. Mais ces titres, sur la page de l'article, et sur la page d'accueil du site aussi, je peux les modifier. Au lieu de mettre un titre en deux points "Mort de Kadhafi : Obama se réjouit", que j'aurais mis dans le titre SEO, je vais mettre "Obama se réjouit de la mort de Kadhafi", ce qui est quand même beaucoup mieux. Mes mots-clés sont là aussi, le lecteur comprend de quoi je parle et ce n'est pas moche. Je peux donc faire du pur informatif à l'extérieur et, à l'intérieur, pour les quelques lecteurs qui arriveront par la homepage, de l'incitatif, ou en tout cas un truc mieux que ce qui apparaît à l'extérieur. »

Lexpress.fr — Front Page Editor — Camille Caldini

Il est intéressant de constater dans le verbatim ci-avant que les lecteurs arrivés sur un article par le biais d'un infomédiaire sont traités autrement que les lecteurs arrivés directement par la page d'accueil. Ces derniers ont droit à un titre incitatif, tandis que les premiers ont un titre informatif sous le format « Mot-clé :... » simplement parce qu'ils n'auraient peut-être jamais eu connaissance de l'article si son titre n'avait pas été optimisé, ce qui rejoint les observations de Murray Dick (p. 473). Le dédoublement du titre permet ainsi aux journalistes de « fuir » du dispositif, et de recouvrer leur créativité en matière d'écriture. En fait, le double titre peut dans certains cas être un triple titre, car les journalistes ont également la possibilité de changer le titre de leur article sur la page d'accueil du site. Il y aura alors un titre éditorial (page article), un titre SEO (moteurs + onglet de navigation) et un titre de Une (page d'accueil).

« C'est une bonne trouvaille parce que ça permet de concilier travail éditorial et travail de référencement. Nous avons trois types de titre : titre de Une, titre de l'article et titre SEO. On peut très bien imaginer un titre SEO qui est écrit complètement pour le référencement et un titre de Une qui va être un jeu de mots, parce que c'est le titre qui apparaîtra sur la page d'accueil. [...] On peut s'amuser et faire quand même des titres à la Canard enchaîné en mettant un titre SEO très informatif, un titre de Une très jeu de mot et un titre de page article qui est informatif sans l'être trop non plus. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Les rédacteurs en chef apprécient le titre SEO, qu'il voit comme le fruit d'une volonté de la part du référenceur de respecter la créativité des journalistes :

« Le titre SEO a la vertu de simplifier notre tâche. Dans l'idéal, on peut avoir un titre éditorial et un titre SEO, avec les mots les plus importants dans les 60 premiers critères. Donc ça conduit souvent à des choses pas jolies, avec des « mots clés, deux points, etc. ». C'est sûr que sur une page d'accueil si on avait tous les titres comme ça, ce serait horrible. Heureusement on a une possibilité d'ajustement avec le titre SEO. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

Certains journalistes, cependant, émettent des réserves vis-à-vis du dédoublement qui, à leur avis, est chronophage et ne les délivre pas de faire des titres très informatifs sur les pages article.

« Ça ne nous arrange pas tant que ça, parce que même dans le titre normal il faut mettre les mots-clés au début. Donc finalement, je me demande parfois si ça a vraiment un sens de différencier l'un et l'autre. De temps en temps, on oublie d'ailleurs. A priori, le titre SEO fait remonter plus rapidement l'article, mais je ne sais pas exactement comment ça fonctionne. La différence, c'est que tu peux faire un truc beaucoup plus long dans le titre SEO, parce que tout le monde ne le voit pas. Donc tu peux faire un titre à rallonge, parce qu'on ne le voit pas vraiment. »

Lentreprise.lexpress.fr — Journaliste (Entreprise) — Emmanuel Colombié

Par manque de temps, et parce qu'ils n'y pensent pas forcément, de nombreux journalistes n'utilisent pas le titre SEO. La ligne de fuite reste vacante. C'est sans doute la raison pour laquelle nous avons observé une forte proportion de titres optimisés sous la forme « Mot-clé :... » dans notre échantillon.

« Tu peux faire des titres SEO et des titres éditoriaux. Donc nous c'est vers ça qu'on voudrait arriver, mais on n'a pas le temps de le faire. Je ne le fais pas systématiquement parce que je n'y pense pas et parce que je n'ai pas le temps. [...] Quand je ne me sers pas du titre SEO, c'est-à-dire presque tout le temps, je fais un titre juste pour Google. C'est un problème sur notre home : les titres se ressemblent, et ça c'est vraiment à cause de Google. »

Lexpress.expansion.fr — Journaliste (Economie) — Julie de la Brosse

Les journalistes se servent finalement assez peu du titre SEO. Aucun ne nous a dit qu'il y avait systématiquement recours, et tous nous ont dit qu'ils regrettaient de ne pas l'utiliser plus souvent.

« J'essaye de travailler le titre pour qu'il soit bon sur Google plutôt que de faire un titre SEO. Le titre SEO, je ne le fais pas systématiquement mais je devrais le faire. Les "pourquoi", "comment", "ce que l'on sait", ça marche plutôt bien. Après, il faut mettre les mots clés. Renaud nous a fait des formations pour se servir de Google Insights et comparer deux mots pour savoir lequel marche le mieux, mais j'ai encore du mal avec cette démarche purement SEO. »

Lexpress.expansion.fr — Journaliste (Economie) — Emilie Lévêque

Le Directeur de l'acquisition d'audience, quant à lui, est satisfait de la mise en place du titre SEO et des résultats que le dédoublement a donnés. Il nous a expliqué que Lexpress.fr avait été le premier SIPI en France à mettre en place un tel outil, avant d'avoir été imité par son concurrent *20minutes.fr*. Il nous a également dit qu'il s'agissait d'une stratégie de long terme, et qu'il faudrait, pour des résultats encore plus probants, de convaincre les journalistes en leur montrant que l'usage du titre SEO leur permettrait d'être à la fois plus créatifs et plus visibles.

« Le titre SEO, ça donne vraiment quelque chose. [Les journalistes] ne s'en servent pas tous, mais ça commence, on y vient, et commence à voir les résultats. D'ailleurs, on était les seuls avant, mais 20 minutes vient de le mettre en place. Ça a été un peu long, mais là on y est allé de manière violente et pour un vrai résultat. On a pris des chiffres, on a convoqué les journalistes, on leur a montré le trafic, "choisis ton camp camarade", et ça les a plutôt convaincus. Je pense vraiment qu'ils vont s'y mettre. A moyen et long terme, tous les journalistes feront un titre pour leur lecteur et un titre pour Google. Ça bouffe du temps mais c'est mieux pour tout le monde. »

Lexpress.fr — Directeur de l'acquisition d'audience — Marc Bougué

Nos observations divergent de celles effectuées par Murray Dick à la *BBC* en cela que le dédoublement du titre est considéré à Lexpress.fr comme un point qui peut être extrêmement bénéfique pour le référencement, tandis que le référenceur de la *BBC* n'y voit qu'un faible impact potentiel (Dick, 2011, p. 468). Murray Dick ne semble en revanche pas avoir noté que le dédoublement du titre était également fait pour que les journalistes échappent aux pressions liées au référencement, et puissent écrire un titre qu'ils jugeront intéressant du point de vue éditorial.

9.1.5 Bâtonnage

Nous avons constaté qu'à Lexpress.fr, les journalistes étaient partagés entre un travail dit de « bâtonnage », qui consiste à retravailler très rapidement les dépêches AFP et à les mettre en ligne, et leur travail de fond, qui consiste à rédiger ce qu'ils nomment eux-mêmes les « papiers à valeur ajoutée » : trouver une information exclusive, adopter un angle original, fournir une analyse. La tâche de publication des dépêches a déjà été plusieurs fois observée (Boczkowski, 2004 ; Estienne, 2007 ; Cabrolié, 2010). Elle peut aboutir à une uniformisation des sujets traités par les sites d'information y ayant recours. Dans notre cas, il a été intéressant de constater que les journalistes se plaignaient du fait que ce soit le plus souvent les dépêches bâtonnées, et souvent optimisées, qui soient mises en avant par Google plutôt que les papiers à valeur ajoutées. En revanche, le bâtonnage est considéré comme une activité apportant du trafic, et donc des revenus, sans demander trop d'efforts, et permettant ainsi aux journalistes d'œuvrer à la santé financière de leur employeur. Ainsi, les journalistes acceptent d'effectuer un travail dont la valeur journalistique est assez faible dans le but de pouvoir continuer par ailleurs à effectuer un travail qu'ils considèrent comme étant de qualité.

« Nous on va faire des trucs pour que ça marche sur Google. On voit dans Google Insights ou dans Google Actu qu'un sujet buzz et qu'on a une dépêche : hop, on bâtonne, et on fait de la page vue facile. A côté de ça, si tu veux faire de la qualité, tu la fais à côté. Tu fais les deux en même temps. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

« Bâtonner » consiste à réécrire le titre et le chapeau d'une dépêche AFP, éventuellement à reformuler le corps du texte, puis à publier une page en ajoutant une illustration. Le fait d'ajouter systématiquement une image nous a clairement été décrit comme un point avantageux en matière de référencement.

« Les images sont aussi référencées par Google. Dans Google News, tu peux avoir une image qui renvoie vers ton article. Donc il faut toujours mettre des images. Ça me prend un peu plus de temps que de faire une dépêche brute, mais ça vaut le coût. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Emilie Lévêque

En revanche, certains journalistes regrettent le temps pris pour ajouter une image à chaque dépêche, qu'il considère comme une action spécifiquement effectuée pour le référencement et dont ils remettent en cause l'efficacité.

« On réécrit les titres et les chapeaux pour les dépêches. Il paraît aussi que mettre une photo systématiquement, ça augmente le référencement ; alors on le fait. Mais j'aimerais bien savoir si ça augmente vraiment le référencement, parce que je n'en suis pas sûr et que ça nous demande vraiment du temps supplémentaire. Bon, et du coup, Google arrive, et voit un titre, un chapeau, une illustration, un début personnalisé et des liens qui renvoient de façon pertinente sur des papiers à propos du sujet, c'est-à-dire qu'on apporte une certaine profondeur à l'information. Cela donne en théorie plus de chances à une dépêche, même si ce n'est qu'une dépêche, de remonter dans Google News. Donc normalement, Google récompense la qualité du travail journalistique, après dans les faits on sait très bien que sur la masse de dépêches qui sont faites dans la journée, certaines sont reprises, d'autres non, et cela nous échappe en grande partie. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

Le rédacteur en chef considère ici que Google récompense la valeur ajoutée par le journaliste à une dépêche. Le fait que le bâtonnage, même rapide, ait consisté en un véritable traitement, et qu'il soit ajouté de la « profondeur » à la dépêche, notamment grâce aux hyperliens, génère théoriquement davantage de trafic en provenance du dispositif. La dépêche, ainsi révisée, *performerait* donc de manière à ce que Google récompense la qualité. Toutefois, d'autres journalistes peuvent penser que le bâtonnage est une manière de « tromper » le lecteur : lui faire croire, grâce à un titre retravaillé, que le sujet a fait l'objet d'un traitement approfondi.

« Quand je batonne, je fais attention au référencement. Ce qui est pratique, c'est que les dépêches AFP ou Reuters sont construites pour le référencement, c'est-à-dire que les infos sont condensées au début. En revanche, c'est important de batonner. Il faut modifier la dépêche pour Google et pour les lecteurs, parce que je n'ai pas envie que les lecteurs se disent : "sur l'Express on trouve exactement la même info que sur le Figaro et le Nouvel Obs" [...] Bon, après, le problème avec Google, c'est qu'en voyant le lien, on ne sait pas si c'est un papier de fond ou pas. C'est pour ça qu'on change le titre : pour que le lecteur se dise que ça a l'air mieux qu'ailleurs. Donc il y a une part de jeu : parfois tu trompes le lecteur quand tu batonnes en refaisant une attaque et que tu ajoutes un para de contexte avec un lien vers un papier de la semaine d'avant. Tu donnes l'illusion que t'as enquêté, alors que t'as juste ajouté un truc pour Google. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

9.1.6 Du papier au web

En plus des articles à valeur ajoutée et du bâtonnage de dépêches, les journalistes sont également en charge de la mise en ligne des articles du magazine. Ils « bâtonnent » les articles de façon à ce qu'ils correspondent aux critères de l'écriture web, et en particulier aux bonnes pratiques en matière de référencement.

« Pour le papier, c'est moi qui m'occupe de l'édition. Ça peut même m'arriver de changer les mots, les tournures, surtout dans les chapeaux. Titres et chapeaux. Il faut que ce soit Google-friendly. Je ne le fais pas tout le temps, mais mes deux chefs reprennent la plupart du temps les titres et les chapeaux parce qu'il faut réussir à glisser un maximum de mots-clés. »

L'entreprise.lexpress.fr — Journaliste — Emmanuel Colombié

Les deux journalistes du magazine *L'Express* interrogés nous ont dit ne pas être gênés par la réécriture de leurs titres et de leurs chapeaux par la rédaction web :

« C'est à nos collègues du site de trouver un titre "Google compatible" et ça arrive qu'ils changent en effet nos titres, et ils peuvent écrire un chapeau, une espèce de lancement du papier, dans lequel je vois bien qu'ils ont glissé les mots qui comptent pour Google.

Ça ne vous a jamais dérangé ?

Non, sauf cas évident où une erreur factuelle ou une faute d'orthographe, ça ne me dérange pas. Je sais bien que les règles d'écriture sont différentes selon les supports. »

L'Express (papier) — Rédacteur en chef du service Monde — Marc Epstein

9.1.7 Liens hypertextes

Nous avons vu dans le chapitre 8 que L'express.fr pratiquait ce que nous avons nommé « l'égo-centrisme hypertexte ». Les articles du site, dans notre échantillon, comportent 315 liens dont 284 (90%) vers des pages de L'express.fr parmi lesquels 131 (42%) pointent vers des pages d'atterrissage (cf. 7.2.2 et 7.2.3). Comme pour la BBC, le *Guardian* et *Northcliffe* (Dick, 2011, p. 471), les liens vers les pages d'atterrissage sont générés automatiquement dès qu'un mot-clé associé à une page est utilisé par un journaliste dans le corps de l'article. En revanche, Murray Dick ne parle que très peu de l'impact que peut avoir la stratégie de référencement sur les liens effectués par les journalistes. Dans son travail de thèse, Juliette de Maeyer a effectué une étude auprès de deux rédactions belges francophones, où elle a mis à jour une

configuration débouchant « sur des dissonances entre ce que plaide le service marketing (les liens externes sont potentiellement mauvais pour le SEO) et ce que croient — de bonne foi — les journalistes (les liens externes sont toujours bons pour le SEO) » (De Maeyer, 2013, p. 242-243). A Lexpress.fr, au contraire, les journalistes pensent que les liens internes avantagent le SEO et que les liens externes distribuent à d'autres sites, parfois concurrents, l'autorité PageRank. Ce n'est pas pour autant que les journalistes ne font pas de liens externes, mais ils le font en étant conscients que cela avantagera le référencement de leurs concurrents.

« On fait en sorte de linker plutôt sur nos papiers. Mais quand on veut faire un lien vers ailleurs, même quand c'est des concurrents, et même si du coup ça leur envoie potentiellement de l'audience et ça aide leur ranking, on le fait, parce que d'abord on va se faire un ami du site vers lequel on pointe, et en plus on va se faire un autre ami : l'internaute, qui va se dire "ouah, c'est vachement fair play" et qui va revenir chez nous. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Le but, c'est de redonner une vie à des papiers que nous avons faits. Pas de la dépêche bâtonnée, mais des papiers avec une analyse. Par contre, pour moi, l'honnêteté journalistique, c'est dès que je cite le papier d'un autre confrère ou un autre article paru dans un autre journal, je fais le lien vers cet article. Toutes les source que je peux linker, je les linke et ça même si ça atténue le PageRank. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Emilie Lévêque

Pour les journalistes, le lien a une signification éditoriale. Ainsi, la bonne pratique recommandée par Google, qui consiste à effectuer le lien depuis des ancres significatives, c'est-à-dire depuis des mots-clés décrivant le contenu pointé par le lien, semble conciliable avec ce que le journaliste pense de l'usage qui doit être fait du lien.

« J'essaye de faire en sorte que le lien corresponde vraiment au mot sur lequel il porte. Ça m'exaspère d'avoir : "le meurtrier présumé d'Agnès" qui renvoie vers "Machin va devoir ouvrir une enquête", alors qu'on a un portrait du meurtrier présumé d'Agnès. Je fais attention à ça. J'ai quelques TOC comme ça. Celui-ci en fait partie, et tant mieux, parce que j'ai l'impression que ça sert au référencement. Je fais aussi attention à ne jamais mettre plus de huit mots pour un lien. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Un journaliste nous a expliqué toutefois qu'étant donné la signification éditoriale qu'il attribuait au lien, les recommandations effectuées par le référenceur pouvaient entrer en conflit avec ce qu'il estimait, lui, être pertinent. En conséquence, il ne respecte pas forcément ces recommandations.

« Parfois, ça peut entrer en conflit avec la signification du lien. Un lien a une signification éditoriale. Si je mets un lien sur "Eva Joly, qui a renoncé à se présenter à la présidentielle, a décidé de devenir présidente de Sortie du Nucléaire", eh bien le lien qu'on met sur "Eva Joly" il faut que ce soit un portrait d'Eva Joly, une fiche de présentation d'Eva Joly ou un groupement thématique à propos d'Eva Joly. En revanche, si j'incorpore "qui a renoncé à se présenter à l'élection présidentielle", il faut que la page derrière le lien explique pourquoi elle a refusé de se présenter, et non pas une page thématique ou un portrait. Et pourtant, c'est ça qu'on nous demande de faire pour le SEO : toujours renvoyer vers une page thématique [i.e. page d'atterrissage]. Donc moi ma politique c'est de prendre "qui a renoncé à se présenter à l'élection présidentielle", et de mettre le lien, tant pis pour "Eva Joly", sur l'ensemble des mots, parce que je trouve qu'éditorialement ça a une signification. »

Lexpress.fr — Rédacteur en chef adjoint — Thomas Bronnec

Le référenceur a également demandé aux journalistes de mettre les liens tôt dans le papier. L'un d'entre eux nous a expliqué que cela avait directement impacté son écriture, sans altérer pour autant le message. En même temps, il nous a également expliqué qu'il était tout à fait libre de suivre ou non une telle recommandation :

« On nous a expliqué que c'était beaucoup mieux de mettre des liens très tôt dans le texte, pour des questions de référencement sur Google, donc ça je le fais dès que je peux. Ça peut influencer l'écriture, je veux dire : ça peut influencer carrément ma manière d'écrire. Par exemple, si tu écris un papier le lendemain du meeting du Bourget d'Hollande, tu peux avoir tendance à dire dès le début "Au lendemain du meeting du Bourget" pour pouvoir mettre le lien sur le papier que t'as fait la veille. Donc effectivement, ça peut impacter l'écriture, mais maintenant c'est un choix. Je suis libre de faire comme je veux. Moi c'est vrai que depuis qu'on m'a dit ça j'essaye de le faire au maximum, donc ça change ma manière d'écrire les attaques de mes papiers. Je fais une attaque très courte pour rentrer tout de suite dans le vif du sujet et tout de suite mettre des liens. Journalistiquement, il faut dire que ça a du sens aussi, donc ça ne me dérange pas. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Les journalistes ne veulent pas qu'il y ait trop de liens dans leurs articles. A leur avis, cela gêne le confort de lecture en altérant la forme de ce qu'ils produisent. En outre, le référenceur demande aux journalistes de ne pas faire trop de liens pour que chacun des liens internes puisse avoir un impact significatif en termes de référencement, car le PageRank distribué n'aura pas été réparti sur des liens trop nombreux. La volonté des journalistes de ne pas faire trop de liens et les conseils du référenceur s'accordent donc parfaitement sur ce point.

« Il n'y a rien de plus horrible que d'écrire un article avec des liens soulignés partout. Donc voilà : on en met 4 ou 5, et il faut que ce soit le plus pertinent possible. On pointe quand même surtout vers nos contenus à nous, à part quand on cite une étude ou un baromètre. Après, quand on envoie vers d'autres sites d'infos, on essaye de rester dans les sites du groupe. On envoie vers L'Express, L'Expansion. On essaie quand même de faire rester les gens chez nous. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

« Finalement, je trouve que c'est affreux d'avoir un paragraphe avec trop de mots soulignés, donc si [le référenceur] me dit de ne pas faire trop de liens moi ça me va très bien. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Cadre 3. Nombre et fonction des liens

Un point extrêmement intéressant nous a été révélé par un consultant de Resoneo ayant travaillé pour différents sites de RMF, ainsi que d'autres SIPI et certains sites nés en ligne. En fait, puisque nous avons vu que des liens pouvaient être générés automatiquement vers des pages d'atterrissage, (cf. 8.2), il se trouve que si les journalistes font trop de liens manuellement, ceux-ci diluent l'autorité distribuée par les liens générés automatiquement vers les pages d'atterrissage. Dès lors, un usage trop intensif de la part des journalistes du lien pourrait nuire à la stratégie de SEO. C'est pourquoi un dispositif d'alerte peut être mis en place visant à limiter l'action des journalistes, sans pour autant que ceux-ci soient destitués de leur « pouvoir de faire ». Une alerte les prévient qu'ils ont dépassé le « quota idéal » de liens, sans que rien ne les empêche de ne pas en faire davantage. Cependant, le consultant nous a expliqué que le dispositif pouvait être transformé de manière à empêcher la validation des articles dont les auteurs auraient dépassé le quota de liens autorisés. Même si un tel dispositif d'alerte n'a pas été mis en place à Lexpress.fr (c'est pourquoi nous le séparons du reste de notre argumentation), il s'agit d'un

excellent exemple des ajustements auxquels la question du référencement peut donner lieu dans le CMS de publication, et de la manière dont un référenceur peut tenter de « conduire la conduite » des journalistes, sans rien leur imposer. L'alerte émise lorsque le quota de lien sera dépassé aura alors pour but de *performer* l'action du journaliste. En revanche, dans le cas d'un dispositif d'empêchement, il n'y aura plus performance mais *restriction* pure et simple.

« Deux choses viennent se battre », nous a expliqué le référenceur, avant de préciser que les liens dont le nombre était limité pouvaient être répartis de la manière suivante : 1/3 pour le référencement et 2/3 pour les journalistes. Ces liens n'auront pas le même but. Les premiers pourront être invisibles, tandis que les seconds seront chargés d'une fonction de monstration. Le référenceur nous a également expliqué qu'en cas de conflit, c'étaient les liens SEO qui céderaient leur place aux liens journalistiques. Dans une telle situation, la fonction éditoriale attribuée par le journaliste au lien primerait sur la fonction de communication attribuée par le référenceur.

Le référenceur a recommandé aux journalistes de Lentreprise.lexpress.fr de pointer systématiquement vers une vingtaine d'articles dont il leur a donné la liste, déjà particulièrement bien référencés. Ainsi, le PageRank est dirigé par les journalistes, dès qu'ils le peuvent, vers ces articles, et ce même si cela n'a pas de signification éditoriale. Les 20 articles font ainsi office de vitrine : ils sont utilisés comme des pages d'atterrissage. Le référenceur fournit également aux journalistes une liste de mots-clés qu'il leur recommande d'utiliser dès que possible, et à partir desquels ils doivent ensuite faire un lien vers l'un des 20 articles.

« Les articles qui marchent bien, il faut qu'on arrive à les glisser dans les autres articles sous la forme de liens, parce que ça fait des espèces d'ascenseurs. Par exemple, j'ai fait un article il y a deux semaines sur le Plan Fillon, qui dit qu'on va raboter le dispositif de défiscalisation des heures supplémentaires. Dans cet article, à un moment je parle du SMIC, donc j'ai mis sur le mot SMIC le lien vers notre article champion sur ce sujet. Après, on ne peut pas vérifier directement que ça marche, mais ce sont des recommandations que nous fait Renaud, qui nous a envoyé une liste de mots-clés et d'articles. J'ai un tableau sur mon bureau avec les 20 articles qui marchent le mieux, comme ça dès que je peux je les glisse, et ça même si l'article que j'écris n'est pas vraiment corrélé à l'article que je linke. En plus de ça, on a les mots-clés, et il faut qu'on fasse l'effort à chaque fois qu'on écrit un papier de penser à ces mots-clés et à ces articles pour essayer de les intégrer sous la forme de liens dans les papiers qu'on écrit. Ça a donc complètement une influence sur ce qu'on fait. D'ailleurs, on travaille à

moitié pour Google hein, malheureusement. Et c'est pénible, parce que ça prend du temps, et en plus tu dois y penser pas à chaque fois. Bon, tu ne vas pas non plus mettre "SMIC" sur un article qui n'a rien à voir. Il faut que tu sois cohérent, sinon c'est de la magouille. On n'est pas dans la magouille. Mais bon, comme consigne, on doit intégrer au moins 3 ou 4 liens par article. C'est le référenceur qui nous a donné ces consignes. On ne le fait pas à chaque fois, mais plus ça va plus on le fait. On a vraiment l'impression de bosser pour Google. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

Le verbatim ci-avant finit par une plainte du journaliste qui a l'impression de travailler pour Google en effectuant certaines actions — ici les liens vers les 20 articles d'atterrissage — qui ne lui paraissent pas relever de sa fonction, et qui lui prennent un temps qu'il aurait justement pu (et dû, selon lui) consacrer à cette fonction. Il demeure ainsi certaines pierres d'achoppement, où les scripts du référenceur et des journalistes s'opposent. C'est alors le journaliste qui a le dernier mot, puisque c'est à lui qu'il appartient de faire ou non un lien.

Finalement, nous avons constaté que le lien hypertexte et Google étaient *liés* dans l'esprit des journalistes. D'autre part, le lien conçu comme un outil de SEO et le lien conçu comme un outil journalistique peuvent s'opposer, le lien à vocation journalistique étant susceptible de diluer l'impact du lien à vocation SEO, tandis que le lien à vocation de SEO peut limiter la possibilité pour le journaliste d'effectuer autant de liens qu'il le veut, ou bien lui prendre du temps et le gêner dans sa pratique. Cependant, les conflits sont très rares et le référenceur et les journalistes réussissent à travailler sur le maillage hypertexte en bonne intelligence. Il s'agit au quotidien d'ajustements et de légères différences de points de vue plutôt que, véritablement, de disputes, de conflits ou de concessions trop lourdes.

9.1.8 Temporalité de publication

La stratégie SEO peut avoir un impact sur la temporalité de publication. Le formateur à l'écriture web nous ont notamment expliqué que les journalistes avaient intérêt parfois à retarder la publication de certaines dépêches bâtonnées, comme de certains articles de fond, de manière à ce qu'une « grappe » ait déjà été créée sur Google Actualités en haut de laquelle, alors, ils auraient une chance de

figurer. En d'autres termes, il s'agit d'attendre que suffisamment de sites concurrents aient traité le sujet pour soi-même publier ce qu'on possède à propos de ce sujet.

« Il ne faut pas publier un article trop tôt, et il ne faut pas non plus le publier trop tard. Parce qu'en fait, pour arriver sur la Une de Google Actu, il faut que Google identifie qu'il y a beaucoup de sites qui traitent ce sujet et que par conséquent ce sujet est important. L'enjeu est de réussir à se positionner dans Google Actualités au bon moment, et donc de réussir à pas forcément être le premier à sortir une info mais plutôt celui qui saura la sortir au bon moment, parce que si t'es le premier tu risques d'être en bas du cluster. Pour faire ça, il faut surveiller. Il faut que tu passes beaucoup de temps sur les sites de tes concurrents pour voir ce qu'ils publient. »

WAN-Ifra — Formateur — Cédric Motte

Nous voyons ici comment la question du référencement peut souffler sur les braises de la circulation circulaire de l'information. Les journalistes interrogés nous ont pourtant dit ne pas appliquer la méthode décrite ici par leur formateur, parce qu'elle leur semblait contraire à la fois à leurs valeurs, à ce qu'on leur avait appris lors de leurs formations respectives et à l'idée d'une information traitée en temps réel. Sur ce point, il y a une vraie divergence entre le fonctionnement de Google Actualités et ce que les journalistes considèrent comme étant la qualité de leur travail.

La temporalité joue également dans l'autre sens : en théorie, pour une stratégie SEO optimale, il ne faut pas publier trop tôt, certes, mais il ne faut pas non plus publier trop tard, car alors la grappe hypertexte ne sera plus en « Une » de Google Actualités. Cela peut poser un problème à certains journalistes :

« Puisqu'on a des contraintes éditoriales importantes, il nous faut du temps pour écrire notre papier. Donc on va avoir tendance à le sortir plus tard que les autres, et peut-être que ça n'apparaîtra plus sur Google News à ce moment-là. Alors là en gros, on nous fait comprendre qu'on sort nos papiers trop tard et qu'il faudrait avoir un petit papier dès le début de l'après midi. Sauf que quand t'es en train de plancher sur un gros papier, tu peux pas en faire deux, parce qu'on est pas très nombreux, donc concrètement il y a un vrai problème de temps de travail. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Finalement, les journalistes ne changent par leur temporalité de publication en fonction de Google, d'une part car ils ne sont pas certains que cela ait réellement un impact, d'autre part parce que cela leur semble incohérent.

« Garder l'info pour ne pas se faire punir par le cluster, moi je ne le fais pas parce que j'ai du mal à comprendre les règles en vigueur là-dessus. Je ne suis pas sûr qu'un papier qui va être publié en dernier va remonter parce que c'est le plus récent. J'ai l'impression que les algorithmes de Google sont tellement complexes qu'on ne sait jamais. Donc je suis plutôt partisan au contraire de le passer tout de suite. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Un journaliste nous a fait part d'une autre justification quant à son refus d'ajuster la temporalité de publication de ses articles en fonction de Google. Cette justification est intéressante, car le journaliste mentionne le conflit qu'il peut y avoir entre le fait de s'adresser aux utilisateurs de Google et le fait de s'adresser aux lecteurs qui ne sont pas arrivés par le biais d'un infomédiaire.

« Je ne cherche pas à jouer sur la temporalité, et j'ai d'ailleurs l'impression que ça ne marche plus sur Google. C'était vrai à une époque, mais j'ai l'impression que ce n'est plus le cas aujourd'hui. J'essaye d'être présent au moment où ça se fait, pour notamment une raison très simple : la consommation de l'info sur l'application. L'application fait que les gens attendent l'info au moment où elle sort. Ils sont sur leurs mobiles et ils veulent l'info en direct. On ne va pas leur dire : "Hé les gars, attendez cinq minutes pour qu'on soit bien classé sur Google, ok ?" Il faut choisir : si on dit qu'on fait de l'info en temps réel, il faut fonctionner en temps réel. »

Lexpress.fr — Journaliste (Sport) — Clément Daniez

9.1.9 Mises à jour

Une des recommandations relatives au SEO concerne la mise à jour des articles. Lorsqu'un article est modifié, et ce même si la modification est minime (un mot du titre ou du chapeau), le crawler de Google scanne de nouveau son contenu, ce qui peut permettre potentiellement à un article d'être repositionné en haut de la grappe de Google Actualités. Il s'agit d'un moyen de jouer avec le critère « degré de fraîcheur » de l'algorithme (cf. 4.3.3), c'est-à-dire qu'il s'agit de faire croire au dispositif que l'information est fraîche alors qu'elle ne l'est pas. Les journalistes nous ont dit manquer de temps pour remettre systématiquement à jour des articles pour Google. Dès lors qu'ils n'ont pas une information à ajouter à ce qu'ils ont écrit, un arbitrage a lieu : ils sont prêts à se mobiliser, c'est-à-dire à effectuer des mises à jour, mais seulement jusqu'à un certain point et à condition qu'ils ne soient pas trop « débordés » par d'autres tâches. Si la mise à jour n'aboutit pas à l'effet escompté,

autrement dit si l'article ne remonte pas dans Google Actualités, il peut arriver aux journalistes de réessayer une fois ou deux, mais pas davantage.

L'une des missions confiées au *Front Page Editors* (FPE) est de seconder ses collègues dans cette tâche de mise à jour. Ainsi, le problème lié au temps du journaliste rédacteur est contourné, dans la mesure où le FPE, lui aussi journaliste, remplit cette tâche à sa place.

« Le boulot des FPE, c'est de faire tourner la moulinette de manière à ce qu'on n'ait pas un papier qui occupe pendant deux heures une place qui aurait dû être occupée par un autre, et que Google repère les papiers ; c'est-à-dire que les papiers soient à la fois assez bien placés et titrés et édités pour qu'ils remontent dans Google News. Donc les FPE ont un œil sur Google en permanence et un œil sur le papier, et se disent : "Tiens, on va changer un mot dans le titre et voir si ça marche. On va changer l'ordre d'édition, on va rajouter un paragraphe, et voir ce que ça fait sur Google News". »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Les FPE, pratiquement toute la journée, remettent à jour les chapeaux, les titres, principalement pour des logiques de référencement. Ils viennent te voir parfois pour en parler. Ils éduquent. Moi après, qu'on touche à un de mes papiers, ça ne me dérange pas. Je n'ai pas d'égo vis à vis de ça. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Le FPE effectue la mise à jour à la place des rédacteurs et communique avec eux au sujet des bonnes pratiques SEO. Ainsi, le FPE seconde le référenceur dans sa tâche, faisant en sorte de sensibiliser les journalistes de la rédaction à la stratégie qu'il a mise en place. Sa fonction est en partie consacrée à l'optimisation pour Google, et à créer des « circonstances favorables » (Austin, 1970, p. 43) à la performance. C'est un exemple de la manière dont le travail peut s'organiser, ou se réorganiser, autour de la question du référencement, certains acteurs jouant le rôle, comme le référenceur, de « centres de traduction », qui renforceront ou au contraire atténuent l'action du référenceur, et, *in fine*, l'influence de Google. Le FPE, parce qu'il est journaliste, arbitre entre le SEO et le journalisme. S'il n'y a pas d'information supplémentaire, et donc aucune valeur ajoutée journalistique, il ne procèdera pas à la mise à jour :

« Si ça ne remonte pas sur Google News, je ne change pas automatiquement. On a appris les limites du SEO en essayant de comprendre le fonctionnement de Google News : on sait qu'il ne suffit pas de re-titrer un papier pour le faire remonter. Parfois ça marche et parfois pas du tout. Donc je suis plutôt pour re-titrer un papier quand j'ai une actu nouvelle, et que j'ai envie de montrer que le papier a changé. [...] C'est toujours un peu délicat, parce que je ne suis pas là pour re-titrer un papier toutes les demi heures juste pour remonter sur Google, d'autant que ça ne fonctionne pas à tous les coups. Il faut trouver une juste mesure là dedans. Tu ne vas pas mentir à ton lecteur en lui faisant croire que ton papier est nouveau, alors qu'il n'a pas évolué. »

Lexpress.fr — FPE — Camille Caldini

Nous retrouvons l'idée déjà mentionnée à plusieurs reprises qu'il ne faut pas, selon les journalistes, que les pratiques de SEO aboutissent à un « mensonge » ou une « tromperie » vis-à-vis du lecteur. Ce n'est jamais à Google qu'on ment directement, mais au lecteur, c'est-à-dire à l'utilisateur des services de Google.

Toujours au sujet de la mise à jour, nous avons recueilli auprès d'une même journaliste deux exemples de situations où une tension se créait, dans le premier entre le référenceur et le journaliste, dans le second entre deux journalistes, pour des raisons opposant la stratégie de SEO aux visées journalistiques. Premier exemple :

« L'affaire du meurtre d'Agnès est sortie vendredi. Samedi, la personne qui était de permanence a fait un "Ce que l'on sait sur le meurtre d'Agnès". C'est un format assez récurrent en Société, où tu dis qui est la victime et où en est l'enquête. Le papier a été très bien référencé sur Google. Il y avait plein de commentaires, genre 200, et l'article remontait bien. Puis le dimanche, il y a eu de nouvelles actus et de nouveaux papiers. Et le lundi, pareil. Alors là, [le référenceur] vient me voir, enfin il vient voir la FPE, qui ensuite me fait passer le message, et il dit : "Les nouveaux papiers qu'on a fait sur le meurtre d'Agnès ne remontent pas, le seul qui est visible sur Google c'est le "Ce que l'on sait". Donc il faudrait faire une actualisation et le remettre en home du site." Très bien, mais c'est un papier de samedi, et il y a eu plein de nouvelles actus entre temps, donc je ne me vais pas remettre en home un vieux papier. Renaud voulait qu'on le mette à jour, vite fait, genre changer un peu le titre, un peu le chapeau et ajouter un para, pour que ça change la date de mise à jour et que Google le reprenne. Tout ça alors que je préparais un gros papier sur l'identité du meurtrier présumé. Je préférerais qu'on mette en avant le papier qui était nouveau avec de nouveaux éléments plutôt que de remonter un papier publié samedi par une personne de permanence. Tu vois, ça, pour moi, c'est aberrant. Je comprends totalement la logique de tenir compte du référencement sur Google, mais je trouve que parfois c'est poussé à l'extrême. Alors voilà, j'ai refusé de le faire, et je pense que j'ai bien fait. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Nous voyons ici apparaître l'interaction des différents centres de traduction : le référenceur (centre n°1) prévient la FPE (centre n°2) qui prévient le journaliste. L'exemple montre par ailleurs comment le script du journaliste quant au traitement de l'actualité et à la temporalité de publication peut entrer en conflit avec le script du référenceur. Ce dernier demande à la journaliste de republier un article vieux de deux jours, alors que d'autres articles ont été publiés entre temps et que la journaliste est elle-même occupée à rédiger un article de fond à propos du même sujet. Le référenceur demande de procéder à des modifications éditoriales de l'article daté, sans que celles-ci n'aient aucune valeur informationnelle. Il demande également au FPE de changer la hiérarchie des informations sur la page d'accueil, pour mettre en haut un article qui n'est pas le plus récent. Finalement, la journaliste refuse de suivre les recommandations, qui lui paraissent « aberrantes ».

La même journaliste nous a fait part d'un deuxième exemple :

« Il m'est arrivé un matin de trouver deux fois sur le site la même dépêche publiée. Je demande à la personne de permanence s'il y a eu un problème, et elle me répond : "Non, la première dépêche n'a pas été vue par Google, donc j'ai refait la même dépêche". C'est complètement stupide. Tant pis, si "Trois morts dans un incendie à Pantin" n'est pas remonté sur Google. On ne va pas perdre notre temps à faire la même dépêche une deuxième fois avec les mêmes titres, les mêmes chapeaux, les mêmes liens, le même sommaire, pour une dépêche en plus ! Il faut arrêter là, parce que ça devient ridicule. Je ne comprends pas du tout, et je refuse de le faire. Imagine un peu pour les lecteurs qui ne viennent pas par Google, c'est ridicule ! Ils se demandent ce qu'il se passe. Alors je comprends pourquoi on peut faire ça, et je sais que [le référenceur] a des bonnes raisons de nous demander de faire des trucs comme ça, mais ce ne sont pas mes bonnes raisons à moi et de mon point de vue, dans des cas extrêmes comme celui-là, ça peut même devenir ridicule. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Nous voyons ici, d'une part, que la journaliste critique une action effectuée par un autre journaliste spécifiquement pour remonter sur Google Actualités, alors que cela n'a aucun sens journalistique. Ainsi, parce qu'ils n'ont pas tous le même degré d'implication dans la stratégie SEO, les journalistes peuvent se disputer entre eux. Cette fois, la négociation a lieu en marge des centres de traduction : les journalistes ne sont pas d'accord quand à l'interprétation qu'il faut avoir vis-à-vis de l'action du référenceur mais ce dernier, lui, n'agit pas, ou plus. La journaliste trouve « ridicule »

le fait de publier deux fois la même dépêche. Elle considère que c'est une perte de temps. Et elle considère que ce qui a été fait pour attirer les usagers de Google Actualités nuit à la qualité de ce qui est fait pour les lecteurs arrivés directement sur le site. Elle explique enfin que les raisons de ceux qui pratiquent une stratégie de SEO sont sans doute de « bonnes raisons », mais que ce ne sont pas les mêmes que les siennes. En quelque sorte, nous pourrions pasticher Pascal en disant que « le SEO a ses raisons que la raison du journaliste n'acceptent pas ».

*

Tout au long de cette sous-partie, nous avons montré comment le cours d'action des journalistes était travaillé par la question du référencement. D'autre part, nous avons vu que les journalistes demeuraient largement indépendants, libres de s'enrôler dans la stratégie du référenceur ou non. Ils choisissent, arbitrent, mesurent, refusent. Certaines actions peuvent parfaitement s'accorder à la fois aux objectifs du référenceur et aux objectifs du journaliste, chacun d'entre eux justifiant une même action différemment. D'autres actions opposent les critères journalistiques aux critères du référencement, et alors certaines concessions peuvent être faites de part et d'autre, sachant qu'en l'absence d'ascendant hiérarchique de la part du référenceur, c'est toujours le journaliste qui aura le dernier mot dès lors qu'il s'agira de l'aspect éditorial de la production. Nous avons également vu que le FPE pouvait jouer un rôle de tampon entre le référenceur et les journalistes, devenant alors lui-même un « centre de traduction », c'est-à-dire, dans le réseau, le lieu où les visées des uns sont *traduites* (i.e. interprétées et déplacées) pour être comprises par ceux-là dont la collaboration est nécessaire à l'aboutissement du projet du référenceur.

Nous allons à présent quitter les pratiques journalistiques pour nous intéresser aux représentations des journalistes de l'action de Google et de la stratégie SEO mise en place chez RMF. En décortiquant ce que nous appelons leur « rapport dialectique », nous pourrions expliquer au nom de quoi les pratiques décrites ci-avant sont effectuées, partiellement effectuées, détournées ou refusées. En d'autres termes, après nous être intéressé à ce que peuvent faire et font les journalistes en fonction de Google et du référenceur, nous nous intéresserons à *ce qu'ils en pensent* et au caractère « psychogène » de la performativité (Muniesa et Callon, 2009).

9.2 Rapports dialectiques

Le mot « dialectique » vient du grec *dialegesthai* : « converser ». Nous le considérons ici dans son sens propre, et non selon les définitions paradigmatiques (et différentes) qu'ont put en donner des philosophes comme Hegel, Feuerbach, Sartre et Adorno. La dialectique désigne donc ici le rapport, direct ou indirect, entre deux entités n'ayant pas les mêmes scripts. De surcroît, le terme a cela de particulier qu'il désigne le dépassement des divergences et l'élaboration d'une synthèse commune, plus ou moins satisfaisante aux yeux des parties prenantes, mais tout de même : satisfaisante.

9.2.1 Meilleur ennemi n°1 : Google

A. Divergences

Comme nous l'avons expliqué pour le niveau mezzo de notre terrain (cf. 7.1.1), les journalistes craignent d'être dépendants à l'action d'un moteur dont ils ignorent le fonctionnement, et dont l'entreprise propriétaire poursuit d'autres buts que les leurs. Le quasi-monopole de Google renforce ce sentiment. Cependant, le rédacteur en chef de Lexpress.fr nous a également expliqué que la firme Google « faisait son métier » et qu'elle ne pouvait par conséquent pas être pointée du doigt. Selon lui, c'est aux propriétaires des SIPI de se dégager de cette dépendance, ou au moins de s'en arranger.

Quant aux journalistes, ils nous ont dit qu'ils avaient l'impression de subir au quotidien le « poids Google », ce qui rejoint les observations menées par Murray Dick (Dick, 2011, p. 472). Ils nous ont également expliqué que les journalistes du titre papier seraient étonnés s'ils savaient à quel point le site web dépendait de l'infomédiaire. En plus de cela, dans les deux verbatim ci-après, le SEO est présenté comme une « technique de vente » qui « n'est pas l'essence du métier ». Les journalistes regrettent par conséquent de dépendre de Google, cependant ils n'imaginent pas non plus qu'eux ou leurs lecteurs puissent se passer de son moteur.

« Google, c'est de l'amour-haine. C'est terrible de dire ça, mais on est obligés de faire avec, parce que tout passe par Google. C'est la googlization des esprits. Si nos collègues du papier voyaient comment on dépend de Google, ils deviendraient fous. [...] Le référencement, c'est pas du journalisme, c'est juste de la technique en plus. C'est de la technique de vente. Ça a beau être ce qu'on nous apprend à l'école maintenant, ce n'est pas l'essence du métier. L'essence du métier, c'est de raconter ce qu'on voit et ce qu'on sait. Raconter et donner les informations. Ça n'a rien à voir avec Google, et pourtant Google occupe une importance considérable pour nous, et je le regrette énormément, parce que ce n'est pas notre métier. »

Lexpress.fr — Journaliste (Sport) — Clément Daniez

« On subit complètement. On ne peut même pas imaginer ce que serait le web s'il n'y avait pas Google. [...] Mais bon, on ne l'a pas non plus vraiment choisi ce truc là. Il est arrivé et il a bouffé tout le monde. Il y a quelques années, je pensais que les autres réussiraient à ne pas se faire avoir comme ça. Nous, quand on parle SEO, on ne parle pas des autres moteurs. On ne parle que de Google. C'est marrant parce qu'ils nous disent les « robots Google », alors tu imagines un robot qui passe le matin, le soir. Tu sais pas comment ça fonctionne mais c'est comme ça. Il faut travailler avec ça. C'est Google qui le veut. Moi, j'ai l'impression de subir. En même temps, je sais pas comment ça fonctionnerait le web sans Google. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

Selon les journalistes interrogés, le trafic apporté par le dispositif n'est pas fonction de la valeur ajoutée du traitement de l'actualité, mais de détails techniques, de réactivité et de performance. Ils déplorent le fait qu'une dépêche bâtonnée puisse être rangée dans une grappe au-dessus des articles de fond différenciant leur production de la concurrence. Malgré tout, ils sont persuadés que Google essaye de remédier à cette situation en paramétrant son algorithme de manière à ce que le contenu de qualité soit avantagé. Ils citent la modification d'algorithme « Panda », visant à sanctionner les fermes de contenus, comme une action exemplaire visant à sanctionner les fermes de contenu.

« Google News permet souvent à de simples dépêches d'être en première position devant des papiers fouillés qu'on a fait. Là, c'est clair, ça nous fait enrager. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

« Il faudrait que le critère primordial dans l'algorithme, ce soit la qualité de ce qu'on fait, la réactivité, l'imagination. Bon, il se trouve qu'il n'y a pas que ça qui est pris en compte pour être bien placé. Mais Google tend vers ça. »

C'est ce qu'ils ont fait avec Panda. Ça marche, ça marche pas, ils essayent... Nous, on s'adapte. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Pour une dépêche AFP traitées par tous les sites d'infos, c'est Midi Libre qui remonte, avec un titre ignoble, trois virgules, deux points, la dépêche pas retravaillée. Je ne comprends pas pourquoi c'est ça qui remonte alors que c'est juste illisible. »

Lexpress.fr — Journaliste (Monde) — Julie Saulnier

« Le problème de Google, c'est qu'il privilégie immédiatement une dépêche. Il hiérarchise très peu. Une dépêche aura la même valeur qu'un papier de fond sur le même sujet. Mais à terme, un papier à VA finit par être repéré par les internautes, partagé sur Facebook, et remontera plus tard par Google [Search] plutôt que par Google News. »

Lexpress.fr — Journaliste (Monde) — Catherine Goueset

La journaliste Julie Saulnier reproche quant à elle à Google l'apparence du lien sur la liste de résultats. Selon elle, le nom de la source (*L'Express*) n'est pas assez mis en valeur : écrit trop petit, en gris, illisible. D'autre part, l'absence de nom d'auteur à côté du lien est elle aussi regrettée. C'est le droit au nom et à la paternité qui est mis en avant ici, composante du droit moral des auteurs :

« Sur Google News, le nom de la source est écrit tellement petit que les gens ne savent même pas de qui est ce qu'ils lisent. Tu ne vois pas bien que c'est L'Express, et tu ne sais pas qui, à L'Express, a écrit ça. Tu ne sais pas si c'est une dépêche AFP, non signée, ou le travail d'un auteur. Ça arrangerait beaucoup les choses, à mon avis, s'ils mettaient ça en avant chez Google. »

Lexpress.fr — Journaliste (Monde) — Julie Saulnier

Ces divergences et ces craintes sont autant d'éléments susceptible de gêner la chaîne d'associations/traductions. Nous pouvons leur ajouter la critique du journaliste anglais interrogé par Murray Dick, qui déplore le fait que dans le cas où tous les éditeurs se plieraient parfaitement aux recommandations de Google, les lecteurs se retrouveraient devant des titres indifférenciés sur la liste générée par le moteur (Dick, 2011, p. 473.). Mais nous allons voir ci-après qu'il existe malgré tout des terrains d'entente entre Google et les journalistes.

B. Dépassement

Le rédacteur en chef de *lexpress.fr* considère que Google est un allié plutôt qu'un ennemi. A son avis, la firme « balise le chemin », de manière à ce que les éditeurs de presse en ligne produisant des contenus originaux soient récompensés. Les ingénieurs placent selon lui des « pièges », de manière justement à privilégier un site comme celui de *L'Express*. C'est ce qui justifie aux yeux d'Eric Mettout que l'environnement soit contraignant : respecter les recommandations de Google, et donc accepter que sa propre capacité d'action soit en partie déterminée par les choix de la firme, permet de montrer patte blanche au dispositif et d'être récompensé sans avoir révisé à la baisse les critères de qualité journalistiques. En outre, dès lors que les lecteurs passent par Google, Eric Mettout, qui considère que son métier est de produire l'information mais aussi de la porter aux lecteurs, nous a expliqué qu'il était fondamental d'adapter son action aux actions de Google. Les journalistes sont en accord avec leur rédacteur en chef, notamment le FPE qui, nous l'avons vu, a un rôle central en matière de SEO. La FPE interrogée nous a même dit que le rédacteur en chef pouvait parfois demander aux journalistes de « lever le pied » du côté des pratiques de SEO dans le cas où les considérations liées au référencement pouvaient empiéter sur la qualité de leur travail. C'est alors au FPE de prendre le relais et d'ajouter à la production du journaliste-rédacteur les éléments nécessaires au SEO.

« Il y a un moment où on a eu l'impression de pédaler dans la semoule avec Google. On s'est dit : "Ça va nous détruire. On va courir après l'actu avec un truc comme ça, courir après le mot-clé, et avoir les mêmes titres et les mêmes infos que tout le monde". A partir du moment où on a eu la formation SEO, tout le monde s'est mis à flipper quand il ne voyait pas son article remonter sur Google News. Eric [Mettout] a été obligé de dire de lever un peu le pied : "Détachez vous de Google News. Laissez ça aux FPE, et faites votre boulot." »

Lexpress.fr — FPE — Camille Caldini

Dans le discours des journalistes, « bien écrire pour le web » et « bien écrire pour Google » ont tendance à se confondre, comme si rédiger un document pour le web signifiait *nécessairement* le concevoir de manière à maximiser la captation de trafic depuis Google. De la même façon « les internautes » et « les utilisateurs du

moteur » s'amalgament : un internaute, dans l'imaginaire des journalistes interrogés, utilise *forcément* les services de Google pour s'informer.

« Ecrire pour le web, c'est écrire pour Google. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Google, quelque part, correspond à ce que les internautes veulent. Tout le monde utilise Google. Nous devons faire avec. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Il semble ainsi qu'une montée en généralité soit opérée à l'issue de laquelle Google ne représente pas seulement les utilisateurs et les partenaires de ses services mais également l'ensemble des internautes et, finalement, le web tout entier. Les journalistes assimilent ce qu'ils nomment eux-mêmes « l'écriture Google » comme une condition qu'ils acceptent dès lors que leur métier est de publier du contenu sur le web, donc pour le web, donc pour Google. Par conséquent, l'écriture SEO est présentée comme relevant d'un réflexe :

« On a nos esprits qui sont un peu cannibalisés par ça. On n'a pas besoin d'y penser tout le temps, tellement c'est naturel. Ça coule de source. C'est une logique qui est assimilée, où on a des styles de titre par exemple qui ressemblent à ce que veut Google ; et on le fait. Mais on n'a pas forcément conscience à ce moment qu'on fait du Google. »

Lexpress.expansion.fr — Journaliste — Julie de la Brosse

Les journalistes considèrent comme étant de leur responsabilité de faire en sorte que leurs articles accèdent au lectorat. Pour eux, écrire en pensant aux recommandations de Google est une réalité indépassable de leur quotidien.

« Google, t'apprends très vite que c'est déterminant. Et tu te rends compte très vite que t'as un patron qui est ton rédac' chef, et un autre patron qui est Google. Donc forcément, t'es obligé de le prendre en considération, parce que sinon ton papier n'est pas vu. Il n'est pas accessible et il n'est pas vu. »

Lexpress.fr — Journaliste — Matthieu Deprieck

Allant également dans ce sens, le formateur en « écriture web » explique aux journalistes que leur but ne doit pas être simplement de produire en pensant à un lecteur *lambda*, mais également en pensant aux dispositifs qui permettront aux lecteurs d'accéder au contenu. De la même manière qu'un scientifique n'écrit pas en pensant seulement aux autres scientifiques auxquels il destine son article, mais

également aux consignes données par les éditeurs de la revue à laquelle il destine son article, le journaliste prend en compte les exigences liées au véhicule de diffusion.

« En fait ce que je leur explique c'est que leur métier sur Internet, ce n'est pas seulement de produire de l'info. Leur métier est aussi de malaxer leur contenu, de le retravailler, de l'éditer au sens américain du terme, c'est-à-dire que tu retravailles ton papier de manière à ce qu'il corresponde à des critères de Google ou de communautés ou autre, mais toujours à des critères qui permettront d'être lu. »

WAN-Ifra — Formateur — Cédric Motte

Le référenceur, quant à lui, lorsqu'il sensibilise les journalistes aux recommandations de Google, le fait en *traduisant* ses objectifs propres dans leur langage à eux. Il leur explique notamment que dans la mesure où Google apporte plus d'un tiers du trafic, et donc est à l'origine d'un tiers des revenus du site d'information, le respect de certaines règles éditoriales leur permettra de continuer à être lus, mais aussi d'assurer la survie financière de leur employeur. Dans ces conditions, ils devraient selon lui être prêts à accepter certaines concessions :

« Les journalistes écrivent également pour les moteurs de recherche, parce que les journalistes qui me disent "on s'en fout, on va pas écrire pour Google", moi je leur réponds que ça les fait bouffer, Google. Je pense qu'on réduirait de moitié les effectifs si on n'était plus référencés, alors je le leur dis. Aujourd'hui, notre audience c'est la moitié ou un tiers de notre trafic qui provient de Google, donc on va optimiser, okay ? »

Lexpress.fr — Référenceur — Renaud Perrin

Il nous a semblé lors de notre enquête que « l'écriture pour Google » constituait pour les journalistes une évidence plutôt qu'une contrainte. En effet, les journalistes acceptent l'idée qu'il puisse y avoir des règles propres à l'univers sociotechnique dans lequel ils évoluent et auquel ils destinent leurs contenus, même si certaines frictions peuvent tout de même avoir lieu. D'autre part, les journalistes ont tout à fait conscience des réalités économiques soulignées par le référenceur, et qui constituent en effet à leurs yeux une justification pour le SEO.

« Pour avoir des annonceurs, il nous faut du clic. Si on n'a pas assez de clics, on n'a pas assez d'annonceurs. Et si on n'a pas assez d'annonceurs, des employés quittent L'Expansion, et on ne peut pas renouveler des CDD. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Le rédacteur en chef Eric Mettout nous a également expliqué qu'écrire pour le web consistait à écrire *avec* Google, c'est-à-dire en utilisant les services de Google activement pour trouver de l'information, la vérifier, la recouper. Ainsi, Google s'intégrerait à chacun des maillons de ce que Denis Ruellan et Daniel Thierry désignait déjà en 1998 comme un « continuum informatique », c'est-à-dire l'utilisation de l'ordinateur (et d'internet) aussi bien pour chercher de l'information que pour la traiter, la publier et la diffuser (Ruellan et Thierry, 1998). Le dispositif d'infomédiation permet en effet de mettre en avant la production journalistique, mais fournit également au journaliste l'information en amont de sa pratique de collecte et de production d'information (Pélissier et Diallo, 2009). La performance est triple : performance de l'énoncé produit par Google sur la recherche d'information, performance de la production journalistique sur l'énoncé produit par Google, performance des recommandations émises par Google sur la production journalistique. Cet usage de Google au quotidien comme outil de collecte plaide en faveur d'un apaisement de ce que les journalistes pourraient reprocher à la firme :

« Ecrire web, c'est écrire pour Google, mais aussi avec Google, en utilisant Google, et avec les réseaux sociaux aussi, avec les agrégateurs — pour le web et avec le web. Cette compétence est obligatoire pour les nouveaux journalistes. »

Lexpress.fr — Rédacteur en chef — Eric Mettout

« Toute la journée je me sers de Google comme d'un outil de veille. Il est toujours ouvert. Je m'en sers 25 fois par jour, au moins. Donc c'est aussi vraiment génial, Google, et il ne faut quand même pas l'oublier ça. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Enfin, nous avons pu interroger les journalistes de Lexpress.fr à propos du « Choix des rédactions ». Nous avons vu en effet dans le chapitre 3 (cf. 3.2.5) que cet outil semblait avoir été intégré par Google à Google Actualités dans le but d'atténuer certains reproches qui pouvaient lui être adressés. Dans « Le Choix des rédactions », le nom de l'auteur figure, la source apparaît clairement et les articles de fond peuvent être mis en avant. A Lexpress.fr, c'est le FPE qui s'occupe de choisir les articles qui figurent dans le module. Au moment de notre enquête, « Le Choix des Rédactions » venait tout juste d'être lancé et apportait très peu de visiteurs au site. Cependant, la FPE nous a expliqué qu'elle appréciait l'outil, car il comblait selon elle certains

manques de Google Actualités et montrait aux journalistes la bonne volonté de la firme à leur égard. Notamment, d'après la FPE, « Le Choix des Rédactions » permet de faire remonter des articles de fond qui ne correspondent pas toujours à des grappes, et donc qui pourront potentiellement combler le manque de « sujets rares » (Marty *et al.*, 2012), ainsi que des formats comme les diaporamas et les vidéos qui sont en général très mal référencés sur Google Actualités.

9.2.2 Meilleur ennemi n°2 : le référenceur

La sociologie des organisations a montré comment les acteurs « n'acceptent les changements imposés que dans la mesure où ils en comprennent la logique, où ces changements permettent de donner un sens à leur travail » (Bernoux, 2010, p. 12). Nous avons déjà décrit les cas où les journalistes et le référenceur pouvaient ne pas être d'accord. Le référenceur nous a résumé ces divergences en nous expliquant qu'il était possible selon lui qu'il y ait une incompréhension de la part des journalistes quant à son rôle exact vis-à-vis d'eux.

« [Les journalistes] pensent que je suis là pour faire leurs titres, même s'ils ont les outils pour faire deux titres différents. Ils pensent que je suis là pour leur dicter leurs contenus. Et je ne suis pas là pour ça. [...] Ils ont l'impression que je vais pouvoir les contraindre dans leur métier parfois, alors que je suis là pour les aider. Personne ne se méfie vraiment de moi, mais c'est plutôt un "Oh je pense pas que ça va marcher", ou bien un "Non, je ne sais pas si je vais mettre ça comme titre". Ils vont te trouver un tas de trucs pour te dire que c'est pas ça qu'il faut faire. »

Lexpress.fr — Référenceur — Renaud Perrin

Nous voudrions rendre compte ici de ce qui, malgré les divergences et les réticences évoquées en 9.1, motive finalement l'adhésion des journalistes à la plupart des changements proposés par le référenceur : comment ils comprennent ces changements et dans quelle mesure cela est susceptible de donner, à leurs yeux, un sens à leur travail.

Tout d'abord, nous avons vu que le référenceur n'avait pas d'ascendant hiérarchique vis-à-vis des journalistes. Il ne leur donne donc pas d'ordre. Sans aucun doute, ce point est fondamental dans leurs rapports :

« [Les équipes de la Direction de l'acquisition d'audience] nous font des suggestions seulement. Ils ne nous donnent pas d'ordre, et ça c'est important. Nous, on suit les suggestions quand on peut. Mais tu ne vas pas non plus recevoir un mail qui te dis : "Fais comme ci ou fais comme ça, t'as pas le choix". On sait ce qu'on a à faire, et on le fait autant qu'on peut. Et si on n'a pas le temps, ou que la recommandation n'a aucun sens, on ne se fera pas taper sur les doigts. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

« Je ne supporterais pas que [le référenceur] ait le pouvoir de me dire ce que je dois faire ou non, quand, et qu'il soit une espèce de deuxième rédac' chef. Il n'est quand même pas journaliste. Donc je veux bien qu'il me donne des conseils, mais pas des ordres. Il est hors de question qu'il me donne des ordres. »

Lexpress.fr — Journaliste (Société) — Julie Saulnier

Le référenceur lui-même accepte tout à fait cette situation, qu'il voit comme étant bénéfique à son action. En effet, c'est selon lui parce qu'il ne donne pas d'ordre qu'il peut être entendu des journalistes. Et c'est en comprenant leur problème de gestion du temps et leurs craintes vis-à-vis d'une écriture qui deviendrait trop contraignante qu'il peut mettre au point certaines mesures, comme le dédoublement du titre, dans le but de poursuivre ses propres objectifs tout en résolvant les « obstacles-problèmes » des journalistes. Ceux-là sont libres, pour chaque article, de suivre ou non les recommandations du référenceur. Le dépassement de leurs divergences est conditionné par cette liberté. En effet, pour qu'une issue satisfaisante aux deux parties puisse être trouvée, il faut que les deux parties en question puissent agir et penser indépendamment l'une de l'autre.

Nous avons déjà cité de nombreux exemples des arbitrages auxquels est confronté le journaliste entre les activités d'information et de communication. En voici un nouveau, où l'on voit très bien de quelle façon le journaliste demeure libre de « trancher », et ne suit les recommandations du référenceur que dans la mesure où elles « ne compromettent pas son travail ». Il est intéressant aussi de voir que sur ce point le journaliste du web ne se sent pas éloigné de celui du titre papier :

« Toute la difficulté, c'est que tu es pris entre la nécessité d'être lu et la nécessité de bien faire ton boulot. Donc tu jongles systématiquement entre les deux. Parfois, il faut faire évoluer son travail sans se compromettre. Mais quelque part, ce sont des règles qui sont valables partout. Nous, c'est Google. Pour la presse magazine, ça va être la pub et l'espace. En presse

magazine, si tu sais qu'il y a de la pub sur la page de droite, tu peux difficilement faire un article de deux pages, donc tu te restreins pour la pub. La chance qu'on a ici, c'est qu'on a un service dédié à ça, Renaud Perrin, Marc Bouguié, etc., chargé de l'acquisition d'audience. Ils vont nous trouver les mots-clés, des sujets, des trucs comme ça. Après, c'est à nous de trancher. Si on estime que ça ne compromet pas notre boulot, on le fait. Sur "Euro 2012", je vois aucun souci à mettre "Championnat d'Europe", ou alors pour les présidentielles par exemple, j'ai aucun souci à marquer "Elysée 2012", plutôt que "Présidentielles". Mais s'ils me disent en revanche : "Là tu ne fais pas de papier sur Bayrou parce que ça ne marche pas sur Google", et si je juge qu'il y a de l'info, je vais quand même le faire. Jusqu'ici, on a toujours eu raison dans ces moments là. Mon travail n'a jamais été empêché par des logiques d'acquisition d'audience. »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

Toujours selon cette idée de rapport dialectique, les journalistes, quand ils ne voient pas l'intérêt d'une proposition du référenceur, peuvent parfois, plutôt que de lui opposer un refus catégorique, accepter de la suivre pour une durée limitée dans le temps. Ainsi, le temps peut servir de variable d'ajustement à la négociation : une concession est faite du côté du journaliste, mais pour un temps déterminé. Ce sont en général les rédacteurs en chef adjoints et les FPE qui procèdent à ces arbitrages.

« Parfois, [le référenceur] vient nous voir et il nous demande de faire des opérations, des trucs, qui n'ont pas de sens éditorial, donc on ne le fait pas forcément, ou pas longtemps. Par exemple une fois, il voulait qu'on mette en Une du site un truc qui n'était pas dans l'actu du moment. Je crois qu'on l'a fait cinq ou dix minutes et c'est tout. Après, on l'a enlevé. »

Lexpress.fr — Rédacteur en chef adjoint — Thomas Bronnec

« [Le référenceur] me propose de temps en temps d'essayer par exemple de mettre le mot "Direct" à cet endroit dans mon titre en majuscule à l'intérieur, pour voir si ça remonte mieux. Bon, je veux bien tester, mais avec certaines limites. Le jour où il m'a dit d'écrire "EN DIRECT" en majuscules sur la Une, je lui ai dit : "Ecoute, je fais ça dix minutes et pas plus, parce que c'est moche". Je veux bien tester comme ça, mais je ne veux pas non plus travailler uniquement pour Google, et je veux que ma Home Page ressemble quand même à quelque chose. »

Lexpress.fr — FPE — Camille Caldini

Le référenceur nous a expliqué que l'un de ses buts était de disparaître progressivement du quotidien des journalistes. Ainsi, une fois que ces derniers auront assimilé ses recommandations en matière éditoriale, il pourra se concentrer d'autant

plus sur les mises en œuvres techniques. Cette disparition progressive du référenceur semble avoir particulièrement marché par exemple dans le cas du journaliste sportif. Ce dernier nous a expliqué qu'après avoir souvent consulté le référenceur pour comprendre comment il était préférable d'agir, il le voyait beaucoup moins souvent. Cela nous a été confirmé par le référenceur, pour qui les journalistes des rubriques Sport et Culture sont les exemples d'une « mobilisation » réussie.

Pour les journalistes les plus difficiles à convaincre, le référenceur souhaiterait utiliser la mesure d'audience comme un levier d'intéressement, en leur montrant sans ambiguïté l'efficacité des mesures qu'il préconise. Il se servirait ainsi des comptes rendus d'audience comme des inscriptions visant à prouver que ceux qui suivront ses recommandations seront « récompensés ». Cependant, la direction ne donne pas accès à tous les journalistes aux résultats des mesures d'audience, parce qu'elle craint, selon le référenceur, de les mettre en concurrence d'une part, et de permettre à certains journalistes dont les papiers génèrent beaucoup de trafic, et donc de revenus, de négocier leur salaire à la hausse. La mesure d'audience est donc un argument que le référenceur souhaiterait, mais ne peut pas, mobiliser pour renforcer l'attrait de ce qu'il propose aux journalistes.

Les journalistes souhaitent que leurs papiers remontent sur Google, jugeant qu'il s'agit là d'une espèce de récompense pour le travail fourni ainsi que d'un gage de reconnaissance et de légitimité.

« Le journaliste est content quand son papier est lu, c'est aussi un de ses buts, c'est pour ça qu'il va optimiser le titre, mettre des mots-clés, choisir l'angle... parce que Google est sensible à ça. »

Lexpress.fr — Rédacteur en chef adjoint — Thomas Bronnec

« Mon but c'est d'être reconnu sur le créneau où je suis spécialisé. Sur Google, je voudrais que quand quelqu'un tape une requête sur un sujet que j'ai traité, ce soit mes articles qui remontent. Ça c'est une vraie reconnaissance. C'est une vraie expertise. J'ai un peu honte de le dire, mais chaque fois que je fais un article je regarde son comportement sur Google. Je regarde s'il apparaît. S'il remonte bien, je suis content. J'ai l'impression d'avoir fait mon boulot. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

Référenceur de presse aguerrie, Virginie Clève a expliqué sur son blog qu'« aussi compétent soit le référenceur, il ne peut pas tout faire tout seul et les résultats ne seront là que si l'entreprise participe activement à la réalisation des objectifs » (Clève, 2011). Nous avons vu ici quels étaient les mécanismes d'argumentation du référenceur dans le rapport qu'il entretient avec les journalistes, et comment, au nom de quoi, à quelles fins, ils pouvaient dépasser leurs contradictions et agir dans la même direction, à laquelle ils attribuent un sens pourtant différent (le référenceur parle de maximisation du trafic, le journaliste de reconnaissance). Petit à petit, le référenceur se fait entendre. C'est une stratégie de persuasion à moyen et long terme : une conversation où chacune des parties fait de petits pas vers l'autre, sans revenir pour autant sur ce qu'il considère comme étant l'essentiel.

9.2.3 Se préserver et se fixer une limite

La stratégie de référencement n'est pas la priorité des journalistes dont la principale mission demeure attachée au traitement de l'information d'actualité. Dès lors, ils se disent préservés, non seulement par eux-mêmes étant donné leur indépendance, mais également par leur direction et leurs collègues non-journalistes. Un champ d'action qui n'appartient qu'à eux est maintenu dans l'entreprise. Le journaliste politique Matthieu Deprieck nous a expliqué que le chef de rubrique et le rédacteur en chef jouaient un rôle de tampon sur les questions liées au référencement, permettant qu'il ne soit pas directement concerné par les objectifs d'audience et par le levier SEO.

La journaliste Julie de la Brosse nous a expliqué quant à elle que lorsqu'un déclassement survenait, les journalistes n'étaient pas directement impliqués. Les techniciens, le référenceur et le rédacteur en chef essayent de voir à quoi est dû le déclassement, et ce qui peut être mis en œuvre pour y remédier sans demander quoi que ce soit aux journalistes. Nous retrouvons l'idée d'un « rôle tampon » joué dans le but de préserver les journalistes.

« Par exemple, la semaine dernière, on a eu un gros bug sur Google. On a perdu la moitié de notre audience venue de Google, parce qu'on avait changé un truc dans notre CMS. A priori, ça a fait un bug sur Google et on

n'était plus du tout référencés. On a eu une audience en chute libre. Donc la technique venait voir tous les jours le red' chef pour voir ce qu'il se passait. Mais nous les journalistes, on était assez préservés. Ils ne venaient pas nous voir directement. C'est moi qui ai vu qu'on n'était plus référencés sur Google News, alors j'ai demandé s'il y avait un problème et on m'a dit que oui, effectivement, il y avait un problème et que ça allait être réglé. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

En plus d'être préservé dans l'entreprise et de se voir accorder une marge d'indépendance qui leur permet de décider eux-mêmes de leur degré d'implication dans la stratégie de SEO, les journalistes se fixent une limite. Cela rejoint l'observation de Murray Dick, selon lequel les journalistes n'utilisent pas un mot-clé à fort potentiel si celui-ci ne correspond pas à la ligne éditoriale (Dick, 2011, p. 473). Le 20 mai 2010, le rédacteur en chef Eric Mettout publiait sur son blog un billet appelé « Zahia n'a pas écrit ce titre », réagissant à la chronique publiée dans le *New York Times* par David Carr (Carr, 2010 ; Mettout, 2010). Tandis que le chroniqueur américain s'inquiétait de l'influence des moteurs sur l'uniformisation des titres, tout en expliquant que cette influence pouvait être dépassée grâce à des titres autant informatifs qu'incitatifs, Eric Mettout se montrait optimiste en prévoyant que les sites capables de « conjuguer personnalité et efficacité » finiraient par gagner la course au référencement. Lorsque nous l'avons interrogé, il a repris ce propos, et nous a expliqué que selon lui il existait des limites que les journalistes de Lexpress.fr ne franchiraient pas et que, d'autre part, les sites qui n'auraient pas franchi ces limites finiraient par être récompensés par Google. Ainsi, cette « limite » que les journalistes se fixent en matière de SEO est précisément justifiée par l'intérêt qu'aurait Google à privilégier ceux qui ne se seront pas livrés à un SEO trop intensif.

9.3 Deux points d'achoppement

Malgré le rapport dialectique présenté ci-avant, certaines actions effectuées à des fins de référencement gênent les journalistes, et nuisent à ce qu'ils se représentent comme leur rôle et à la qualité de ce qu'ils produisent dans le but de remplir ce rôle. Dans le cas des deux « points d'achoppement » que nous présenterons ici, les journalistes ont subi un programme d'action s'opposant au leur, contre lequel ils ne

pouvaient rien ou pas grand chose, contrairement aux cas dans lesquels nous avons vu qu'ils étaient libres d'arbitrer entre ce qui leur était demandé et ce qu'ils se représentaient eux-mêmes comme la bonne chose à faire.

9.3.1 Des fissures à la Coupole

Certains journalistes de *L'Expansion* et de *L'Entreprise* nous ont dit regretter le rapprochement effectué de l'adresse URL de leur titre avec celui de *L'express.fr*. Comme nous l'avons expliqué (cf. 7.4), ce rapprochement a été effectué par la direction de RMF afin de mutualiser l'impact des sites du groupe sur le marché publicitaire et d'augmenter leur visibilité de l'ensemble des sites sur les moteurs de Google. Cependant, des journalistes, qui n'ont pas été consultés pour ce projet, nous ont dit que cela pouvait brouiller les pistes. Les magazines *L'Expansion* et *L'Entreprise* sont parfaitement différenciés de *L'Express* en kiosque, cependant sur le web ils n'ont pas (plus) d'existence propre, devenus des sous-domaines de *L'express.fr*. Nous avons vu qu'étant donné un tel rapprochement, le nom de *L'Express* apparaît dans Google Actualités, quel que soit le sous-domaine. Comme nous l'a expliqué le journaliste Emmanuel Colombié, de la rédaction de *L'Entreprise*, le titre pour lequel il travaille a une légitimité sur certains sujets, reconnus notamment par les professionnels du monde des PME, alors que *L'Express* ne traite pas, ou peu, ces sujets. Dès lors, la stratégie de Coupole, tout en augmentant la visibilité des articles sur Google, pourrait nuire à la légitimité du titre en général et du journaliste en particulier. Emmanuel Colombié nous a expliqué qu'il ne voyait que très rarement les journalistes de *L'express.fr*, et qu'il estimait travailler exclusivement pour le site web de *L'Entreprise*, pourtant trop peu visible en tant que tel depuis le rapprochement des adresses URL. Le journaliste reconnaît toutefois que la Coupole a abouti à une vraie augmentation de la visibilité sur Google, et que par conséquent, les reproches qu'il lui adresse sont contrebalancés par cet aspect positif.

« Avant, on s'appelait Lentreprise.com. On s'appelle toujours comme ça, seulement maintenant l'adresse c'est Lentreprise.lexpress.fr. Donc on profite de la puissance de L'Express pour gagner en référencement. Ce qui se passe, c'est quand on fait un article qui marche bien, quand il remonte sur Google et sur Google Actu, il n'y a pas écrit en dessous L'Entreprise mais L'Express. Donc on a le sentiment de bosser pour L'Express alors qu'en

réalité on travaille pour L'Entreprise. C'est des combines ça tu vois. Pour l'identité de la marque, je ne suis pas sûr que ce soit une très bonne chose, parce que le lecteur il voit L'Express, alors que ce n'est pas la rédaction de L'Express mais la rédaction de L'Entreprise qui a fait ce papier là. Après, c'est bien parce que ça nous fait gagner des visiteurs, donc au final c'est peut-être positif. Mais c'est un exemple d'un truc qui n'a été fait que pour le référencement tu vois, et pas du tout pour les journalistes. Nous on a une expertise PME que L'Express n'a pas, donc il faut qu'on reste sur notre créneau. Mais en même temps sur le référencement ça ne peut que nous aider, donc on ne va pas se plaindre. »

Lentreprise.lexpress.fr — Journaliste — Emmanuel Colombié

La ligne éditoriale des rédactions dont les articles sont intégrés à la Coupole peut différer de la ligne éditoriale de Lexpress.fr. Cela n'est pas forcément évident à détecter pour les lecteurs qui naviguent parmi des sous-domaines différents, et pourrait nuire selon certains journalistes à la cohérence de la ligne éditoriale du sous-domaine, autant qu'à la cohérence de la ligne éditoriale de Lexpress.fr.

« On a deux visions qui ne sont pas les mêmes entre Lexpansion.com et Lexpress.fr. A L'Express, il y a un rédacteur en chef qui est hyper web, qui veut du buzz, de l'info en continu et surtout de la réactivité. A L'Expansion, au contraire, c'est l'héritage du papier. Le rédacteur en chef est plutôt un vieux de la vieille dans le type journaliste. C'est le genre à revérifier l'info qui est balancée sur l'AFP, tu vois ? Il revérifie tout le temps. Et lui, ce qu'il veut, c'est du gros papier où t'as toutes les infos dedans. [...] Les deux rédacs chef de L'Express et de L'Expansion ont des personnalités très différentes, et ça peut rentrer en conflit puisque L'Expansion maintenant est intégrée à L'Express. Les deux visions s'embrouillent l'une l'autre, genre strabisme, et le lecteur ne voit plus clairement où est la ligne éditoriale de qui. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Julie de la Brosse

Une autre journaliste de *L'Expansion* nous a expliqué qu'elle avait effectivement trouvé étrange que le magazine auquel elle était affiliée n'ait plus de site dédié. Cependant, elle nous a dit comprendre qu'une telle stratégie était liée au référencement et à la maximisation du trafic, indispensables à la fois à la survie de l'entreprise et à la visibilité de son travail. Elle trouve donc la fusion légitime mais regrette qu'elle n'ait pas été accompagnée de la création pour le sous-domaine de *L'Expansion* d'un « univers dédié » différencié de celui de *L'Express*.

« L'Expansion a clairement été intégré au site de L'Express pour le référencement. Au début, c'était quand même bizarre pour moi de voir qu'on n'avait plus de site attribué. Mais en fait, c'est mieux pour le référencement. Donc finalement ça ne me dérange pas trop, ou plus trop. Les papiers, les dépêches qu'on peut faire, Google va les référencer comme Lexpress.fr, c'est-à-dire que quand tu vois un de nos papiers sur Google, ça apparaît comme L'Express, pas comme L'Expansion. Mais ça a un vrai intérêt, et la fusion est légitime, donc tant mieux. Enfin je veux dire, finalement, ça ne me dérange pas. Mais bon quand même j'aurais voulu qu'on soit L'Express Eco dans ces cas-là, avec un site dédié à l'économie. Tu vois comme L'Express Style et L'Express Culture, où ça t'emmène sur un univers dédié. C'est pas juste des rubriques. J'aurais aimé avoir ça pour l'éco. »

Lexpansion.lexpress.fr — Journaliste (Economie) — Emilie Lévêque

Nous voyons ici que c'est parce que la journaliste comprend pourquoi la Coupole a été mise en place qu'elle accepte les changements induits. Et nous voyons comment, en revanche, elle voudrait que d'autres changements soient faits pour contrebalancer la perte d'identité.

Le rédacteur en chef de *L'Expansion* considère quant à lui que les effets de la Coupole sont positifs. Il accepte que le nom du SIPI soit rapproché de celui de *L'Express* dans l'adresse URL et confondu avec lui sur Google, dans la mesure où l'impact en matière de trafic apporté par Google est fort. Ainsi, la concession faite en matière d'identité vaut à ses yeux la récompense obtenue en termes de trafic.

« Une fois rattachés à L'Express, on peut espérer être mieux référencés dans Google News. Il y a eu un vrai impact là-dessus. Ça a été gagnant pour nous : on a eu accès à une audience à laquelle on n'avait pas accès auparavant. Dans une vie antérieure, L'Expansion a été blacklisté par Google News. Ça nous faisait bien grogner parce qu'on voyait remonter des dépêches sans intérêt alors qu'on avait des papiers à valeur ajoutée sur ces sujets. Donc il y avait un déficit d'exposition. Et c'est ce que nous a donné la Coupole : une voie d'accès plus sûre à Google News. »

Lexpansion.lexpress.fr — Rédacteur en chef — Yves Akaden

La stratégie de Coupole est selon nous un bon exemple de la manière dont la question du référencement peut influencer aussi bien l'organisation de presse que le dispositif de publication et, ici en particulier, jusqu'à l'identité d'une rédaction. Nous évoquerons ci-après un second point d'achoppement, où cette fois le changement critiqué par les journalistes n'a pas été opéré spécifiquement pour le référencement, mais est critiqué étant donné précisément ce qu'il induit en matière de référencement.

9.3.2 *Express Yourself* et les poissons-pilote

Nous avons présenté dans le chapitre 3 le dispositif de participation « Express Yourself » (cf. 3.3.3), qui consiste à inviter certains collaborateurs extérieurs — spécialistes ou amateurs, anonymes ou non — à publier des articles sur les pages de Lexpress.fr après validation des journalistes. Nous avons également parlé des réticences parfois très virulentes des journalistes vis-à-vis de ce dispositif. Or, il se trouve que pendant nos interviews, le fait de mentionner « Express Yourself » soulevait de vives réactions car les articles des rédacteurs extérieurs apparaissaient sur Google sous la même forme que les articles écrits par les journalistes.

Voici par exemple comment apparaît dans la rubrique « Education » du site lexxpress.fr un article publié le 28 juin 2013 à propos d'une mesure annoncée en France par le Ministre de l'Education Vincent Peillon, et rédigé par « Guy2Mop ». L'internaute sait simplement que l'auteur « Guy2Mop » est enseignante (cf. chapeau), mais l'auteur demeure anonyme : son profil, que l'auteur peut pourtant renseigner, est vide (dernière vérification : 30 juin 2013).

Figure 47. Capture d'écran : Express Yourself (28/06/13)



Source :

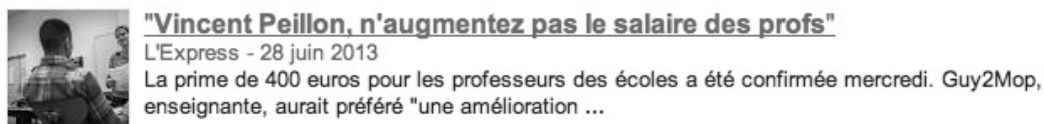
http://www.lexpress.fr/actualite/augmenter-le-salaire-des-profs-n-est-pas-une-priorite_1261770.html

Nous voyons en haut à gauche le logo « Express Yourself » et en haut à droite le logo « Vérifié et mis en Une ». A part ces logos, et la parenthèse près du nom de l'auteur, rien ne différencie la page d'un article publié par un journaliste. Comme nous le verrons dans les verbatim ci-après, les journalistes craignent que les lecteurs n'étant pas habitués au site, et en particulier les lecteurs arrivés depuis Google, ne fassent pas attention au logo et croient, par conséquent, que c'est un journaliste de *L'Express* qui a traité l'information. Voici à présent comment ce même article apparaîtra sur Google Search et Google Actualités :

Figure 48. Captures d'écran : Express Yourself sur Google

"Vincent Peillon, n'augmentez pas le salaire des profs" - L'EXPRESS...
www.lexpress.fr/actualite/augmenter-le-salaire-des-profs-n-est-pas-une-p... ▾
 Il y a 2 jours – "Vincent Peillon, n'augmentez pas le salaire des profs". Dans
 l'Education nationale, "le fond du problème ne se trouve pas dans les salaires", ...

Google Search



Google Actualités

Cette fois-ci, plus rien ne différencie le lien vers le contenu effectué par un contributeur extérieur à la rédaction des liens effectués vers les contributions des journalistes. En fait, le crawleur n'a aucune façon de les différencier étant donné que les formats de leurs adresses URL sont similaires. Les articles « Express Yourself » jouissent par conséquent du même référencement sur le dispositif Google que les articles écrits par les journalistes. Nous avons vu d'ailleurs dans le chapitre 8 que les articles de « Express Yourself » étaient rangés de la même façon que ceux de la rédaction dans ce que nous avons appelé le « sitemap à tiroirs » (cf. 8.4.2).

Lors de nos entretiens, le dispositif « Express Yourself », d'abord consacré aux seuls contenus liés à la musique, à la culture et à la mode, était en train d'être étendu aux autres rubriques du site. Le sujet était par conséquent brûlant, et les

journalistes très virulents à son endroit. Certains nous ont dit qu'ils trouvaient scandaleux que des amateurs, parfois anonymes, jouissent de leur autorité et de leur visibilité. Métaphoriquement, les amateurs jouent les « poissons-pilotes », profitant de leur arrimage au « grand requin » pour bénéficier de sa visibilité et de la crédibilité qu'on lui accorde. S'ils avaient créé leurs propres blogs, ils auraient été beaucoup moins visibles sur Google Search et n'auraient sans doute jamais intégré Google Actualités. Par ailleurs, les journalistes craignaient que, sous couvert de pseudonymes, des partis politiques désireux de faire campagne ou bien des entreprises de vente, ne vinssent profiter également de la crédibilité et de la visibilité de Lexpress.fr³¹⁰. Il s'agit clairement, pour les journalistes opposés au projet, d'un péril pour leur propre légitimité et d'une atteinte au rôle social joué par *L'Express*. Parmi eux, le plus vindicatif était le Rédacteur en chef du service Monde du magazine *L'Express*. Nous retranscrivons ci-dessous une partie seulement de la longue tirade qu'il nous a adressée à l'encontre de « Express Yourself » :

« Je comprends l'intérêt qu'il y a à faire participer des lecteurs. Le Plus de l'Obs, après tout, c'est pas mal, mais c'est une URL différente. Alors que nous, c'est bien pire : c'est la même adresse URL, et quand vous êtes sur un document de « Express Yourself », vous avez juste un petit logo « lu et validé ». Qui voit ce logo ? Un gars qui ne nous lit pas souvent et qui arrive par Facebook ou par Google, vous croyez qu'il va le voir le logo ? Il ne le verra pas. Et si l'article est nul, il pensera qu'on est nul. Je trouve que c'est potentiellement très emmerdant pour Lexpress.fr, mais je trouve aussi que c'est très emmerdant pour L'Express tout court. Je ne réagis pas comme ça par corporatisme. Je crois qu'il y a un certain nombre de choses qu'on attend malgré tout d'un journaliste, et notamment qu'il exerce sa profession, c'est-à-dire qu'il vérifie, qu'il croise les sources. Au-delà de ça, il y a tout simplement le fait qu'à partir du moment où on écrit soit pour l'hebdo soit pour le site web, on peut vous demander des comptes. Je peux être sanctionné parce que je me suis planté. Je peux avoir une mauvaise réputation. Il y aura des conséquences. Ce n'est pas pas le cas des contributeurs extérieurs, je veux dire pour les anonymes, parce que si le gars est prof et qu'il signe avec son nom, c'est différent. Mais si je peux signer

³¹⁰ Ces craintes ont été confirmées très récemment lorsqu'il a été avéré, le 9 septembre 2013, que des entreprises de communication s'étaient saisies des tribunes libres de *L'Express*, des *Echos*, de *Capital*, du *Nouvel Observateur* et du *Journal du Net*. En se faisant passer pour des amateurs, ces entreprises ont profité de ces dispositifs de participation pour faire la publicité de produits. « Contournant les règles imposant d'intervenir sous sa véritable identité, nombre de prestataires ou de services de communication ont choisi d'opter pour l'intox à grande échelle. Ou profitent des audiences de ces sites médias pour doper leur référencement en plaçant des liens pointant vers les pages de leurs clients sur des mots-clés précis. » Source : <http://www.journaldunet.com/ebusiness/crm-marketing/les-pros-de-la-e-reputation-infiltrent-les-medias-web.shtml> (dernière visite le 19 septembre 2013).

avec un pseudo, donc disons que je m'appelle Gros Lapinou, je peux raconter n'importe quelle connerie et il n'y aura pas de conséquence, à part pour nous : les journalistes. [...]

Nous créons du bruit et nous mettons en danger. Imaginez que Gros Lapinou est dans une boîte qui a pour mission de faire la promo d'un produit. Si ça se trouve, Gros Lapinou bosse même pour l'UMP, qu'est-ce que j'en sais ? Il est dans la cellule « riposte » de Sarkozy. Jusqu'ici, il pouvait écrire dans les commentaires. Mais à partir du moment où ces gens ont accès à l'écriture de papiers, ils bénéficient de notre légitimité. Et ils vont abîmer notre légitimité. La légitimité, ça existe quand on la défend. Sur Google, par exemple, tant qu'on n'a pas cliqué, on a aucun moyen de savoir qui, comment, quoi. On a juste « L'Express ». Et Gros Lapinou se retrouve en haut des résultats, comme moi, comme nous, avec le nom de l'entreprise dans laquelle j'ai bossé pendant vingt-six ans, et il peut abîmer ce nom, ce média, mon travail, ma crédibilité, la nôtre à tous, aussi bien au papier qu'au web. »

L'Express (papier) — Rédacteur en chef du service Monde — Marc Epstein

Nous retrouvons dans cette citation chacun des points évoqués plus haut à propos des « poissons-pilotes » anonymes profitant de l'autorité et de la crédibilité du « grand requin ». Nous y voyons également une différenciation faite entre l'expert, dont le journaliste encourage la collaboration, et l'anonyme. Et nous trouvons une réelle inquiétude du « requin » quant au devenir de sa visibilité et de sa crédibilité.

Dans la rédaction web, nous n'avons pu interroger hélas que deux journalistes au sujet de « Express Yourself », car nous avions déjà interrogés les autres lorsque l'ouverture du dispositif à toutes les rubriques fut rendue publique (fin janvier 2012). Même s'ils étaient moins virulents que Marc Epstein, ces journalistes nous ont également fait part d'inquiétudes.

« Je suis okay si ce sont des mecs qui t'apportent réellement de l'info, ce qui pourrait être le cas par exemple en politique : un blogueur local par exemple, qui est bien implanté dans la politique du coin et qui va nous faire un papier sur la restauration d'une façade dans sa rue qui tourne au combat politique entre gauche et droite. Ça moi ça m'intéresse, parce qu'on ne l'aurait pas vu, donc très bien. Le type nous écrit, et à nous de prendre contact avec lui. Dans ce cas, on a un papier qui pour moi mérite d'être dans L'Express, et d'être identifié comme tel sur Google. On va bosser avec le mec, presque comme avec un correspondant, et le mec sera content parce qu'il aura médiatisé son truc, et nous on sera contents parce qu'on aura une vraie info qu'on aurait loupée dans tous les cas. Mais si c'est un anonyme, qui veut écrire à propos de politique, et qu'il n'apporte rien, et qu'il répète ce que les autres disent, ce qu'on pourrait très bien faire nous-mêmes avec

une dépêche, je ne vois pas l'intérêt. Ou alors, le mec dit des conneries et donne son avis, dont nos lecteurs se fichent. Pourquoi ferait-on appel à lui ? Et pourquoi servirait-on de vitrine sur Google à ces conneries ou à ces banalités ? »

Lexpress.fr — Journaliste (Politique) — Matthieu Deprieck

« A vrai dire, j'appréhende un peu. Peut-être que les articles d'Express Yourself seront happés par le référencement, et peut-être que ça ne se passera pas. Il y a beaucoup d'inquiétudes là-dessus. [...] Les journalistes du papier sont furieux, et ils ont raison à mon avis. Moi-même il y a des choses qui ne me plaisent pas. Je ne suis pas contre le fait que les amateurs écrivent, mais pas comme ça, pas avec la même URL. Il faut voir ça comme un enjeu d'audience : des papiers supplémentaires, c'est plus d'audience. Ils ne se disent rien d'autre au niveau des éditeurs. Les amateurs vont avoir une audience significative parce qu'ils seront référencés par la machine L'Express sur Google notamment, et Google News en plus. Sur Google News ! C'est le rêve pour eux.

On verra, une fois que ce sera là, comment on va s'ajuster. On peut regretter ce fonctionnement mais bon, moi je n'ai pas le choix là-dessus. »

Lexpress.fr — Journaliste (Sport) — Clément Daniez

Cadre 4. Les poissons-pilotes de Mediapart

Il nous a été raconté une anecdote intéressante par un journaliste de Mediapart, alors que nous menions l'enquête préliminaire à ce travail. Même si, pour des raisons que nous avons déjà expliquées, nous avons finalement décidé de ne pas inclure les sites nés en ligne dans notre terrain, cette anecdote mérite à notre avis d'être rapprochée du cas des « poissons-pilotes » de *L'Express*. Le journaliste de Mediapart Vincent Truffy, interrogé le 10 septembre 2011, nous a expliqué que les articles des journalistes de Mediapart, indexés sur Google Search et sur Google Actualités, avaient un classement « désastreux », car ils n'étaient pas optimisés d'une part, et, d'autre part, qu'ils étaient payants. En revanche, les articles du « Club », c'est-à-dire les articles publiés par les abonnés sur leurs blogs hébergés par Mediapart remontent dans les classements parce qu'ils sont, eux, gratuits, et qu'un modèle unique d'adresse URL ne permet pas aux crawlers de Google de les différencier des articles publiés par les journalistes. D'autre part, les contenus publiés sur le « Club » par des abonnés ne sont supprimés que dans le cas où ils contreviennent à la loi. Autrement dit, les propriétaires de blogs disposent d'une totale liberté d'expression quant à la publication de leurs opinions et de leurs réflexions. Les internautes qui ne sont pas abonnés au site peuvent avoir l'impression selon Vincent Truffy, quand ils y accèdent via le dispositif Google, que « Médiapart, ce n'est que ça ». Mais le journaliste nous a

exposé un cas qui a finalement obligé les dirigeants et les journalistes de Mediapart à modifier leur site, c'est-à-dire à effectuer un « détour » dans le but de corriger une situation qui n'avait d'abord pas été prévue.

Philippe Val, alors chroniqueur sur *France Inter*, a pointé du doigt un jour la faible qualité de Mediapart en citant l'article d'un auteur d'après lequel le Sida ne serait pas causé par le virus VIH mais par une mauvaise hygiène de vie. Cet article n'avait cependant pas été rédigé par une journaliste mais par un membre du « Club Mediapart » et ne contrevenait aucunement à la loi. Philippe Val a donc fait une confusion entre le contenu produit par les journalistes et le contenu produit par les abonnés, sûrement selon Vincent Truffly parce qu'il y avait eu accès *via* Google. A la suite de cette « mésaventure », les journalistes de Mediapart ont demandé à Google de différencier les articles du « Club » des articles de la rédaction, sans pour autant déréférencer le Club (ne souhaitant pas *punir* leurs abonnés). Mais Google a répondu qu'une telle opération était impossible, puisque les adresses URL des blogs et celles de la rédaction n'étaient pas différenciées. Mediapart a par conséquent révisé les adresses URL du Club, qui ont pris la forme « <http://blogs.mediapart.fr/...> », de façon à ce que Google puisse explicitement afficher dans le nom de source « Mediapart » pour les articles de la rédaction et « Club Mediapart » pour les articles des blogs. Cependant, Vincent Truffly nous a expliqué que des récriminations d'abonnés avaient alors été adressées à Mediapart, dans la mesure où la présence du mot « Club » dans les résultats de Google et du mot « blog » dans les adresses URL leur ôtaient la légitimité dont ils avaient joui jusque là (et pour laquelle, en partie, ils avaient payé). En effet, certains des blogueurs « poissons-pilotes » étaient précisément intéressés par l'ambiguïté qui pesait sur leurs contenus auprès des lecteurs, et qui pouvait les faire passer pour de « vrais journalistes ».

Ce cas est intéressant à rapprocher d'Express Yourself, car nous retrouvons l'idée d'une légitimité journalistique travaillée par la présence de rédacteurs extérieurs. Dans le cas de Mediapart, cette question a abouti à une différenciation des adresses URL et aurait provoqué, selon Vincent Truffly, le départ de certains « poissons-pilotes ». Après avoir mis un terme à leur abonnement, ces derniers sont peut-être partis trouver d'autres « requins » dont ils pourraient pleinement jouir de la puissance et de la légitimité. Peut-être Express Yourself...

CONCLUSION PARTIE 3

Nous avons voulu montrer dans cette partie, au niveau mezzo d'une part, pour des SIPI français comparables, puis dans le cas de Lexpress.fr en particulier, comment la question du référencement sur le dispositif décrit dans la Partie 2 pouvait travailler à la fois l'organisation de presse, la coordination des métiers, le site lui-même, sa structure, l'emploi du code HTML, l'apparence des pages, la mise en forme de l'information, le maillage hypertexte et, finalement, les tensions entre des représentations et des intérêts divergents et la manière dont ces divergences pouvaient être dépassées. Nous avons décrit les frictions autant que les ajustements, les luttes autant que les ententes, les points d'achoppement, les mutations, les concessions et les choix effectués par les journalistes étant donné leur marge de manœuvres.

Nous avons vu que le profil du référenceur pouvait parfois ne pas être directement lié à des activités techniques, mais qu'au contraire certains référenceurs avaient des profils commerciaux ou littéraires étant donné un champ d'action triple : technique, marketing et éditorial. Même s'il est mieux payé que la plupart journalistes, le référenceur n'a pas d'ascendant hiérarchique à leur endroit et doit par conséquent les persuader qu'ils auront raison de suivre ses recommandations. De sa « compétence particulière » et de sa « spécialisation fonctionnelle » (Crozier et Friedberg, 1977, p. 83), il tire une forme de *pouvoir* dans l'organisation, qui n'est jamais autoritaire ou incontournable. Au contraire, ce pouvoir est de l'ordre de la conduite des conduites d'acteurs dont le référenceur doit obtenir la collaboration par d'autres moyens que par des ordres formels. Il tente ainsi de *traduire* son projet dans la « langue » des autres, et de rendre compatibles les actions qu'il propose avec les buts poursuivis par ses partenaires, sans quoi il échouera. Il le fait quotidiennement, au cas par cas, en s'immisçant dans les projets techniques ou éditoriaux dès leur lancement afin d'y participer selon une méthode que nous avons qualifiée « d'agile », et en formant les journalistes, en leur répondant, en les sensibilisant dans le but que les actions qu'il préconise deviennent chez eux des réflexes et qu'il puisse ensuite disparaître de la salle de rédaction et se concentrer sur les aspects techniques de son activité. Dans ces conditions, l'un des buts poursuivis par le référenceur est de faire en sorte qu'il soit facile pour les journalistes de suivre ses recommandations. C'est la

raison pour laquelle le référenceur de Lexpress.fr a injecté l'outil « Tendance de recherche » dans le système de gestion de contenus. Les journalistes peuvent ainsi coopérer sans que cela ne leur prenne trop d'un temps précieux, et choisir certains sujets ou certains mots recommandés sans avoir à visiter le site web « Tendances de Recherche ».

Par ailleurs, nous avons vu que les référenceurs de la presse en ligne étaient des « chapeaux gris », c'est-à-dire qu'ils ne suivaient pas forcément les recommandations de Google, sans pour autant se livrer à des procédés intensifs de « *spamming* ». Ils développent ce qu'ils nomment eux-mêmes une éthique, c'est-à-dire qu'ils fixent une frontière entre ce qui est acceptable et ce qui ne l'est pas en matière de SEO. Enfin, nous avons vu que le métier du référenceur pouvait comporter une part de « hors-ligne » : il téléphone à Google et négocie de vive voix avec les employés de la firme lors d'un déclassement ou d'une modification d'algorithme. Jouant un rôle de tampon entre Google, la direction, le service technique, les équipes commerciales et les journalistes, le référenceur est ainsi un « centre de traduction » en contact avec l'ensemble des porte-parole du réseau, capable d'actionner, directement ou indirectement, les leviers influençant le processus de communication. Si le référenceur peut, avec plus ou moins d'efficacité, actionner ces leviers, c'est précisément « parce que le réseau sociotechnique existe, fournissant les lignes d'action possible et autorisant leur accomplissement » (Callon, 2006, p. 270).

Nous nous sommes ensuite intéressés à l'effet tangible de la manipulation des leviers d'action, afin de juger de la performativité des recommandations de Google et de celles du référenceur, et de voir comment les éditeurs faisaient en sorte que leur propre production performe. Nous avons décrit et analysé chez six SIPI comparables ce qu'il y avait « en façade » (ce que l'internaute *voit*) et ce qu'il y avait « derrière » (ce que le navigateur et le moteur *traduisent*). Nos observations nous ont appris d'une part que de nombreux leviers architextuels — notamment les balises méta-tags et les microdonnées, dont l'utilisation était souvent simpliste et incomplète, c'est-à-dire permissive — n'étaient pas mobilisés. Ainsi, il semble que les éditeurs « abandonnent » (ou *délèguent*) une partie du pouvoir qu'ils pourraient exercer quant à la diffusion de leurs contenus. Les *Sitemaps*, en revanche, sont largement et

finement utilisés par les éditeurs, qui réussissent même à contourner la limite imposée par Google de 1000 adresses URL par *Sitemap* et développent des stratégies de valorisation des archives grâce à ce que nous avons appelé les « *Sitemaps à tiroirs* ». D'autre part, nous avons vu avec les stratégies de *PageRank Sculpting* comment étaient déployées des outils de création automatique de liens visibles ou invisibles vers des pages d'atterrissage à des fins de référencement. Ce levier est massivement actionné, et ce en dépit des annonces des porte-parole de Google selon lesquelles cela ne servirait (presque) à rien. Le *PageRank Sculpting* est opéré pour augmenter la visibilité de la production journalistique mais aussi, dans le cas des conjugueurs, pour redistribuer l'autorité du contenu journalistique à des contenus à faible valeur ajoutée mais générant des revenus. Nous avons vu comment et dans quelle mesure l'automatisation de la production de liens pouvait diluer l'autorité des journalistes et même, dans certains cas extrêmes, entrer en conflit avec l'édition manuelle par les journalistes de liens hypertextes. Par ailleurs, nous avons vu comment les SIPI pouvaient se positionner dans l'arbre de navigation décrit dans le chapitre 5. En effet, les SIPI génèrent des revenus via DoubleClick et/ou AdSense qu'ils partagent avec Google. Ils se situent ainsi du côté des éditeurs que Google aurait intérêt à avantager dans le cas d'un biais, tandis que certains sites affichent même des publicités *pour* les produits de Google, ce qui augmente encore le poids des incitations.

Pour ce qui est des journalistes, nous avons vu avec notre monographie que, sur de nombreux points, comme par exemple la pyramide inversée, leur action était compatible avec l'action du référenceur. Dès lors qu'une recommandation a d'après eux une justification éditoriale, c'est-à-dire dès lors qu'elle peut être *traduite* dans leur propre « grammaire d'action », ils adhèrent à la stratégie de SEO. Finalement, les journalistes sont prêts à laisser en partie leur conduite être conduite (à se *mobiliser*), mais à condition que cela conserve une valeur éditoriale selon des critères qui leur sont propres, que la possibilité d'arbitrer demeure de leur côté et qu'ils puissent continuer à traiter les sujets qui leur semblent importants même si leur trafic potentiel est faible. Dans le cas où une recommandation du référenceur s'oppose au projet et aux représentations des journalistes, ceux-là sont libres de refuser de la suivre, soit parce qu'elle leur semble injustifiée, soit parce qu'elle est trop chronophage. Dans les

faits, quand ils ne trouvent pas justifiée une préconisation du référenceur, ils coopèrent tout de même partiellement : soit ils suivent la recommandation pour un temps limité, soit ils la suivent de manière incomplète. Finalement, étant donné ce que nous avons nommé un « rapport dialectique », les divergences entre les journalistes et le référenceur, ainsi qu'entre les journalistes et Google, réussissent à être dépassées, dans la mesure où les journalistes parviennent à justifier dans leurs propres termes, et selon leurs propres valeurs, leur adhésion à la stratégie de SEO. Ils peuvent également s'appuyer sur certains « centres de traduction », notamment les FPE, dont une des fonctions consiste à décharger partiellement les journalistes de ce qu'une telle adhésion implique. Il est toutefois nécessaire de souligner que les journalistes se fixent une limite qui n'est pas la même que celle du référenceur quant à ce qui est acceptable ou non. D'autre part, nous avons vu avec l'exemple du dédoublement du titre comment des espaces étaient aménagés par le référenceur dans le but que les journalistes y soient libérés des pressions relatives au SEO. Même si une forme de pression peut malgré tout demeurer dans ces espaces, notamment en raison du temps supplémentaire qu'il s'agit d'y consacrer, de telles modifications du CMS font office de « lignes de fuite » le long desquelles les actions des journalistes ne sont plus concernées par le dispositif. Le référenceur peut donc remplir une fonction qui ne consiste pas à faire pression sur les journalistes pour des questions de SEO, mais au contraire à les en libérer.

En outre, nous avons vu que certaines mesures étaient prises par la direction sans que l'avis des journalistes ne soit sollicité, et pouvaient aboutir à un sentiment de perte d'identité ou de menace vis-à-vis de leur légitimité. Enfin, il ressort de ce dernier chapitre un point fondamental au sujet de la ligne éditoriale. Celle-ci n'est pas modifiée directement pour des questions liées au référencement mais écartelée entre les sujets qui auraient été traités de toute façon, les sujets qui n'auraient pas été traités sous cet angle si la question du référencement avait été ignorée et les sujets qui n'auraient tout simplement pas été traités. Ainsi, la ligne éditoriale définie par les journalistes a beau continuer d'exister, certains regrettent qu'elle soit diluée parmi des sujets qu'idéalement ils n'auraient pas traités de cette façon ou qu'ils n'auraient pas traités du tout.

Conclusion

Nous avons déployé le dispositif puis réassemblé ses composants autour de ce que nous observions en ligne et de ce que nous en disaient les acteurs, dans le but de jauger la performativité de son énonciation et de sonder les rapports de pouvoir en son sein. Nous reprendrons ici les différentes dimensions conceptuelles du dispositif pour dégager les lignes de notre conclusion, après quoi nous reviendrons sur nos hypothèses et notre problématique. Nous présenterons enfin les limites de notre contribution ainsi que les perspectives susceptibles de nourrir de futurs travaux.

Déploiement du dispositif

Nous avons présenté dans notre premier chapitre le dispositif comme une grille de lecture analytique utilisée pour saisir la composante instrumentale d'une situation de communication en caractérisant l'hétérogénéité des éléments, la nature des liens entre ces éléments, les déplacements stratégiques parmi ces éléments et leurs résultats : un faire dire, un faire voir, des forces et des lignes de fuite. Nous avons ensuite mobilisé ces notions dans le but d'établir « la rationalité à l'œuvre dans la coordination des relations et des médiations qui s'établissent entre les différents éléments [...] et la manière dont elle contraint les acteurs selon leur situation au sein du dispositif » (Monnoyer-Smith, 2013). Nous cherchions ainsi à comprendre comment le dispositif d'infomédiation pouvait faire « fonctionner des relations de pouvoir dans une fonction, et une fonction par ces relations de pouvoir » (Foucault, 1975, p. 208), en même temps que nous cherchions à sonder les intentions concrétisées à travers la mise en place d'« environnements aménagés » (Peeters et Charlier, 1999, p. 18). Nous reprendrons ici les dimensions conceptuelles du dispositif pour montrer à quelle compréhension notre analyse aboutit.

Rencontre d'éléments hétérogènes

La nature et le résultat du processus d'infomédiation dépendent de l'interaction de l'infrastructure Internet, du code informatique, des protocoles de communication, des crawlers, d'un algorithme d'appariement, d'un indexeur, d'une interface, des concepteurs, des journalistes, des internautes, des annonceurs et des référenceurs. Joignant ces éléments humains et non-humains en un même réseau hybride, le dispositif produit un énoncé en traduisant, c'est-à-dire en interprétant et en déplaçant, leurs actions sous forme de liens hiérarchisés et éditorialisés. Son action se déroule en plusieurs étapes, sans que les relais qui en assurent le déroulement n'aient la même qualité matérielle ni la même charge intentionnelle. Les crawlers de Google consomment les ressources des éditeurs, les éditeurs optimisent leurs sites, les internautes utilisent les outils de Google, les annonceurs paramètrent des enchères, l'indexeur compile et ordonne les informations, l'algorithme les apparie, l'interface donne corps à l'énoncé, l'internaute se déplace et interagit avec cette interface, tandis que Google, les éditeurs et les annonceurs tentent de mesurer et de monétiser le trafic. C'est ainsi que, lors de la *médiation sociotechnique*, les *programmes d'action* se rencontrent, que des *interférences* ont lieu, et, finalement, que des *compositions*, ou *recompositions*, aboutissent à l'émergence d'un produit éditorial déterminé : la liste de liens générée à un instant t , pour un internaute i , en un lieu l . Outre le fait que les éléments agencés pour produire la liste (*inputs*) soient hétérogènes, il se trouve que cet agencement lui-même (*output*) file entre les doigts de l'analyste. Les caractéristiques de la liste hypertexte changent notamment selon le point de vue adopté : bien public du point de vue de l'utilisateur, la liste est constituée par un ensemble de positions privatives du point de vue des éditeurs et des annonceurs. De surcroît, du point de vue de tous, et de l'analyste lui-même, on n'en connaît la valeur qu'après l'avoir consommée et elle n'existe qu'à condition d'être expérimentée. Par ailleurs, le matériau informationnel constitutif de la liste a été produit par les éditeurs et mis en mémoire sur leurs serveurs, puis recopié par les crawlers de Google et mis en mémoire sur les serveurs de la firme, et finalement transformé sous forme de lien et affiché sur le terminal d'un internaute. La liste hypertexte provient donc de trois lieux, procède de trois champs d'action et résulte de trois moments, sans

être réductible à un seul de ces lieux, un seul de ces champs, un seul de ces moments. Cela fait d'elle un produit éditorial évanescent, pour lequel les notions de valeur, de qualité, de propriété et de temporalité semblent inopérantes.

En ajoutant le raisonnement économique à l'analyse dispositionnelle, nous avons montré comment il existait des incitations susceptibles de « peser » sur le processus d'infomédiation, et comment les liens de la liste ne pointaient pas seulement vers des pages, mais également vers des « ensembles navigationnels », c'est-à-dire vers différentes possibilités de navigation. Pour l'internaute, cliquer sur un lien généré par le dispositif, c'est faire un choix, et sortir du dispositif pour aller vers d'autres choix possibles, d'autres dispositifs, d'autres incitations, d'autres lignes de code, d'autres rapports de pouvoir. Nous avons ainsi présenté le web sous les traits de ce que l'on nomme en théorie des jeux un « arbre de décision », et montré que l'économie avait une dimension communicationnelle dans la mesure où les incitations pouvaient influencer la nature et le résultat du processus de communication et l'« être ensemble » activé par le processus. Les chemins de navigation, les incitations qu'ils sous-tendent et les potentialités qu'ils ouvrent font ainsi partie du produit éditorial. La liste ne pointe pas seulement vers des opportunités de navigation : elle les *crée*.

Rendre compte de l'hétérogénéité des éléments et de l'évanescence de leur agencement nous a permis d'aborder l'infomédiation en considérant ce qui pouvait l'influencer et ce qui pouvait permettre que cette influence s'exerce — à la fois l'action, *ce qu'elle* suscitait et *ce qui* la suscitait. Cependant, il ne suffisait pas de faire état de l'hétérogénéité et du caractère *non-isotropique*, *non-synchrone* et *non-synoptique* de l'interaction communicationnelle, mais également de savoir à quel point les éléments identifiés pouvaient infléchir sa nature et son résultat, c'est-à-dire comment chaque élément pouvait entrer en résonance avec d'autres éléments et les influencer ou être influencé par eux. Car l'interaction dont la liste est le résultat est également *non-isobarique* : « Certains participants imposent leur présence avec force, exigent d'être entendus, pris en considération, tandis que d'autres ne sont plus que des routines relayées par des habitudes [...] sans aucune visibilité » (Latour, 2007, p. 295). C'est pourquoi nous avons observé la capacité des actants, et leur *capacitation*, afin de tracer les contours des espaces stratégiques où ils évoluaient.

Espaces stratégiques

Le dispositif comble un manque en positionnant et en agréant les éléments de manière à répondre à un besoin qui correspond dans notre cas au besoin d'information exprimé par les internautes aussi bien qu'au besoin des éditeurs de recevoir du trafic sur leurs pages. Dès lors, il a une dimension stratégique forte, et certains acteurs tentent d'agréer les éléments autour de leurs intérêts, quitte à s'opposer pour cela aux intérêts d'autres parties. Le dispositif est ainsi indissociable de la notion de rapports de force, et son agencement dépend du champ d'action de *ceux qui exercent le pouvoir*, et de *ceux par qui* et *ce par quoi* le pouvoir s'exerce.

Nous nous sommes intéressés à la manière dont les compétences et les performances étaient distribuées et à ce que celles-ci pouvaient induire en termes de « manipulation de rapports de forces, [...]et d']intervention rationnelle et concertée dans ces rapports de forces, soit pour les développer dans telle direction, soit pour les bloquer, ou pour les stabiliser, les utiliser. » (Foucault, 1994, vol.3, p. 299). Nous avons répertorié pour cela l'ensemble des actions que pouvaient faire les éditeurs : leur « pouvoir faire », lié aux caractéristiques du web considéré comme un espace de possibilités d'action, et leur « pouvoir de faire », accordé par les concepteurs du dispositif. Cela nous a donné un aperçu de leur marge de manœuvres, c'est-à-dire de la manière dont les éditeurs pouvaient établir, spécifier ou renforcer l'association de leur cours d'action à l'action du dispositif, et de quelle manière il leur était également possible de réduire, d'interrompre ou de rompre cette association. Nous avons notamment expliqué comment le texte aussi bien que l'architexte et l'hypertexte pouvaient être utilisés. Nous avons montré qu'il fallait pour optimiser le rapport site/moteur acquérir certaines compétences et développer certains arts opératoires, et qu'en cela « les stratégies de rapports de forces supporta[ien]t des types de savoir, et [étaient] supportés par eux » (*ibid.*). A l'aide d'une analyse détaillée des propos des porte-parole de Google, nous avons également identifié ce que la firme tentait de faire faire aux éditeurs, et ce qu'elle pouvait leur interdire, de quelle manière elle pouvait limiter leur action, les sanctionner ou les récompenser. Nous avons ainsi vu comment Google cherchait à conduire la conduite des éditeurs, ou bien au contraire comment la firme pouvait exercer un pouvoir contraignant à leur endroit, et comment le pouvoir

pouvait être déployé par les éditeurs pour maintenir ou améliorer la visibilité de leurs contenus en activant les lignes d'action possibles ou rendues possibles.

Une fois que nous connaissions les champs d'action de Google et des éditeurs, nous nous sommes intéressés à ce qui justifiait que les éditeurs établissent, spécifient ou renforcent leur implication dans le dispositif, ou qu'au contraire ils la réduisent, l'interrompent ou la rompent. Nous avons vu qu'il existait différents degrés d'enrôlement et de mobilisation, dont dépendaient *in fine* la solidité de l'association et l'ampleur de la traduction. De la même manière, nous avons sondé l'intérêt que pouvait avoir Google à influencer, limiter, sanctionner ou récompenser l'action des éditeurs. Nous avons également montré comment les espaces juridictionnels pouvaient entrer en concurrence, obligeant les acteurs à certains déplacements visant à se prévaloir de l'un ou l'autre de ces espaces. Enfin, nous avons vu comment le dispositif pouvait donner lieu à certaines stratégies de montée en généralité. Google, les éditeurs, les journalistes et parfois certains juges, certains élus, certains scientifiques, certains essayistes, tentent de s'exprimer au nom des internautes, de l'espace-web en général et du bien commun, pour définir ce qu'il convient de faire en matière de production et de mise en circulation de l'information d'actualité.

Machine à faire dire (et à ne pas faire dire)

A partir des contenus produits par les éditeurs de presse en ligne, et selon des modalités qui lui sont propres, le dispositif *met en récit* l'actualité. Il génère alors un traitement de l'actualité qui n'est imputable ni aux éditeurs, ni directement à la firme Google, mais qui est le sien *avec* les leurs et les leurs *à travers* le sien. La méta-éditorialisation de l'actualité est ainsi conditionnée en partie par ce que les concepteurs croient être une bonne manière de s'informer, et affectée par ce qui, à leurs yeux, peut faire qu'une source, un document ou un auteur sera plus pertinent qu'une autre source, un autre document, un autre auteur, en même temps qu'elle est affectée par ce qui, aux yeux des éditeurs, constitue les bons choix en matière de traitement de l'actualité et le bon arbitrage en matière d'activités d'information et de communication.

Certains éditeurs peuvent agir de façon à maximiser leurs chances d'être lus, c'est-à-dire les chances pour leurs contenus de figurer « en bonne place » dans le méta-énoncé.

Ils suivent pour cela les recommandations données par les porte-parole de Google et procèdent à leurs propres ajustements selon ce qu'ils savent de la manière dont le dispositif fonctionne. C'est en cela qu'il est possible de considérer que le dispositif contribue à « façonner le discours public lui-même »³¹¹ (Grimmelmann, 2010, p. 456). Nous disons alors que le méta-énoncé (la liste) et les énoncés produits par Google (les recommandations) *performent* les circonstances de production et de circulation du contenu journalistique, et que le dispositif dans son ensemble est une « machine à faire dire » instituant un « régime d'énoncé » (Bonaccorsi et Julliard, 2010, p. 5).

« L'acte de dire est un usage *de* la langue et une opération *sur* elle » (De Certeau, 1980, p. 82). Dans notre cas, nous avons étudié l'acte de dire dans le cas d'une langue dont la syntaxe, la grammaire, la conjugaison et le vocabulaire sont fixés *a priori* par Google (pouvoir de parler) et par des possibilités d'action propres au web (pouvoir parler). Les éditeurs font usage *de* cette langue et opèrent *sur* elle. La manière dont ils s'en saisissent peut modifier la langue elle-même, dès lors qu'elle n'existe, comme n'importe quelle langue, que par la *médiation* de ceux qui la parlent et peuvent la déformer.

Les concepteurs de Google Search et Google Actualités créent et paramètrent des indicateurs qui sont les porte-voix de ce que les éditeurs et les internautes expriment quand ils agissent. La portée de chacun de ces indicateurs est définie par le coefficient de pondération que les concepteurs leur attachent dans l'algorithme d'appariement. Certains indicateurs se font l'écho de la performance technique des sites des éditeurs, d'autres de sa position dans le graphe hypertexte, d'autres de l'architexte et du texte, et d'autres, enfin, du comportement des internautes. Les indicateurs sont régulièrement réajustés par les concepteurs, tandis que de nouveaux indicateurs sont créés et pondérés. De nouvelles données s'ajoutent alors à l'inscription, la complètent, la renforcent et circulent avec elle. Les éditeurs et les internautes ignorent en grande partie quels sont les indicateurs de l'algorithme et, donc, lesquelles de leurs actions influencent le méta-énoncé. Par ailleurs, quand ils connaissent un indicateur, ils ignorent quelle est sa pondération dans l'algorithme, c'est-à-dire l'influence effective de l'action dont cet indicateur est le porte-voix. Le dispositif laisse ainsi en coulisses les conditions d'énonciation et ouvre la voie

³¹¹ « Search engines are the new mass media [...] capable of shaping public discourse itself. » (Grimmelmann, 2010, p. 456, notre traduction)

aux spéculations. Nous avons vu avec notre analyse du PageRank que cette mise en coulisse pouvait être justifiée par le fait que certains indicateurs, étant donné les actions qu'ils encouragent, étaient susceptibles de nuire à la bonne marche de l'énonciation et même d'empêcher que l'énonciation ait lieu. Des employés de Google communiquent à propos des critères et de leurs pondérations, cherchant ainsi à s'ériger en porte-parole de ce qu'il se passe dans les coulisses. Quand ils prétendent qu'ils ont modifié l'algorithme ou qu'il est préférable d'agir d'une certaine manière, sans que personne ne puisse le vérifier, leur discours peut influencer les actions des éditeurs. Ainsi, l'algorithme et le discours à propos de l'algorithme sont indissociables et performant, ensemble, et pour partie, la réalité qu'ils traduisent. Ils ne font pas état de la hiérarchie des documents : ils la *créent*.

S'il est une manière faire dire, le dispositif est également une manière de ne pas faire dire. Un éditeur, en optimisant le texte, l'architexte et l'hypertexte peut « couper la parole » à ses concurrents. Etant donné la nature hiérarchique du méta-énoncé, la félicité de l'un a ainsi lieu grâce à l'« infélicité » (Austin, 1970) d'un autre. Cependant, nous avons vu que les éditeurs identifiaient souvent les mêmes indicateurs et que les stratégies de SEO pouvaient conduire à une uniformisation des sites et des contenus. Ainsi, le « faire dire » du dispositif est loin d'être optimal en matière de diversité. Nous avons également vu que des quiproquos pouvaient avoir lieu et que des fautes journalistiques pouvaient même être « couronnées » par le dispositif d'infomédiation, étant donné précisément les stratégies de prise de parole. Ainsi, le dispositif « donne la parole » à certaines contre-vérités ou certaines ambiguïtés, et « ôte la parole » à certains contenus de qualité.

Toujours à propos du faire dire, nous avons vu comment le dispositif pouvait faire parler le besoin d'information des internautes par l'intermédiaire de l'outil « Tendances de Recherche », susceptible d'être utilisé par les éditeurs pour définir quels sujets traiter et quels mots employer. Dans ce cas, ce sont les requêtes des internautes qui performant : en disant « nous nous intéressons à ce sujet » par la médiation de « Tendances de Recherche », les internautes suscitent le traitement du sujet en question. Mais nous avons vu qu'il y avait là également un « ne pas faire dire », dès lors que Google ne publiait que partiellement les informations concernant les requêtes effectuées.

Machine à faire voir (et à ne pas faire voir)

Le dispositif n'est pas un *gate-keeper* (gardien du pont) mais un *gate-watcher* (surveillant du pont) : sa fonction n'est pas de garder mais de re-garder, et de faire voir une partie de ce qu'il a vu. Les crawleurs lui servent d'yeux. Ils se déplacent grâce aux hypertextes et rendent compte aux serveurs de Google du texte, des hypertextes et des architextes, qui seront ensuite mis en mémoire, indexés et éventuellement appariés à une requête. De la manière dont les contenus d'un éditeur et leur position sont « vus » par les crawleurs dépendra *in fine* leur mise en visibilité par le dispositif. Dès lors, les éditeurs de presse, s'ils veulent établir un régime « d'exhibition d'une information authentique et fiable » (Le Grignou et Neveu, 1993, p. 943), et pour « labelliser » leurs contenus (Wolton, 2003), peuvent agir de manière à ce que les hypertextes, le texte et l'architexte soient le « mieux vus » possible par les crawleurs. C'est la raison pour laquelle ils positionnent des mots-clés le plus tôt possible dans leurs titres et leurs chapeaux, implémentent des méta-tags permettant au crawleur d'obtenir une information complète à propos du contenu, implémentent des *Sitemaps* de manière à aplanir la structure hypertexte dans laquelle le crawleur se déplace, et optimisent le temps de chargement des pages. Les éditeurs peuvent également agir de manière à être vus d'une certaine façon plutôt qu'une autre, par exemple en mettant en avant certains contenus en particulier. Ainsi, grâce aux méta-tags et aux *Sitemaps*, ils peuvent spécifier au crawleur quels sont les textes et les hypertextes les plus pertinents. Enfin, grâce au protocole d'exclusion robots.txt, ils peuvent agir de manière à ce que leurs contenus, ou une partie de leurs contenus, ne soient pas vus par les crawleurs.

Nous avons décrit comment, avec les méta-tags et les microdonnées, les éditeurs pouvaient tenter de prendre la main sur la manière dont les liens vers leurs contenus apparaîtraient. Avec le « rel=author », nous avons expliqué comment l'auteur pouvait se mettre en scène en faisant en sorte que sa photo soit ajoutée aux liens générés vers le contenu qu'il a produit. Avec les pages d'atterrissage, nous avons décrit comment les éditeurs pouvaient positionner certaines pages destinées expressément à être vues par les crawleurs, et à leur « plaire », puis à être vues par les utilisateurs de Google, alors que ces pages ne présentent que peu d'intérêt aux yeux des journalistes et sont rarement mis en avant sur le site lui-même. Nous avons également expliqué comment les éditeurs pouvaient cacher des liens hypertexte, cachant ainsi à l'internaute ce qui est vu par le crawleur.

Finalement, nous avons expliqué comment les éditeurs pouvaient faire en sorte de construire leurs sites et leurs documents de façon à les mettre en lumière (intégrer le champ), à les mettre « le plus possible » en lumière (optimiser la captation) et à maîtriser la façon dont ils reçoivent la lumière (spécifier les zones de crawling et les modalités d’affichage). La vision du web produite par le dispositif influence donc le web, comme si le fait de regarder un objet modifiait les caractéristiques de cet objet (phénomène bien connu par les physiciens et les biologistes).

Le dispositif est une machine à faire voir autant qu’à ne pas faire voir. Certains points de vue et certains énoncés journalistiques sont en effet exclus. Les documents qui ne sont pas reliés à d’autres par des liens hypertextes sont invisibles pour les crawlers. Les sites dont le contenu est payant sont moins visibles que ceux dont le contenu est gratuit. Quant aux sujets qui ne sont pas traités par de nombreux éditeurs, ils seront eux-aussi moins visibles, ou pas du tout. Le dispositif ne nous montre que ce que sa lumière est en mesure de dévoiler. Il ne révèle jamais l’existence des zones inaccessibles. Il ne révèle pas non plus à l’internaute les chemins de navigation vers lesquels la liste s’ouvre. Il ne révèle pas les scores *PageRank* obtenus. Et il ne révèle pas, dans « Tendances de Recherche », les valeurs absolues du nombre de requêtes effectuées.

Rapports de pouvoir

Nous avons montré que les rapports de pouvoir entre Google et les éditeurs donnaient lieu à ce que nous avons nommé un consensus unilatéral et léonin. Les éditeurs ont beau attaquer Google en justice, s’unir pour peser dans la balance, ou pratiquer le *black hat SEO*, les dirigeants et employés de Google sont ceux qui choisissent les conditions et les modalités du référencement sur leurs services. Ils ne décident certes pas seuls de ce que sera la hiérarchie des liens hypertextes, mais décident de ce que sera la nature de leur relation avec les éditeurs et de la marge de manœuvres dont ces derniers jouiront. Ils les « habilitent », c’est-à-dire qu’ils rendent possibles leurs actions, « et ce faisant leur assignent une limite et une portée » (Badouard, 2012, p. 481). Finalement, la plupart des éditeurs n’utilisent pas le robots.txt pour exclure leur site, et n’implémentent pas non plus l’*Automated Content*

Access Protocol, puisque les moteurs ne le respectent pas. Aussi acceptent-ils les dispositions et le résultat d'un consensus unilatéral et léonin.

Cependant, même si Google a le pouvoir d'exclure un éditeur du champ du dispositif (auquel cas la sanction est disciplinaire), l'essentiel du pouvoir exercé par la firme n'est pas totalisant, mais relève plutôt du « gouvernement des conduites ». En effet, même si les éditeurs se déplacent dans un champ qui les régule à un premier niveau, « ils y introduisent une façon d'en tirer parti qui obéit à d'autres règles et qui constitue comme un second niveau imbriqué dans le premier » (De Certeau, 1980). Les éditeurs sont libres de se conformer ou non aux recommandations de Google, peuvent parfois y désobéir délibérément, peuvent à tout moment exclure leurs documents du champ de crawling et créer certains espaces exempts de l'autorité de Google. Dès lors, un jeu d'intéressement et d' enrôlement se crée. Des aller-retour ont lieu, des ajustement, des négociations, dont dépendront *in fine* la mobilisation des éditeurs, la satisfaction des intérêts de chacun et le degré de performativité du dispositif. A tout moment, si un acteur estime que ce qui lui est demandé ne peut pas (ou plus) satisfaire ce qu'il considère comme étant son propre intérêt, il peut rompre la chaîne d'associations. La firme Google, si elle veut continuer d'orienter la performativité dans un sens qui lui est favorable, pourra revoir la matérialité du dispositif de manière à diminuer la portée de ce que les concepteurs considéreront, étant donné leur programme d'action, comme des trahisons, des contestations ou des malversations vis-à-vis de leur projet. Ainsi, « chaque effet positif ou négatif [du dispositif], voulu ou non voulu, vient entrer en résonance, ou en contradiction, avec les autres, et appelle à une reprise, à un réajustement, des éléments hétérogènes » (Foucault, 1994, vol. 3, p. 299). Et dès lors, il n'y a pas de pouvoir absolu, ou de performativité unidirectionnelle, mais toujours : des influences mutuelles (performations/contre-performations, compositions/recompositions).

Lignes de fuite

Nous avons cherché les lieux où le pouvoir de Google n'opérait plus, et où les éditeurs effectuaient une subjectivation visant à sortir du dispositif. Nous avons vu que le rôle du référenceur pouvait consister à dégager certains espaces où les

journalistes n'étaient plus contraints. Ainsi, le dispositif peut « fuir », c'est-à-dire que les journalistes peuvent se soustraire aux « rapports de force établis comme des savoirs constitués » (Deleuze, 1989, p. 187). Nous avons notamment expliqué comment le référenceur pouvait utiliser le code HTML pour contre-performer, ou bien pour rendre invisible la performativité du méta-énoncé. Dans ce cas, les journalistes peuvent ne plus se soucier de la performativité des textes qu'ils produisent eux-mêmes sur les conditions d'énonciation du méta-énoncé, dès lors que c'est l'architexte et l'hypertexte automatisé qui se chargeront, seuls, de performer. La firme Google peut elle-même prévoir des lignes par lesquelles les éditeurs pourront « fuir » : le module « Choix des rédactions » a ainsi été mis en place pour que des contenus puissent apparaître sans passer par l'action d'un algorithme, et donc sans être optimisés.

Réassemblage du dispositif

Nous avons observé en ligne la nature et le résultat du processus de communication effectivement opéré, en même temps que nous interrogeons les acteurs concernés (Partie 3). Il s'agissait de sonder l'« être ensemble » sous-tendu par le processus en « réassemblant le social » autour des éléments hétérogènes, des espaces stratégiques, de la machine à faire dire/voir/faire et des lignes de fuites constitutifs du dispositif. Métaphoriquement, après avoir mis à jour le *squelette* de la situation de communication, nous avons cherché à comprendre comment la *chair* (le social) s'y agençait, et quels étaient les effets *organiques* (sociaux) d'un tel agencement. C'est ainsi que nous pouvons désormais évaluer nos hypothèses.

Hypothèse 1

Le dispositif d'infomédiation Google contribue à créer de nouvelles formes de valorisation de la production journalistique.

Le dispositif est un réservoir susceptible d'apporter un trafic important, très peu qualifié, et dont la venue, même si elle n'est jamais assurée, est plus facile à provoquer que de nouveaux accès directs. La possibilité d'être bien placé dans cette

vitrine et le fait d'avoir intégré le champ d'indexation de Google Actualités sont valorisés par l'éditeur sur le marché de la publicité de la même manière que sont valorisés en presse écrite l'autorité d'un titre et son appartenance à un réseau de distribution. En revanche, certains éléments inédits apparaissent dans notre cas. Nous avons vu notamment que le dispositif pouvait donner lieu à l'organisation d'un marché de l'autorité *PageRank*. Certains éditeurs de presse en ligne, en vendant des liens à d'autres éditeurs, valorisent leur position dans l'arbre de navigation. Une telle pratique n'est pas encouragée par la firme Google, car le *PageRank* perd de la crédibilité en tant qu'indicateur de pertinence. C'est pourquoi la firme Google menace, sanctionne et communique, désireuse d'éviter autant que possible qu'un tel marché ne s'organise. Apparemment, les principaux éditeurs français issus de la presse imprimée respectent les préconisations de la firme sur ce point.

Nous avons également révélé une forme de valorisation que nous avons nommé « stratégie du conjugueur ». La production journalistique peut être agrémentée de liens automatiques effectués vers des pages qui ne sont pas produites par des journalistes, et notamment vers les pages d'un conjugueur. Sur celles-ci, des publicités figurent, rémunérées au coût par affichage. Ainsi, l'autorité *PageRank* est redistribuée par l'éditeur à ses propres pages, qui n'ont que peu de valeur ajoutée, tandis que l'autorité des liens effectués par les journalistes est diluée. Les pages du conjugueur seront presque toujours visitées par des utilisateurs de Google. Dans le cas du *Monde*, les liens vers le conjugueur sont effectués automatiquement sur chaque verbe employé par un journaliste à l'infinitif, et sont invisibles. Le lecteur ne voit pas la valorisation dont le contenu fait l'objet, et ce n'est pas son attention qui est valorisée, comme pour une bannière publicitaire classique, mais l'autorité de la page qu'il visite. La firme Google est théoriquement opposée à ce genre de pratique, mais ne semble pas sanctionner les entreprises dont la marque est forte et qui jouissent ainsi d'une forme d'impunité et d'un avantage concurrentiel non-négligeable.

Pour conclure quant à notre première hypothèse, nous dirions qu'il existe de nouvelles formes de valorisation du contenu journalistique induites par le dispositif d'infomédiation, puisque celui-ci donne de la valeur à la position topographique de la production, et non pas seulement à la marque ou à son contenu. Etant donné le

fonctionnement du dispositif, les liens effectués depuis un contenu journalistique ont un *effet*, et donc, potentiellement, un prix. En plus d'être des « signes passeurs », ils deviennent des *services*, et, dès lors, même si c'est encore peu le cas en France, un marché est susceptible de s'organiser.

Hypothèse 2

Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google entraînent une reconfiguration des organisations de presse.

En France, les principaux sites d'information politique et générale adossés à des titres de presse imprimée ont embauché un référenteur. L'internalisation de la stratégie d'optimisation nous a été présentée comme une condition nécessaire à son efficacité sur le long terme étant donné la veille et les échanges quotidiens que celle-ci demande. Une telle internalisation coûte cher, sans certitude quant au retour sur investissement. Ainsi, il semble que seules certaines organisations de presse aient les moyens de supporter de tels coûts (et de tels risques) et que les plus riches soient ainsi potentiellement les plus visibles sur Google.

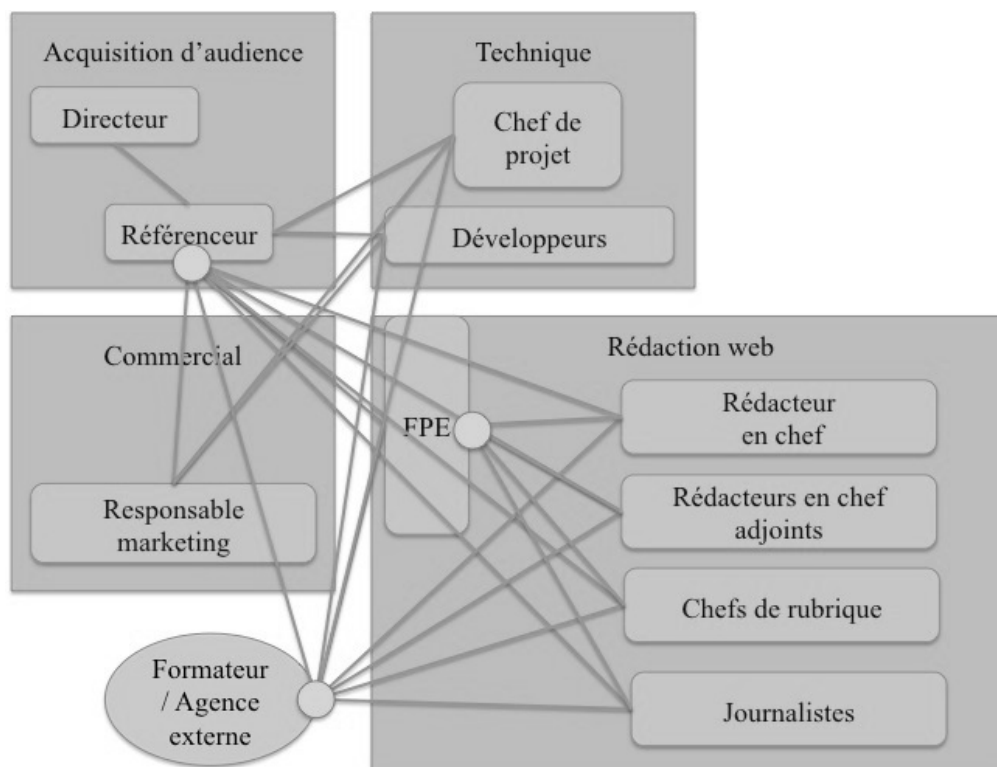
Le référenteur a un profil hybride, fruit de la mise en tension de trois dimensions de l'activité de production : technique, marketing et éditorial. Il lui faut à la fois agir mais aussi négocier, persuader, comprendre, veiller. Afin de définir une stratégie de maximisation du trafic, il capitalise les outils de description et d'analyse à sa disposition, son propre savoir, le langage HTML, les conseils donnés par Google, le contenu du site pour lequel il travaille, les objectifs de ses dirigeants, les intentions de ses collègues, les analyses publiées par des experts, les compte-rendu de mesure d'audience, l'observation des sites concurrents, les listes de résultats et ce qu'elles lui apprennent à propos de la situation à optimiser. Le rôle du référenteur est de *traduire* son propre projet et de rendre compatibles les actions qu'il propose avec les buts poursuivis par ses partenaires, et avec leurs intérêts, leurs marges de manœuvres, leurs représentations, sans quoi son action échouera. Du succès de cette traduction dépendront d'une part la maximisation du trafic, d'autre part les effets concrets de la stratégie de référencement et le contrôle exercé par l'entreprise sur la mise en lumière

de ses contenus, et d'autre part enfin l'issue heureuse du rapport dialectique entretenu entre l'éditeur et Google.

Le référenceur n'a d'ascendant hiérarchique sur aucun de ses partenaires et travaille sur un mode « agile », incorporé dès le début des nouveaux projets techniques, éditoriaux et publicitaires. Il peut recevoir le soutien d'une agence externe, spécialisée en SEO, qui lui apporte des outils et une expertise supplémentaires : d'autres lignes d'action, d'autres informations. Il bénéficie également de certains alliés privilégiés, notamment le formateur et le *Front Page Editor* (FPE). Le formateur est extérieur à l'entreprise et donne aux journalistes des conseils quant à l'écriture adaptée au web en général et à Google en particulier. Il peut être un allié du référenceur, en confirmant et en complétant ses conseils d'ordre éditorial auprès de la rédaction, comme il peut être un atténuateur ou même un contradicteur. Le FPE, quant à lui, est un journaliste intégré à la rédaction. Son rôle vis-à-vis du référencement peut consister à veiller à ce que ses recommandations soient suivies par les rédacteurs comme il peut consister à les en décharger en effectuant à leur place ces actions. L'agence externe (consultants en SEO), le formateur et le FPE sont, comme le référenceur, des centres de traduction : ils interagissent à plusieurs niveaux et constituent des nœuds stratégiques du réseau d'attachements dont dépend la bonne marche de la stratégie d'optimisation et les conditions de félicités de sa performance.

Si nous représentons schématiquement l'organisation de la stratégie de référencement, cela donne :

Figure 49. SEO : échanges stratégiques



Sur le schéma ci-dessus, chaque trait représente des échanges formels ou informels susceptibles d'induire des actions destinées à optimiser le référencement. Très peu d'ordres formels sont donnés, étant donné l'absence d'ascendant hiérarchique des centres de traductions (marqués par des ronds). Chaque acteur a sa propre marge de manœuvres et ne se retrouve jamais dans un environnement purement contraint. Des négociations ont lieu, durant lesquelles les programmes d'action des centres de traduction rencontrent ceux de leurs partenaires, et de l'issue desquelles dépendent la nature et les effets de la stratégie de captation. Nous voyons sur ce schéma par quelles interactions passe et circule le pouvoir. Nous comprenons, étant donné la multiplicité des interactions, qu'il peut y avoir certains points d'achoppement dès lors que le pouvoir d'un centre de traduction peut être renforcé, atténué ou contré par le pouvoir d'un autre centre de traduction, ou par l'un des acteurs concernés. Il y a des zones d'incertitude, des compétences, des arts de faire, de part et d'autre des interactions. Un jeu s'instaure entre les acteurs, durant lequel

chacun pourra être tiré par le réseau dans un sens qui n'est pas tout à fait le sien, ou pas du tout, ou au contraire tirer le réseau dans le sens de ses intentions. Au final, l'ampleur de la traduction effectuée, son efficacité et la satisfaction des acteurs concernés dépendront de la coordination des champs d'action des acteurs ici représentés, elle-même rendue possible par l'efficacité des centres de traduction.

Hypothèse 3

Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent les choix quant à la structure du site, des pages et du code HTML.

Nous avons expliqué comment les sites de presse en ligne pouvaient être structurés selon des impératifs liés au référencement, et comment le code HTML pouvait être rédigé de manière à performer le dispositif. D'autre part, nous avons également constaté que de nombreuses possibilités d'action étaient laissées de côté par les éditeurs, soit par manque de temps, soit parce qu'ils en ignoraient l'existence, soit parce qu'elles leur semblaient inutiles. Ainsi, dans l'ensemble, l'usage fait des balises méta-tags est permissif : les éditeurs délèguent à l'ensemble du dispositif (concepteurs, algorithmes, crawlers, indexeurs, autres éditeurs, internautes...) une partie de leur capacité à influencer les modalités d'infomédiation de leurs documents. Par ailleurs, il existe certains « pouvoir de faire » attribués spécialement par Google aux éditeurs de presse en ligne, qui permettraient à la firme, si ceux-là s'en saisissaient, d'affiner ses résultats. C'est le cas des balises « news_keywords », « syndication-source », « rel=author » et « standout ». En se saisissant de ces balises, les éditeurs pourraient potentiellement recevoir davantage de trafic, tandis que Google affinerait la qualité de ses résultats selon ce que la firme considère elle-même comme étant cette qualité. Mais les éditeurs ne s'en saisissent pas, ou très peu, ou de manière incomplète, le *Sitemap* étant le seul outil utilisé par tous assez finement.

Par ailleurs, pour ce qui est de la structure des sites, nous avons vu avec le cas de la stratégie de « Coupole » chez Roularta Media France que plusieurs sites pouvaient être rassemblés sous la même adresse URL pour des raisons liées en partie

au référencement. Pour les trois titres de presse imprimée mensuels *L'Expansion*, *L'Entreprise* et *Votre Argent*, trois sous-domaines du site Lexpress.fr ont été créés. Du point de vue SEO, autant que du point de vue de la stratégie commerciale, cela permet d'avoir davantage de force : mutualiser le PageRank et l'AuthorRank, rassembler le maillage interne à l'intérieur d'un seul domaine et peser plus lourd sur le marché de la publicité en termes de nombre de documents mis en ligne, de pages vues et de visiteurs uniques. Tout cela est susceptible de contribuer à maximiser la visibilité des différents titres sur les dispositifs Google Search et Google Actualités. De la même manière que les « pertes et profits » sont mutualisés, permettant aux recettes des sites les plus rentables de soutenir les sites qui le sont moins, la mutualisation de l'autorité permet aux sites les mieux classés par Google d'entraîner dans leur sillage les sites qui, autrement, seraient moins visibles.

Hypothèse 4

Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation influencent les webjournalistes dans leurs pratiques de production et de publication.

Ecrire pour Google consiste en premier lieu à rédiger des titres et des chapeaux où les mots-clés apparaîtront le plus tôt possible. Il s'agit d'éviter les jeux de mots ou les ambiguïtés, de manière à ce que le crawleur identifie tout de suite le sujet d'une page. L'écriture du reste du texte est, quant à elle, dégagée des pressions liées au référencement. Cela n'est pas inconciliable avec les règles enseignées dans les écoles de journalisme mais rejoint au contraire le procédé de « pyramide inversée » qui consiste à donner l'essentiel de l'information dès le début d'un article. Ainsi, le caractère informatif des premières lignes, et le caractère substituable des différents niveaux de lecture (titre, sous-titre, légende, accroche) constituent des points de convergence dans la rencontre des scripts des concepteurs, du référenceur et des journalistes.

La pratique d'écriture des journalistes est particulièrement influencée lors du « bâtonnage » de dépêches d'agence : celles-ci sont publiées dans une forme

optimisée pour Google, leurs titres et leurs chapeaux sont reformulés, et elles sont accompagnées d'une image et de liens hypertextes vers des pages d'atterrissage ou d'autres articles du site. D'autre part, plusieurs journalistes nous ont expliqué qu'ils étaient particulièrement attentifs aux pratiques relatives au référencement lorsqu'un article avait à leurs yeux une valeur ajoutée importante, et qu'ils souhaitaient par conséquent qu'il soit lu par de nombreux internautes, quitte pour cela à faire certaines concessions quant à la formulation du titre et du chapeau.

Outre l'influence exercée sur l'écriture elle-même, il y a une influence possible sur les pratiques de mise à jour. Mais les cas où une mise à jour est faite simplement pour remonter dans les grappes de Google Actualités, alors qu'aucune information n'a été ajoutée, sont perçus comme des pertes de temps par les journalistes, qui n'y voient pas d'intérêt éditorial et ne l'acceptent par conséquent que dans le cas où ils n'ont pas autre chose à faire de plus important. Le FPE peut éventuellement les seconder, ou même les remplacer dans cette tâche, en remettant lui-même les articles à jour à des fins de référencement. Il s'agit d'un exemple où le rôle du centre de traduction est de permettre aux différents scripts d'être conciliés. Par ailleurs, le référenceur peut demander aux journalistes de favoriser le maillage interne et d'effectuer des liens hypertextes sur certains mots-clé en particulier, vers certaines pages d'atterrissage. Les journalistes feront ces liens s'ils y voient un sens éditorial, et parfois, plus rarement, utiliseront un mot plutôt qu'un autre, simplement parce qu'ils savent qu'une page d'atterrissage existe vers laquelle il sera intéressant de pointer depuis ce mot en particulier.

Nous avons vu également comment sont créés des espaces où les pratiques journalistiques sont dégagées de toute influence de Google. Grâce au dédoublement du titre, notamment, les journalistes peuvent écrire un titre incitatif pour le lecteur, tout en écrivant *aussi* un titre informatif et optimisé pour le moteur. Ce qui est écrit « pour Google » est ainsi dissocié de ce qui est écrit « pour l'internaute », de manière à ce que, théoriquement, la stratégie de maximisation du trafic n'empiète pas sur la créativité des journalistes. La performativité du dispositif est ainsi *localisée* en un seul endroit, afin d'en exempter d'autres lieux. Le référenceur fait en sorte ainsi de concilier le script des journalistes et le sien. Il peut être secondé dans cette tâche par

le FPE qui rédige parfois lui-même le deuxième titre, ou recommande aux journalistes d'utiliser cette option.

Finalement, l'influence des conditions et des modalités de captation du trafic sur les pratiques de production et de publication des journalistes est toujours négociée, jamais imposée, jamais certaine, cela car les journalistes arbitrent entre ce qui leur est demandé et ce qu'ils pensent qu'il convient de faire, en fonction du temps qu'ils ont à accorder à chacune de leurs tâches.

Hypothèse 5

Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent les lignes éditoriales des sites de presse en ligne.

Nous avons vu que le dispositif d'infomédiation pouvait être prescripteur, notamment par l'entremise de « Tendances de Recherche », qui peut servir aux journalistes d'outil de décision quant au traitement de tel ou tel sujet. Le référenceur s'en fait le porte-voix en assistant directement aux conférences de rédaction ou bien en téléphonant aux journalistes pour leur dire que tel ou tel sujet pourrait générer un trafic substantiel. Il peut également injecter « Tendances de Recherche » dans le système de management de contenus dans le but que les journalistes l'aient sous les yeux lorsqu'ils publient leurs articles, pour choisir les mots-clés qu'ils utiliseront, mais aussi pour qu'ils constatent quels sujets sont en vogue. Ainsi, la demande, par la médiation de l'outil « Tendances de Recherche » et du référenceur, parfois du FPE, ou même du rédacteur en chef, peut susciter l'offre. Aux yeux des journalistes, certains des sujets suggérés sont intéressants, tandis que d'autres sujets sont sans valeur éditoriale, ou bien concernent des événements à propos desquels ils n'auront pas d'information intéressante à révéler par rapport à ce qui a déjà été relaté par la concurrence. Le référenceur, dans le but de mobiliser les journalistes, leur explique que traiter certains sujets ne leur coûtera pas beaucoup de temps, et que cela pourra même à la rigueur n'avoir « aucune valeur éditoriale », mais qu'ils pourront ainsi « remplir leur quota » et être libérés alors des exigences liées au SEO vis-à-vis du reste de leur production.

Ainsi, certains sujets peuvent être traités « uniquement pour Google », mais cela demeure rare d'après nos observations, étant donné une réticence de la part des journalistes. En effet, si les journalistes peuvent accepter de faire certaines concessions quant à la ligne éditoriale, il n'en demeure pas moins qu'ils continueront de traiter les sujets qui leur semblent importants, et ce que cela soit rentable ou non. On assiste alors non pas à un effacement de la ligne éditoriale, mais à une certaine dysharmonie entre d'une part les sujets traités à des fins de captation, pouvant ou non être perçus comme dignes d'intérêt, et, d'autre part, les sujets traités indépendamment de toute considération liée au trafic qu'ils rapporteront. Là encore, la performance est *localisée*, puisque seule une partie de la ligne éditoriale est influencée.

Enfin, nous avons expliqué que si tous les éditeurs de presse en ligne choisissaient les sujets selon leur seul potentiel de succès sur Google, nous assisterions sans doute à un appauvrissement considérable de la diversité. En effet, l'absence de diversité chez l'infomédiaire pourrait provoquer une absence de diversité chez ceux qui ont besoin de passer par son infomédiation pour accéder à leur lectorat. Le fait que certains choix éditoriaux puissent être faits indépendamment de la question du référencement nous semble dès lors fondamental.

Problématique & hypothèse générale

Problématique. La mise en énoncé effectuée par le dispositif d'infomédiation Google performe-t-elle le processus de production et de valorisation du contenu mis en ligne par les entreprises de presse ?

Hypothèse générale. Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation Google influencent radicalement la manière dont le fait d'actualité est mis en récit, mis en circulation et valorisé par les entreprises de presse sur le média web.

Le dispositif crée la hiérarchie dont il rend compte. En cela, il performe l'espace web, et son action est susceptible d'influencer l'ensemble du processus de production, de communication et de valorisation de l'information. Mais il est aussi lui-même performé par les actions de nombreux acteurs, qui pourront se conformer aux conditions et aux modalités de captation définies par les concepteurs, mais qui pourront aussi tenter de contre-performer le champ d'énonciation institué.

En éfinitive, la performance est double : performance de l'énoncé produit par Google sur les conditions de production et de valorisation de l'information, performance de la production et de la valorisation de l'information sur l'énoncé produit par Google. Des aller-retour ont lieu entre les différents niveaux de performance. Dès lors, chaque acteur concerné par la production du méta-énoncé peut jouer un rôle qui ne sera pas forcément celui qui a été prévu ou souhaité par les autres, et influencer ainsi non seulement le résultat du processus mais aussi la nature même de ce processus, tout en influençant les actions, les possibilités d'action et les résultats des actions des autres.

Grâce à nos cinq hypothèses, nous avons expliqué dans quelle mesure la valorisation du contenu journalistique, l'organisation de sa production, les caractéristiques des sites de presse en ligne, les pratiques journalistiques et les lignes éditoriales pouvaient être influencés par les conditions et les modalités de captation du trafic sur Google. Mais la performance n'est pas pour autant « radicale » comme le suggérait notre hypothèse générale. Elle est localisée, traduite, choisie, instable, négociée. En conséquence, la phrase de notre hypothèse devrait, pour devenir exacte, être formulée ainsi : « Les conditions et les modalités de captation du trafic en provenance du dispositif d'infomédiation influencent *partiellement et selon les intérêts et la marge de manœuvres de chaque acteur concerné* la manière dont le fait d'actualité est mis en récit, mis en circulation et valorisé par les entreprises de presse sur le média web. »

Limites

Chaque thèse a ses limites, sources pour nous à la fois de projets futurs et d'une certaine dose de frustration concernant le parcours effectué. Nous présenterons ici six limites à notre avis significatives. La première concerne le choix de nous intéresser aux seuls sites français issus de la presse imprimée d'information politique et générale. Nous avons effectué ce choix car les entreprises propriétaires de ces sites étaient comparables et, d'autre part, car elles avaient en commun d'avoir embauché un référenceur au moment où nous commençons notre enquête. Il serait intéressant d'effectuer une analyse semblable auprès des sites nés en ligne, des sites issus de

stations de radio et de chaînes de télévision, comme il serait intéressant d'analyser les stratégies de référencement des amateurs, des fermes de contenus, et également des professionnels de la communication publique ou des entreprises. Il serait intéressant par ailleurs d'analyser la situation dans plusieurs autres pays et de voir ainsi si les raffinements locaux auxquels procède Google pour son algorithme d'appariement aboutissent à des différences significatives en termes de SEO. Nous pourrions de la sorte analyser les différences et les points communs entre ces sites et ces pays, les mises en œuvre techniques et éditoriales et leurs effets en termes de mise en récit de l'actualité. Nous aboutirions à une compréhension plus complète et plus nuancée de *ceux qui font* et de *ce que fait* le processus d'infomédiation.

La deuxième limite de ce travail concerne notre échantillon. L'analyse de 252 articles prélevés en octobre 2012 nous a permis de dégager certaines tendances quant au texte, à l'architexte et à l'hypertexte de 6 sites d'information politique et générale issus de la presse imprimée et comparables. Or il se trouve que nous aurions pu raffiner notre observation de trois façons différentes. D'une part, nous aurions pu amplifier notre échantillon en prélevant davantage d'articles dans le but de gagner en représentativité. D'autre part, nous aurions pu élargir notre échantillon en intégrant à notre panel d'autres sites issus de la presse imprimée et comparables à ceux que nous avons choisis. Enfin, nous aurions pu répéter notre échantillonnage dans le but de donner un caractère dynamique à nos observations en identifiant non pas seulement ce que les éditeurs font, mais également comment leurs actions évoluent. Amplifier, élargir, répéter : ce sont les prolongements classiques d'une enquête par échantillon.

La troisième limite concerne nos entretiens semi-directifs. Nous n'avons rencontré des dirigeants d'entreprises de presse, des journalistes, des référenceurs et des techniciens. Nous ne recherchions pas alors la représentativité mais plutôt à percevoir la diversité des mécanismes. En conséquence, interroger davantage d'acteurs aurait certainement pu nous permettre de révéler certains éléments passés entre les mailles de notre filet analytique, ou bien de nuancer certaines de nos conclusions. Par ailleurs, pour des raisons que nous avons exposées dans notre premier chapitre, nous n'avons pas pu accéder aux ingénieurs travaillant chez Google, qu'il aurait pourtant été extrêmement intéressant d'interroger. En outre, nous n'avons

pas interrogé d'employés du *Monde* et du *Nouvel Observateur*, alors que ces deux titres ont révélé certaines particularités lors de nos observations à propos desquelles nous aurions souhaité obtenir des commentaires de vive voix.

La quatrième limite concerne notre monographie. L'approche monographique a été privilégiée pour révéler certains points qui exigeaient que de nombreux acteurs soient rencontrés et observés au sein d'une même organisation. Cela nous a permis de constater comment, très concrètement, le référencement donnait lieu à des négociations, des détours, des points d'achoppement et certaines concessions. Nous devrions faire d'autres monographies et les comparer à la nôtre, comme nous avons déjà comparé la nôtre au travail publié pendant notre parcours de thèse par Murray Dick (Dick, 2011).

La cinquième limite concerne la performance des stratégies décrites. Idéalement, il aurait fallu rendre compte des résultats de l'optimisation en termes d'apport de trafic, et non pas seulement du résultat espéré et des effets de l'optimisation sur le texte, l'architexte et l'hypertexte. Hélas, la mesure des effets concrets des tâches relevant du SEO en termes de trafic est extrêmement difficile à obtenir, voire impossible, étant donné le secret dont l'algorithme fait l'objet. En effet, comment dire quel a été l'impact de telle ou telle action quand on ne sait pas avec certitude qu'elle a un impact, quelle est l'amplitude potentiel de cet impact (la pondération du critère), et quels autres critères, avec quelle amplitude, peuvent également être à l'origine de la captation. La corrélation est toujours supposée, jamais avérée. Il faudrait pour cela effectuer des mesures sur de nombreux sites et tester différentes variables une à une, en les faisant varier sans toucher aux autres variables, de manière à obtenir des coefficients qui ne seraient potentiellement plus exacts aussitôt que les équipes de Google auraient re-paramétré l'algorithme (ce qui arrive plusieurs fois par mois).

Enfin, la sixième limite de ce travail est liée à notre choix de nous positionner du côté de la production journalistique et de la conception du moteur, mais jamais du côté de leur réception et de leur usage. Idéalement, pour finir de réassembler le dispositif autour du processus dont nous avons fait état, il faudrait que nous utilisions les clefs théoriques et méthodologiques de la sociologie des usages pour étudier la

« relation » que les internautes entretiennent avec le dispositif, les « mobiles, les formes d'usage et le sens accordé à la pratique » et, enfin, la « dynamique du social » et « l'évolution des modes de vie » sous-tendus par l'infomédiation, ainsi que le préconise Josiane Jouët (Jouët, 1993).

Perspectives

Nous souhaitons à l'avenir continuer d'étudier l'infrastructure Internet et le média web à partir du positionnement théorique et des outils méthodologiques que nous avons forgés et opérationnalisés ici. Nous souhaitons étudier comment, dans un mouvement indifféremment technique, économique, juridique, politique, discursif, culturel, le réseau d'information et de communication est conçu, négocié, équipé, mobilisé, reconfiguré. Nous souhaitons continuer d'identifier les réseaux d'action et les ensembles navigationnels, d'analyser les rapports de pouvoir et leurs enjeux, leurs effets, leurs points d'achoppements, et continuer pour cela à considérer que le pouvoir n'est que très rarement totalisant, mais qu'au contraire un jeu se crée dont l'issue dépend des marges de manœuvres des uns et des autres, des tentatives de capacitation, des arts de faire, des détours effectués et des buts poursuivis. En étudiant les centres de traduction, et en analysant ce qui influence leur action, ce qui permet leur action, et *ce sur quoi* et *ceux sur qui* leur action influe, nous pourrions ainsi continuer notre tentative de compréhension du web et collaborer aux travaux menés à ce sujet dans les laboratoires de sciences humaines et sociales.

Les pistes que nous voudrions poursuivre et approfondir peuvent être résumées par quatre points :

Premièrement, nous souhaiterions poursuivre notre étude de la manière dont le code informatique est créé, institué, utilisé et détourné, ainsi que rendre compte des causes et des effets des différentes actions de communication qu'il induit. Nous pourrions par exemple nous intéresser en particulier à la manière dont le code HTML (en ce moment HTML 5) est conçu, par qui, comment, selon quels intérêts, pour quelles marges de manœuvres, pour quels comportements juridiques et économiques, et de quelle manière les acteurs se saisissent de ce code et suscitent grâce à lui des situations non-prévues par ses concepteurs. Quelle grammaire pour quelles actions ?

Quelles fautes de grammaire pour quelles contestations ? Quels « être ensemble » enfin ? Quelles *conversations* ?

Deuxièmement, nous souhaiterions continuer à nous intéresser aux moteurs de recherche algorithmiques et, de manière générale, à tout ce qui prétend évaluer, hiérarchiser et apparier automatiquement les informations. « Evaluer, a écrit Nietzsche, c'est créer », or c'est précisément à cet acte de création que nous nous intéressons, ainsi qu'à ce qu'il devient une fois effectué et à ce que devient l'espace sociotechnique où il a été, et par lequel il a été, effectué. En considérant les algorithmes comme des formules mathématiques qui, tour à tour, performant et sont performées, nous pourrions tenter d'analyser comment les mathématiques combinatoires interprètent et déplacent (i.e. *traduisent*) le social et comment le social, à son tour, traduit les objectivations et les itérations propres aux algorithmes. Il s'agirait également de voir comment les algorithmes et le code cohabitent et se performant mutuellement. De surcroît, comme nous l'avons fait ici pour le SEO, nous étudierions ce que les acteurs peuvent faire, savent faire et font dans un environnement truffé d'algorithmes dont les détails sont souvent inconnus et dont dépendent *in fine*, et en partie seulement, la satisfactions des intérêts des uns et des autres, quels qu'ils soient.

Troisièmement, nous voudrions étoffer notre concept d'arbre de navigation afin de représenter le web comme une dynamique d'ouverture et de fermeture de divers champs d'opportunités, et considérer ce que cela induit en matière d'incitations et de déplacements pour les acteurs, selon l'écosystème informationnel dont ils proviennent (la racine de l'arbre). Nous pourrions travailler avec des économistes spécialistes de la théorie des jeux, avec qui nous modéliserions les ensembles navigationnels comme autant d'espaces de décision, induisant diverses situations de communication.

Enfin, nous souhaiterions continuer à nous intéresser au traitement et à la circulation de l'information relative aux événements d'actualité, cela car nous pensons qu'il s'agit d'un enjeu majeur. Les webjournalistes et leurs partenaires, leurs sources, leurs intermédiaires, leurs lecteurs, ne forment un collectif que dans la mesure où chacun y a un intérêt et y joue un rôle qui non seulement ne nuit pas à la

démocratie mais, en plus, et surtout, l'aide, la nourrit, l'équipe. Nous voudrions par conséquent étudier ce qui permet, encourage ou peut gêner la formation d'un tel collectif, et identifier, dans l'idéal, les points où les programmes d'action sont susceptibles d'être conciliés.

Ces projets sont ambitieux, nous en sommes conscients. Ils constituent, pour ainsi dire, notre propre « programme d'action », assez ouvert pour laisser place à l'étonnement et aux rencontres, assez précis pour nous servir de guide dans les années à venir. D'un point de vue strictement personnel, et presque *intime*, ces projets sont l'aboutissement dont nous sommes le plus fier. Notre thèse s'achève grâce à eux non pas sur une impasse mais par une aspiration. Bien plus que de satisfaire la curiosité qui nous animait depuis l'enfance, elle l'aura ainsi justifiée, suscitée, nourrie, intensifiée, concrétisée et canalisée.

Paris/Toulouse : Septembre 2010 – Septembre 2013

Epilogue

Prendre la parole sur le web ne se fait pas à l'aide de la seule langue maternelle d'un individu mais également du code, qui est, pour ainsi dire, la « langue maternelle » des dispositifs informatiques. C'est en apprenant cette langue que les webjournalistes pourront se saisir pleinement d'une marge de manœuvres qui leur permettra aussi bien de trouver l'information que de la traiter et de la publier.

Les systèmes de management de contenu ont pour but de simplifier la publication grâce à des interfaces dont la prise en main est plus ou moins instinctive. Cependant, ces interfaces sont des machines à traduire les actions des journalistes en code et dès lors, comme pour n'importe quelle machine, elles pourront être contraignantes et offrir un champ d'action limité, tandis que, comme pour n'importe quelle traduction, il y aura interprétation et déplacement du *sens*. Le risque pour le journaliste est d'être mal compris, de perdre en subtilité, en contrôle, et même, dans certains cas, de dire l'inverse de ce qu'il a voulu signifier. C'est pourquoi nous pensons que le webjournaliste n'a pas d'autre choix que d'apprendre lui-même la langue utilisée dans l'espace où il évolue. Connaître le code lui permettrait en effet de s'affranchir de nombreuses contraintes et d'optimiser ses chances de dire ce qu'il veut dire, comme il veut le dire et d'être entendu.

D'autre part, le journaliste qui a conscience de ce que représente un lien hypertexte — quels chemins de navigation, quels échanges d'autorité — et connaît les différents types de liens — même onglet, nouvel onglet, nouvelle fenêtre, *nofollow* — pourra tenter de positionner le document qu'il produit à l'endroit de l'arbre de navigation où il sait que la greffe pourra prendre. Ainsi, sur le web, si le texte incarne l'acte de communication, c'est l'architexte qui le rend soluble et l'hypertexte qui le situe. Le *pouvoir* d'un individu passe dès lors par la maîtrise de ces trois composantes, qui lui permettront de faire preuve de créativité et de contester ou de contourner le pouvoir des autres. A condition d'acquérir une telle maîtrise, le webjournaliste sera capable de mettre en forme et de partager ce qui, à ses yeux, *doit* être mis en forme et partagé. Le lecteur le trouvera, sans que le journaliste l'ait *mal* ou *trop* cherché.

Bibliographie

Tous les liens ont été visités pour la dernière fois le 10 septembre 2013.

Acap (2009), *Guide to implementation of ACAP Version 1.1 Communication with Crawlers. A component of the ACAP Technical Framework*, Issue 1, 9 Octobre.

Akrich M. (1987), « Comment décrire les objets techniques ? », *Techniques & Culture*, vol. 1, n°54-55, p. 205-219.

Akrich M. (1991), « L'analyse socio-technique », in : Vinck D. (dir.), *La gestion de la recherche*, Bruxelles, De Boeck, p. 339-353.

Akrich M. (1993), « Les formes de la médiation technique », *Réseaux*, n°60, p. 87-98.

Akrich M. (2006), « La description des objets techniques », in : Akrich M., Callon M., Latour B. (dir.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses des Mines de Paris, p. 159-178.

Akrich M., Latour B. (1992), « A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies », in : Bijker W. E., Law J. (dir.), *Shaping Technology/Building Society. Studies in Sociotechnical Change*, Cambridge MA, MIT Press, p. 259-164.

Alami S., Desjeux D., Garabua-Moussaoui I. (2009), *Les méthodes qualitatives*, Paris, P.U.F., coll. « Que sais-je ? », 128 p.

Albert P. (2002), « L'abandon du monopole postal et ses effets (1850-1890) », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon-Assas, p.143-157.

Albert P. (2010), *Histoire de la presse* (première édition en 2003), Paris, PUF, coll. « Que sais-je ? », 128 p.

Anderson C. (2006), *The Long Tail. Why the Future of Business is Selling less or More*, New York, Hyperion, 256 p.

Anderson C. (2009), « Web Production, News Judgement, and Emerging Categories of Online Newswork in Metropolitan Journalism », texte présenté lors d'une communication au *Online Journalism Symposium*, University of Texas, Austin.

Anderson, S. P., Foros Ø., Kind H. J., Peitz M. (2012), « Media market concentration, advertising levels, and ad prices », *International Journal of Industrial Organization*, 30(3), p. 321-325.

Appel V., Boulanger H., Massou L. (2010), *Les dispositifs d'information et de communication. Concepts, usages et objets*, Bruxelles, De Boeck, 246 p.

Argenton C., Prüfer J. (2012), « Search engine competition with network externalities », *Journal of Competition Law and Economics*, 8(1), p. 73-105.

Armstrong M. (2006), « Competition in Two-Sided Markets », *RAND Journal of Economics*, 37(3), p. 668-691.

Arnesen M. (2012), « How to Prepare for AuthorRank and Get the Jump on Google », *SEOMoz*, 21 sept. 2012, disponible ici : <http://www.seomoz.org/blog/how-to-prepare-for-authorrank-and-get-the-jump-on-google>

Attias D. (2007), *L'impact d'Internet sur l'économie de la presse : quel chemin vers la profitabilité ?*, Thèse en Sciences Economiques, Université Paris-X Nanterre, 232 p.

Augey D., Rebillard F. (2009), « Introduction : la dimension économique du journalisme », *Les Cahiers du Journalisme*, n°20, p. 10-21.

Auletta K. (2010), *Googled. The End of the World as We Know It*. London, Penguin Books, 432 p.

Austin J. (1970/1991), *Quand dire c'est faire*, Paris, Seuil, 208 p.

Bachimont B. (2010), *Le sens de la technique : le numérique et le calcul*, Paris, Les belles lettres, coll. « encre marine », 190 p.

Badouard R. (2012), *Les « technologies politiques » du web. Une étude des plateformes participatives de la Commission Européenne et de leurs publics*, Thèse en Sciences de l'Information et de la Communication, UTC Compiègne, 536 p.

Balle F., Cohen-Tanugi L. (2001), *Dictionnaire du Web*, Dalloz, 322 p.

Band J. (2008), « Google and fair use », *Journal of business and technology law*, vol.3, n°1.

Bardini T. (1996), « Changement et réseaux socio-techniques : de l'inscription à l'affordance », *Réseaux*, vol. 14, n°76. p. 125-155.

Barnhurst K. (2012), « The form of online news in the mainstream U.S. press 2001-2010 », *Journalism Studies*, 13(5), p. 791-800.

Barthes R. (1970), *Mythologies*, Paris, Seuil, coll. « Points », 234 p.

Barzi R. (2011), « PME et agilité organisationnelle : étude exploratoire », *Innovations*, vol. 2, n°35, p. 29-45.

Bataillé O. (2002), « La Poste et le transport des journaux, de l'arrêté du 27 prairial an IX à la loi du 6 avril 1878. La lente érosion d'un monopole », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon-Assas, p. 55-75.

Battelle J. (2005), *The search: how Google and its rivals rewrote the rules of business and transformed our culture*, Boston / Londres, Nicholas Brealey Publishing, 288 p.

Baudrillard J. (2002), *L'esprit du terrorisme*, Galilée, 45 p.

Bayard S. (2009), « Do Ads in Google News Search Change the Fair Use Analysis? », *Citizen Media Law Project*, Brekman Center for Internet & Society.

Benghozi P.-J. (2011), « Le deuxième choc de l'économie de la culture », *Esprit*, juillet 2011, p. 111-125.

Benhamou F., Farchy J. (2007), *Droit d'auteur et copyright*, Paris, La Découverte, 124 p.

Benkler Y. (2006), *The Wealth of Networks. How Social Production Transforms Markets and Freedom*, New Heaven / London, Yale University Press, 528 p.

Berman R., Katona Z. (2011), « The Role of Search Engine Optimization in Search Marketing », Working Paper, disponible ici : <http://www.ron-berman.com/wp-content/uploads/2011/02/seo29.pdf>

Bernoux P. (2010), *Sociologie du changement dans les entreprises et les organisations* (1^{ère} édition en 2004), Paris, Points, 368 p.

Beuscart J.-S., Peerbaye A. (2006), « Histoires de dispositifs (introduction) », *Terrains& travaux*, 2 (11), p 3-15.

Bianchini M., Gori M., Marco Gori, Scarselli F. (2005), « Inside PageRank. », *ACM Transactions on Internet Technology*, 5, p. 92-128.

Blandin C. (2004), « Ne quittez pas L'Express ! », *Le Temps des médias*, vol. 2, n°3, p. 60-73.

Boczkowski P.J. (2004), *Digitizing the news : innovation in online newspapers*, Cambridge MA, MIT Press, 264 p.

Boczkowski P.J. (2010), « Ethnographie d'une rédaction en ligne Argentine. Les logiques contraires de la production de l'information chaude et froide », *Réseaux*, n°160-161, p. 43-78.

- Bomsel O.** (2007), *Gratuit ! Du déploiement de l'économie numérique*, Paris, Gallimard, coll. « Folio », 304 p.
- Bonaccorsi J., Julliard V.** (2010), « Dispositifs de communication numériques et médiation du politique. Le cas du site web d'Ideal-Eu », in : Aghababaei, M., Bonjour A., Clerc, A., Rauscher G., (dir.), *Usages et enjeux des dispositifs de médiation*, Nancy, Presses Universitaires de Nancy, p. 65-78.
- Bougnoux D.** (1995), *La communication contre l'information*, Paris, Hachette, 144 p.
- Bouquillion P.** (2008), *Les industries de la culture et de la communication. Les stratégies du capitalisme*, Grenoble, PUG, 306 p.
- Bourdieu P.** (1982), *Ce que parler veut dire. L'économie des échanges linguistiques*, Paris, Fayard, 246 p.
- Boure R.** (2000), « L'interdisciplinarité en débat », *Sciences de la Société*, n°50/51, p. 7-19.
- Boutet C.-V., Ben Amor S.** (2010), « Vers l'active SEO 2.0 », *Les Cahiers du numérique*, vol. 6, p. 179-198.
- Breed W.** (1955), « Social control in the newsroom: A functional analysis », *Social Forces*, 33 (1/4), p. 326-335.
- Brin S., Page L.** (1998), « *The Anatomy of a Large-Scale Hypertextual Web Search Engine* », Computer Science Department, Stanford University, disponible ici : <http://infolab.stanford.edu/~backrub/google.html>
- Brown J. S., Duguid P.** (2000), *The Social Life of Information*, Boston, Harvard Business School Press, 336 p.
- Bruns A.** (2008), « The Active Audience : Transforming Journalism from Gatekeeping to Gatewatching », in : Paterson C., Domingo D., *Making Online News. The Ethnography of New Media Production*, New York, Peter Lang, p. 171-184.
- Bustamante E.** (2004), « Cultural industries in the Digital Age: some provisional conclusions », *Media, Culture and Society*, vol. 26, n°6, p. 803-820.
- Cabrolié S.** (2010), « Les journalistes du Parisien.fr et le dispositif technique de production de l'information », *Réseaux*, n°160-161, p. 135-161.
- Caillaud B., Jullien B.** (2003) « Chicken & Egg: Competition among Intermediation Service Providers », *RAND Journal of Economics*, 34(2), p. 309-28.
- Caillaud B., Trégouët T.** (2006), « Les marchés bifaces », *Revue Lamy de la Concurrence*, 9, p. 106-112.

Callon M. (1986), « Eléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc », *L'Année Sociologique*, 36, p. 169-208.

Callon M. (1987), « Society in the Making: The Study of Technology as a Tool for Sociological Analysis », in : Bijker W.E., Hugues T.P., Pinch T. (dir.), *The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology*, Cambridge MA, MIT Press, p. 83-103.

Callon M. (1995), « Four Models for the Dynamics of Science », in : Jasanoff S., Markle G., Petersen J., Pinch T. (dir.), *Handbook of Science and Technology Studies*, Thousand Oaks, Sage, p. 29-63.

Callon M. (dir.) (1998), *The Laws of the Markets*, Blackwell Publishers, Oxford, 277 p.

Callon M. (2006), « Sociologie de l'acteur réseau », in : Akrich M., Callon M., Latour B. (dir.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses des Mines de Paris, p. 267-276.

Callon M., Latour B. (2006), « Le grand Léviathan s'apprivoise-t-il ? », in : Akrich M., Callon M., Latour B. (dir.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses des Mines de Paris, p. 11-32.

Calman C., Balin R.D. (2011), « The Future of the Hot News Misappropriation Tort after Barclays Capital Inc. v. TheFlyontheWall.com », *Media Law Monitor*, disponible ici : <http://www.medialawmonitor.com/2011/09/the-future-of-the-hot-news-misappropriation-tort-after-barclays-capital-inc-v-theflyonthewall-com/>

Cardon D. (2010), *La démocratie Internet. Promesses et limites*, Paris, Seuil, 102 p.

Cardon D. (2013), « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, 1, n°177, p. 63-95.

Carlson M. (2007), « Order versus access: news search engines and the challenge to traditional journalistic roles », *Media Culture Society*, 29(6), p. 1014-1030.

Carr N. (2010), « Taylor Momsen Did Not Write This Headline », *New York Times*, 16 mai 2010, disponible ici : http://www.nytimes.com/2010/05/17/business/media/17carr.html?_r=0

Caves R. E. (2002), *Creative Industries : Contracts between Art and Commerce*, Cambridge MA, Harvard University Press, 464 p.

Cayrol R. (1966), « Nouvel Express et Nouvel Observateur », *Revue française de science politique*, n°3, p. 493-520.

Chakrabarti S., Van den Berg M., Dom B. (1999) « Focused crawling: A new approach to topic-specific web resource discovery », *Proceedings of the 8th International World Wide Web Conference*, p. 543- 562.

Chantepie P. (2009), « Web 2.0 : les économies de l'attention et l'insaisissable internaute-hypertexte », *Esprit*, mars-avril 2009, p. 107-127.

Charon J.-M., Le Floch P. (2011), *La presse en ligne*, Paris, La Découverte, coll. « Repères », 126 p.

Cho J., Garcia-Molina H. (2000), « The evolution of the web and implications for an incremental crawler », *Proceedings of 26th International Conference on Very Large Data Bases*, p. 117–128.

Clève V. (2010), « SEO Google News : une compétence rare qui se monnaie cher », *Café référencement*, disponible ici : <http://www.cafe-referencement.com/recherche-universelle/seo-google-news-une-competence-rare-qui-se-monnaie-cher-100#axzz2TGagwxFA>

Clève V. (2011), « Réussir son référencement : internalisation ou sous-traitance ? », *Café référencement*, disponible ici : <http://www.cafe-referencement.com/seo-management/reussir-son-referencement-internalisation-ou-sous-traitance-392#axzz2TGagwxFA>

Cochoy F. (2010), « Comment faire des affaires avec des mots-choses : presse commerciale et performative », *Réseaux*, 5, n°163, p. 77-103.

Coddington M. (2012), « This Week in Review: A Limp Response to Plagiarism, and a Proposed Tax to Save Newspapers », *Nieman Journalism Lab*, disponible ici : <http://www.niemanlab.org/2012/09/this-week-in-review-a-limp-response-to-plagiarism-and-a-proposed-tax-to-save-newspapers/>.

Cooper-Richet D. (2002), « Distribution, diffusion et circulation du *Galvani's Messenger* (1814-1890), premier quotidien de presse anglais », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p.121-139.

Coquaz V. (2013), « Presse web : un rédacteur en chef nommé Google », *Arrêt sur image*, disponible ici : <http://www.arretsurlimages.net/contenu.php?id=5848>

Crowell G. (2005) « Copyrights, Trademarks and Search Engines », *Search Engine Watch*, disponible ici : <http://searchenginewatch.com/showPage.html?page=3557706>

Crozier M., Friedberg E. (1977), *L'acteur et le système. Les contraintes de l'action collective*, Paris, Seuil, 500 p.

Curien N. (2005), *Economie des réseaux*, Paris, La Découverte, coll. « Repères », 123 p.

Cusset Y. (2006), « Pédagogie et politique. Quelques réflexions à partir d'une remarque de Maria Montessori », *Le Télémaque*, n°2, p. 107-116.

Currah A. (2009), *What's Happening to Our News? An investigation into the likely impact of the digital revolution on the economics of news publishing in the UK*, Oxford, Reuters Institute for the Study of Journalism, 166 p.

D'Almeida F., Delporte C. (2003), *Histoire des medias en France de la Grande Guerre à nos jours*, Paris, Flammarion, 506 p.

Davallon J. (1992), « Le musée est-il vraiment un média ? », *Publics et musée*, n°2, p. 99-117.

Davallon, J. (2004), « Objet concret, objet scientifique, objet de recherche », *Hermès*, n°38, p.30-37.

Davallon J., Jeanneret Y. (2004). « La fausse évidence du lien hypertexte », *Communication et langages*, n°140, p. 43-54.

Davallon J., Jeanneret Y. (2006), « Le lien discontinu de la communication », XIXth Congress International Association of Empirical Aesthetics, *Culture and communication*, Avignon, IAEA, p. 9-16.

De Certeau M. (1980), *L'invention du quotidien, I. arts de faire*, Paris, UGE/10-18, 348 p.

Degand A., Grevisse B. (dir.) (2012), *Journalisme en ligne. Pratiques et recherches*, Bruxelles, Deboeck, coll. « Info&Com », 384 p.

De la Haye Y., Miège B. (1984), « Les sciences de la communication : un phénomène de dépendance culturelle ? », in De la Haye Y. (dir.), *Dissonances, Critique de la communication*, Grenoble, La Pensée Sauvage, p. 157-164.

De Maeyer J. (2009), « Journalism practices in an online environment. An investigation into the evolving specificities of newswork in Belgium, as challenged by technological changes », *Journalism in crisis conference*, University of Westminster.

De Maeyer J. (2010) « Être journaliste dans un environnement 2.0 », *Les Cahiers du Numérique*, 6(1), p. 157-177.

De Maeyer J. (2012) « Towards a hyperlinked society: A critical review of link studies », *New Media & Society*, vol. 15, n°5, p. 737-751.

De Maeyer J. (2013), *L'usage journalistique des liens hypertextes. Étude des représentations, contenus et pratiques à partir des sites d'information de la presse belge francophone*, Thèse de doctorat, Université Libre de Bruxelles, Faculté de Philosophie et de Lettres, 295 p.

Deleuze G. (1986), *Foucault*, Paris, Editions de Minuits, 144 p.

Deleuze G. (1989), « Qu'est-ce qu'un dispositif ? », in : *Michel Foucault philosophe. Rencontre internationale. Paris 9, 10, 11 janvier 1988*, Paris, Seuil, p. 185-195.

Delforce B. (2010), « La responsabilité sociale du journaliste : donner du sens », *Les cahiers du journalisme*, n° 2, p. 16-32.

Dellarocas C., Katona Z., Rand, W. M. (2010), « Media, Aggregators and the Link Economy: Strategic Hyperlink Formation in Content Networks » NET Institute Working Paper n°10-13 ; Boston U. School of Management Research Paper n°2010-30 ; Robert H. Smith School Research Paper n°RHS 06-131.

Denis J. (2006), « Préface : Les nouveaux visages de la performativité », *Études de communication*, 29.

Denis J., Pontille D. (2010), « Performativité de l'écrit et travail de maintenance », *Réseaux*, vol.5, n° 163, p. 105-130.

Derieux E. (2002), « Le statut de la distribution de la presse : le régime des messageries », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 381-390.

Derrida J., Habermas J. (2004), *Le concept du 11 septembre*, Paris, Galilée, 244 p.

Derville G. (1999), « Le journalisme et ses contraintes », *Les cahiers du journalisme*, n°6, p. 152-177.

Deslauriers J.-P. (1991), *Recherche qualitative : Guide pratique*, Montréal, McGraw-Hill, 142 p.

Devine K. L. (2008), « Preserving Competition in Multi-Sided Innovative Markets: How Do You Solve a Problem Like Google? » *North Carolina Journal of Law & Technology*, vol. 10, n°59.

Diaz A. (2008), « Through the Google Goggles: Sociopolitical Bias in Search Engine Design », in : Zimmer M., Spink A. (dir), *Web Search. Multidisciplinary Perspectives*. Berlin / Heidelberg, Springer, p. 11-34.

Dick M. (2011), « Search engine optimisation in UK news production », *Journalism Practice*, vol. 5, n°4, p. 462-477.

Diligenti M., Coetzee F., Lawrence S., Giles C., Gori. M. (2000), « Focused crawling using context graphs », *Proceedings of 26th International Conference on Very Large Data Bases*, p. 527-534.

Dimitrova D. V., Neznanski M. (2006) « Online Journalism and the War in Cyberspace: a comparison between U.S. and international newspapers », *Journal of Computer-mediated Communication*, 12(1), p. 248-263.

- Dimitrova D. V., Connolly-Ahern C., Williams A. P., Kaid L. L., Reid A.** (2003) « Hyperlinking as Gatekeeping: online newspaper coverage of the execution of an American terrorist », *Journalism Studies*, 4(3), p. 401-414.
- Dooley B.** (1999) « De bonne main : les pourvoyeurs de nouvelles à Rome au XVIIIe siècle », *Annales. Histoire, Sciences Sociales*. 54, n°6, p. 1317-1344.
- Duffez O.** (2005), « Google Sitemaps expliqué pas à pas », *Webrankinfo*, disponible ici : <http://www.webrankinfo.com/dossiers/sitemaps/presentation-google-sitemaps>
- Duguid P.** (2008), « The World According to Grep », *Deep Search*, Vienne, novembre 2008, disponible ici : http://world-information.org/wii/deep_search/en/videos
- Dupuy-Busson S., Sonnac N.** (2002), « Les relations entre éditeurs et marchands de journaux », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 391-408.
- Duval J.** (2004), *Critique de la raison journalistique. Les transformations de la presse économique en France*, Seuil, Paris, 107 p.
- Eck H.** (2002), « Enjeux et débats politiques autour de la naissance des NMPP », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 253-265.
- Edwards D.** (2011), *I'm Feeling Lucky: The Confessions of Google Employee Number 59*, Boston, Houghton Mifflin Harcourt, 432 p.
- Edwards J., McCurley K. S., Tomlin J. A.** (2001), « An adaptive model for optimizing performance of an incremental web crawler », *Proceedings of the Tenth Conference on World Wide Web* (Hong Kong: Elsevier Science), p. 106-113.
- Engbretsen M.** (2006), « Shallow and Static or Deep and Dynamic? Studying the state of online journalism in Scandinavia », *Nordicom Review*, 27, p. 3-16.
- Ertzscheid O., Gallezot G., Boutin E.** (2009), « PageRank : entre sérendipité et logiques marchandes », in : Simonnot B., Callezot G. (dir.), *L'entonnoir. Google sous la loupe des Sciences de l'Information et de la Communication*, Caen, C&F editions, p. 113-136.
- Esquenazi J.-P.** (2006), « Les médias et leurs publics. Le processus de l'interprétation », in : Olivesi S. (dir.), *Sciences de l'Information et de la Communication*, Grenoble, Presses Universitaires de Grenoble, p. 11-26.
- Estienne Y.** (2007), *Le journalisme après Internet*, Paris, L'Harmattan, 320 p.
- Evans D. S.** (2003), « Some Empirical Aspects of Multi-Sided Platform Industries », *Review of Network Economics*, 2(3), p. 191-209.

- Eveno P.** (2002), « La distribution du *Monde* : un problème économique, rédactionnel, industriel et commercial », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 355-369.
- Fallows D.** (2005), « Search engine users », *Pew Internet & American Life Project*, disponible ici : http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf
- Farchy J.** (2001), « Le droit d'auteur est-il soluble dans l'economie numerique ? », *Réseaux*, vol. 6, n°110, p. 16-40.
- Farchy J.** (2006), « Les industries culturelles à l'heure de la numérisation », *Esprit*, 5, p. 142-153.
- Farchy J., Méadel C.** (2013), « Moteurs de recherche et référencement : chassez le naturel... », *Hermès*, n°66.
- Feng F., Bhargava H., Pennock D.** (2007), « Implementing Sponsored Search in Web Search Engines : Computational Evaluation of Alternative Mechanisms », *INFORMS Journal on Computing*, 19(1), p. 137-148.
- Feyel G.** (dir.) (2002), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, 256 p.
- Feyel G.** (2007), *La presse en France des origines à 1944. Histoire politique et matérielle* (première édition en 1999), Paris, Ellipses, 192 p.
- Fitzgerald B., O'Brien D., Fitzgerald A.** (2008), « Search Engine Liability for Copyright Infringement », in : Spink A., Zimmer M. (dir.), *Web Search: Multidisciplinary Perspectives*, p.103-120.
- Flichy P.** (1991), *Une histoire de la communication moderne, espace public et vie privée*, La Découverte, Paris, 281 p.
- Flichy P.** (1997), « La question de la technique dans les recherches sur la communication », *Sociologie de la communication*, vol. 1 n°1, p. 243-270.
- Flichy P.** (2003), *L'innovation technique. Récents développements en sciences sociales. Vers une nouvelle théorie de l'innovation*, Paris, La Découverte, coll. « Sciences et société », 256 p.
- Foucault M.** (1975/1993), *Surveiller et punir, naissance de la prison*, Paris, Gallimard, coll. « Tel », 360 p.
- Foucault M.** (1994), *Dits et écrits* (vol. 1, 2 et 3), Paris, Gallimard, coll. « Quarto ».
- Freyssenet M.** (1990), *Les formes sociales d'automatisation, Cahiers du GIP Mutations industrielles*, Paris, 1990, 47 p.

Fukuyama F., Shulsky A. (1997), *The « Virtual Corporation » and Army Organization*, Santa Monica, Rand, 97 p.

Furhoff L. (1973). « Some reflections on newspaper concentration », *Scandinavian Economic History Review*, 21(1), p. 1-27.

Gabszewicz J., Laussel D., Sonnac N. (1999), « TV-broadcasting Competition and Advertising », *CORE Discussion Paper*, 6, Université catholique de Louvain et Lucarnes Bleues n°9972 CREST/LEI.

Gabszewicz J., Garela P. Sonnac N. (2007), « "Newspapers" Market Shares and the Theory of the Circulation Spiral », *Information Economics and Policy*, 19, p. 405-441.

Garber M. (2011), « Krishna Bharat on the evolution of Google News and the many virtues of "trusting in the algorithm" », *Nieman Journalism Lab* (Harvard), <http://www.niemanlab.org/2011/05/google-news-founder-krishna-bharat-we-see-ourselves-as-the-yellow-pages/>

Garcia T. (2010), *Forme et objet. Un traité des choses*, Paris, PUF, coll. « MétaphysiqueS », 492 p.

Gaudrat P. (2007), « Le droit d'auteur au lendemain de la transposition », *RTD Com.* 2007, p. 107.

Gavillet I. (2010), « Michel Foucault et le dispositif : questions sur l'usage galvaudé d'un concept », in : Appel V., Boulanger H., Massou L. (dir.), *Les dispositifs d'information et de communication. Concepts, usages et objets*, Bruxelles, De Boeck, p. 17-38.

Geiger C. (2007), « From Bern to national law, via the Copyright Directive: the dangerous mutations of the three-step test », *European intellectual property review*, 29(12), p. 486-491.

George E. (2004), « La place de la démarche critique dans les études en communication », *Colloque du département des communications de l'Université du Québec à Montréal*, 15 p.

Gerhart S. (2004) « Do Web search engines suppress controversy? », *First Monday*, vol. 9, n°1, disponible ici : http://firstmonday.org/issues/issue9_1/gerhart/index.html

Gervais D. (2005), « Towards a New Core International Copyright Norm: The Reverse Three-Step Test », *Marquette Intellectual Property Law Review*, vol. 9, n°1, p. 1-37.

Girard B. (2008), *Une révolution du management - le modèle Google*, Paris, M21, 215 p.

- Goldman E.** (2010), « Search Engine Bias & the Demise of Search Engine Utopianism », in : Szoka B., Marcus A. (dir.), *The Next Digital Decade*. Washington DC, TechFreedom, p. 461-474.
- Goody J.** (1979), *La raison graphique*, Paris, Les Editions de Minuit, 272 p.
- Goyette-Côté M.-O.** (2012), « Le retraitement automatisé de l'information d'actualité en ligne : analyse des mécanismes socio-techniques mis en place par les infomédiaires », *Communication*, vol. 29/2.
- Grefe X., Sonnac N.** (dir.) (2008), *Culture Web : Création, contenus, économie numérique*, Paris, Dalloz, 886 p.
- Grimmelmann J.** (2009) « The Google Dilemma », *New York Law School Law Review*, 53, p. 939-950.
- Grimmelmann J.** (2010) « Some skepticism about search neutrality », in : Szoka B., Marcus A. (dir.), *The Next Digital Decade*. Washington DC, TechFreedom, p. 435-460.
- Gross M., Lentin A.** (1967), *Notions sur les grammaires formelles*, Paris, Gauthier-Villars, 200 p.
- Hadopi** (2012), « Contribution de Google France à la consultation de la Hadopi sur l'exercice effectif des exceptions au droit d'auteur et aux droits voisins », 36 p.
- Halavais A.** (2009), *Search Engine Society*, Cambridge, Polity Press, 196 p.
- Hargittai E.**, (2007). « The social, political, economic, and cultural dimensions of search engines: An introduction », *Journal of Computer-Mediated Communication*, 12(3).
- Harman G.** (2002), *Tool-being: Heidegger and the Meta-physics of Objects*, Chicago, Open Court Publishing, 256 p.
- Hennion A., Latour B.** (1993), « Objet d'art, objet de science. Note sur les limites de l'anti-fétichisme », *Sociologie de l'art*, 6, p. 7-24.
- Herbert J., Thurman N.** (2007), « Paid Content Strategies for news Websites », *Journalism Practice*, 1 (2), p. 208-226.
- Hesmondhalgh D.** (2007), *The Cultural Industries*, London / Los Angeles / New Delhi, Sage, 360 p.
- Hess A.** (2008), « Reconsidering the Rhizome: A Textual Analysis of Web Search Engines as Gatekeepers of the Internet », in : Spink A., Zimmer M. (dir.), *Web Search: Multidisciplinary Perspectives*, Berlin / Heidelberg, Springer, p. 35-50.

- Heydon A., Najork M.** (1999), « Mercator: a scalable, extensible web crawler », *World Wide Web*, 2 (4), p. 219-229.
- Himmelboim I.** (2010), « The International Network Structure of News Media: An Analysis of Hyperlinks Usage in News Web sites », *Journal of Broadcasting & Electronic Media*, 54(3), p. 373-390.
- Hindman M., Tsioutsoulis K., Johnson J.A.** (2003), « "Googlearchy": How a Few Heavily Linked Sites Dominate Politics on the Web », *Annual meeting of the Midwest Political Science Association* (Chicago), 33 p.
- Holscher C., Strube G.** (2000), « Web search behaviour of Internet experts and Newbies », *Computer Networks*, 33 (1-6), p. 337-346.
- Hughes T. P.** (1979), « The Electrification of America the System Builders », *Technology and Culture*, 20 (1), p. 124-162.
- Hughes T. P.** (1983), *Networks of Power. Electrification in Western Societies, 1880-1930*, Baltimore, Johns Hopkins University Press, 474 p.
- Introna L., Nissenbaum H.** (2000) « Shaping the web: why the politics of search engines matters », *The Information Society*, 16, p. 169-185.
- Jamet M.** (1983), « L'Express, du journal d'opinion au news-magazine », *Communication et langages*, n°56, p. 85-97.
- Jansen J., Pooch U.** (2001), « A review of web searching studies and a framework for future research », *Journal of the American Society for Information Science and Technology*, 52, p. 235-246.
- Jansen J.** (2007), « The comparative effectiveness of sponsored and non-sponsored links for Web e-commerce queries », *ACM Transactions on the Web*, 1(1).
- Jansen J., Zhang M., Zhang Y.** (2007), « The Effect of Brand Awareness on the Evaluation of Search Engine Results », *WWW 2007*, p. 1139-1140.
- Jarvis J.** (2009), *What Would Google Do?*, New York, Harper Business, 272p.
- Jeanneret Y., Souchier E.** (1999), « Pour une poétique de l'écrit d'écran », *Xoana*, n°6, p. 97-107.
- Jeanneret Y., Souchier E.** (2005), « L'énonciation éditoriale dans les écrits d'écran », *Communication et langages*, n°145, p. 3-15.
- Jeanneney J.-N.** (2011), *Une histoire des médias*, (première édition en 1996), Paris, Seuil, 446 p.
- Jeon D.-S., Esfahani N. N.** (2013), « News Aggregators and Competition Among Newspapers in the Internet », *Working Paper* n°388, Toulouse School of Economics.

- Johnson S.** (1997), *Interface Culture. How New Technology Transforms the Way We Create and Communicate*, New York, Basic Books, 272 p.
- Jouët J.** (1993), « Pratiques de communication et figures de la médiation », *Réseaux*, n°60, p. 99-120.
- Jouët J.** (2000), « Retour critique sur la sociologie des usages », *Réseaux*, n°100, p. 487-521.
- Kaplan A.** (2001), *Intelligence avec l'ennemi : le procès Brasillach*, Paris, Gallimard, 480 p.
- Kaser D.** (1962), « In principium erat verbum », *Peabody Journal of Education*, 39(5), p. 258-263.
- Kawaguchi A., Mowshowitz A.** (2002), « Bias on the Web », *Communications of the ACM*, vol. 45, n°9, p. 56-60.
- Kenney K., Gorelik A., Mwangi S.** (2000), « Interactive Features of Online Newspapers » *First Monday*, 5(1).
- Kilgour G.** (1998), *The Evolution of the Book*, New York, Oxford University Press, 192 p.
- Kleinberg J.** (1999), « Authoritative sources in a hyperlinked environment », *Journal of the ACM*, 46(5), p. 604-632.
- Kleinberg J., Kumar R., Maghoul F., Raghavan P., Rajagopalan S., Tomkins A., Wiener J.** (1999), « The web as a graph : measurements, models, and methods », *Proceedings of the 5th International Conference on Computing and Combinatorics*, p. 1-17.
- Kleinberg J., Lawrence S.** (2001) « The structure of the web », *Science*, 294, p. 1849-1850.
- Kohn A.** (2012), « Author Rank », *Blind Five Year Old*, disponible ici : <http://www.blindfiveyearold.com/author-rank>
- Kowalski R.** (1979), « Algorithm = logic + control », *Magazine Communications of the ACM*, 22 (7), p. 424-436.
- Kyrou A.** (2010), *Google God. Big Brother n'existe pas, il est partout*, Paris, Inculte, 288 p.
- Isbell K.** (2010), « The Rise of the News Aggregator: Legal Implications and Best Practices », *Berkman Center Research Publication*, n°2010-10.

- Lahaie S.** (2006), « An Analysis of Alternative Slot Auction Designs for Sponsored Search », *Proceeding*, EC '06 Proceedings of the 7th ACM conference on Electronic commerce, pp.218 - 227
- Lamy A.** (2006), « Les spécificités du traitement médiatique dans l'urgence. L'exemple des attentats du 11 septembre 2001 », *Communication et organisation*, n°29, p. 108-122.
- Langville A. N., Meyer C. D.** (2006), *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton, Princeton University Press, 240 p.
- Larsson A. O.** (2012), « Staying in or Going Out? » *Journalism Practice*, p. 1-17.
- Lasica J. D.** (2004), « Balancing Act: How News Portals Serve Up Political Stories », *Online Journalism Review*.
- Latour B.** (1991), *Nous n'avons jamais été modernes*, Paris, La Découverte, 206 p.
- Latour B.** (1993), *La clef de Berlin et autres leçons d'un amateur de sciences*, Paris, La Découverte, 252 p.
- Latour B.** (1996), *Petites leçons de sociologie des sciences*, Paris, Seuil, 252 p.
- Latour B.** (1999a), *Pandora's Hope. Essays on the Reality of Science Studies*. Cambridge MA / London, Havard University Press, 336 p.
- Latour B.** (1999b), « On Recalling ANT », in : Law J. and Hassard J. (dir.), *Actor Network Theory and After*. Oxford, Blackwell Publishers, 264 p.
- Latour B.** (2005), *La Science en action. Introduction à la sociologie des sciences* première édition en anglais en 1987, Paris, La Découverte, 664 p.
- Latour B.** (2006), « Le prince : machines et machinations », in : Akrich M., Callon M., Latour B. (dir.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses des Mines de Paris, p. 87-107.
- Latour B.** (2007), *Changer la société, refaire de la sociologie* (première version publiée en anglais en 2005), Paris, La Découverte, 400 p.
- Latour B.** (2010), *Cogitamus. Six Lettres sur les Humanités Scientifiques*, La Découverte, Paris, 246 p.
- Latour B.** (2012), *Enquête sur les modes d'existence*, Paris, La Découverte, 498 p.
- Latour B., Mauguin P., Teil G.** (1991), « Une méthode nouvelle de suivi des innovations : le graphe socio-technique », in : Vinck D. (dir.), *La Gestion de la recherche. Nouveaux problèmes, nouveaux outils*, Bruxelles, De Boeck, p. 419-567.

Law J. (dir.) (1986), *Power, Action and Belief: a News Sociology of Knowledge?*, London, Routledge and Kegan Paul.

Law J. (1991), *Sociology of Monsters: Essays on Power, Technology and Domination*, Londres, Routledge.

Le Coadic, Y.-F. (2004), *La science de l'information* (première édition en 1994), Paris, PUF, coll. « Que sais-je ? », 128 p.

Lee-Wright P. (2010), « Culture Shock: new media and organizational change in the BBC », in : Fenton N. (dir.), *New Media, Old News: journalism & democracy in the digital age*, London: Sage.

Le Floch P. (1997), *Economie de la presse quotidienne régionale: Déterminants et conséquences de la concentration*, Paris, L'Harmattan, 392 p.

Le Floch P., Sonnac N. (2013), *Economie de la presse à l'ère numérique*, Paris, La Découverte, coll. « Repères », 126 p.

Le Grignou B., Neveu E. (1993, « Intimités publiques. Les dynamiques de la politique à la télévision », *Revue française de science politique*, n°6, p. 940-969.

Lemieux C. (2007), « À quoi sert l'analyse des controverses ? », *Mil neuf cent*, 1, n°25, p. 191-212.

Lemieux C. (2009), *Le devoir et la grâce*, Paris, Economica, 256 p.

Lemieux C. (2010), *Un président élu par les médias ?*, Paris, Presse des Mines, coll. « Sciences sociales », 208 p.

Lessig L. (1999), « Code is Law. On Liberty in CyberSpace », *Harvard Magazine*, traduction française : <http://www.framablog.org/index.php/post/2010/05/22/code-is-law-lessig>

Lévêque F. (2009), « Google double clique et gagne », *Revue Lamy de la Concurrence*, n°18.

Levie F. (2006) *L'Homme qui voulait classer le monde. Paul Otlet et le Mundaneum*, Bruxelles, Les Impressions nouvelles, coll. « Réflexions faites », 352 p.

Lévy P. (1991). « L'hypertexte, instrument et métaphore de la communication », *Réseaux*, vol. 9, n°46-47, p. 59-68.

Levy S. (2009), « Secret of Googlenomics: Data-Fueled Recipe Brews Profitability », *Wired* 16(7).

Levy S. (2011), *In The Plex: How Google Thinks, Works, and Shapes Our Lives*, New York, Simon & Schuster, 432 p.

- Liu D., Chen J.** (2006), « Designing online auctions with past performance information », *Decision Support Systems*, 42, p. 1307–1320.
- Lochard G., Soulages J.-C.** (2006), « Sémiotique, sémiologie et analyse de la communication médiatique », in. Olivesi S. (dir.), *Sciences de l'Information et de la Communication*, Grenoble, Presses Universitaires de Grenoble, p. 229-242.
- Lucas A., Sirinelli P.** (2006), « Droits d'auteur et droits voisins », *Propriétés intellectuelles*, n°20, p. 297-339.
- Luchetta G.** (2012), « Is the Google Platform a Two-Sided Market? », *Working Paper*, Centre for European Policy Studies, disponible ici : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2048683
- Lyons J.** (1978), *Éléments de sémantique*, Paris, Larousse, 296 p.
- Mabi C.** (2011), « Les conceptions de la participation citoyenne inscrites au sein d'un dispositif de concertation », *Journées doctorales GIS Participation*.
- Macchiati A.** (2010), « I motori di ricerca su Internet e il mercato delle news. Profili antitrust e regolamentari », *Mercato Concorrenza Regole*, 3, p. 469-490
- MacDonnell J. C.** (2012), « The Continuing Viability of the Hot News Misappropriation Doctrine in the Age of Internet News Aggregation », *Northwestern Journal of Technology & Intellectual Property*, 255.
- MacGregor P.** (2007) « Tracking the Online Audience », *Journalism Studies*, 8(2), p. 280-98.
- Machill M., Neuberger C., Schweiger W., Wirth W.** (2004), « Navigating the Internet », *European Journal of Communication*, 19, p. 321–347
- Mackenzie D., Millo Y.** (2003), « Construction d'un marché et performance théorique. Sociologie historique d'une Bourse de produits dérivés financiers », *Réseaux*, vol. 21, n°122, p. 13-61.
- MacQuail D.** (2005), *Mass Communication Theory*, 5th ed. London, Sage.
- Mager A.** (2009), « Mediated health: sociotechnical practices of providing and using online health information », *New Media & Society*, 11, p. 1123-1142.
- Mager A.** (2010), « Mediated Knowledge. Sociotechnical practices of communicating medical knowledge via the web and their epistemic implications », Thèse de doctorat, Université de Vienne, Autriche.
- Manara C.**, (2011) « Le droit d'auteur contre l'accès à l'information mondiale ? Le cas des moteurs de recherche », *Revue Internationale de Droit Economique*, 2/t.XXV, p. 143-164.

Marty Emmanuel, Rebillard F., Pouchot S., Lafouge T. (2012), « Diversité et concentration de l'information sur le web. Une analyse à grande échelle des sites d'actualité français », *Réseaux*, vol.6, n°176, p. 27-72.

Mathias P. (2009), *Qu'est-ce que l'Internet ?*, Paris, Vrin, coll. « Chemins Philosophiques », 128 p.

Mathien M. (2007), *Les Journalistes. Histoire, pratiques et enjeux*, Paris, Ellipses, coll. « Infocom », 272 p.

Mattelart A. (2009), *Histoire de la société d'information*, 4^e édition, La Découverte, Paris, coll. « Repères », 128 p.

McCombs M., Shaw D. L. (1972), « The agenda-setting function of the mass media », *Public Opinion Quarterly*, 36, p. 176-185.

McBride K. (2011), « SEO Makes It Too Late for Truth for 'Ground Zero Mosque' », *Poynter*, disponible ici : <http://www.poynter.org/latest-news/making-sense-of-news/105201/seo-makes-it-too-late-for-truth-for-ground-zero-mosque/>

Meadel C. (2010), *Quantifier le public : Histoire des mesures d'audience de la radio et de la télévision*, Paris, Economica, 284 p.

Meisel J. B., Sullivan T. S. (2000). « Portals: The new media companies », *Info - The journal of policy, regulation and strategy for telecommunications*, 2(5), 477-486.

Mercier A. (dir.). (2009). *Le journalisme*. Paris, CNRS Editions, coll. « Les essentiels d'Hermès », 168 p.

Mercier A. (2010). « Défis du nouvel écosystème d'information et changement de paradigme journalistique », *Obsweb*, 42 p., disponible ici : <http://obswebjournalisme.files.wordpress.com/2010/12/dc3a9fis-du-nouveau-ecosysteme-d-information.pdf>

Mettout E. (2010), « Zahia n'a pas écrit ce titre », *Nouvelle formule*, disponible ici : http://blogs.lexpress.fr/nouvelleformule/2010/05/20/zahia_na_pas_ecrit_ce_post/

Mettout E. (2012), « The Google Menace », *Nouvelle formule*, disponible ici : <http://blogs.lexpress.fr/nouvelleformule/2012/10/19/the-google-menace/>

Metzger J.-P. (2006), « L'information-documentation », in : Olivesi S. (dir.), *Sciences de l'Information et de la Communication*, Grenoble, Presses Universitaires de Grenoble, p. 43-61.

Miège B. (2000), « Les apports à la recherche des Sciences de l'Information et de la Communication », *Réseaux*, n°100, p. 549-567.

Miège B. (2004), *L'information-communication, objet de connaissance*, Bruxelles, De Boeck/INA, 248 p.

Miège B. (2006), « Les industries culturelles et médiatiques : une approche socio-économique », in : Olivesi S. (dir.), *Sciences de l'Information et de la Communication*, Grenoble, Presses Universitaires de Grenoble, p. 163-180.

Mill J.S. (1978), *On liberty* (première édition en 1859), Indianapolis, Hackett.

Moe W., Fader P. S. (2004), « Dynamic conversion behavior at ecommerce sites », *Management Science*, vol. 50, n°3, p. 326-335.

Mœglin P. (2012), « Une théorie pour penser les industries culturelles et informationnelles ? », *Revue française des sciences de l'information et de la communication* [En ligne], 1.

Monnoyer-Smith L. (2007), *Les voies de l'expression citoyenne dans les sociétés modernes. Formes et contraintes des dispositifs du débat public*, Mémoire d'Habilitation à Diriger des Recherches, Sciences de l'Information et de la Communication, Laboratoire Costech, UTC Compiègne, soutenu en novembre 2007.

Monnoyer-Smith L. (2013), « Le web comme dispositif: comment appréhender le complexe ? » in : Barats C. (dir.), *Analyser le Web en sciences humaines et sociales*, Paris, Armand Colin.

Morgan W. (2002), « Heterotropes: learning the rhetoric of hyperlinks », *Education, Communication & Information*, 2(2-3), p. 215-233.

Morienval J. (1934), *Les créateurs de la grande presse en France : Emile de Girardin, H. de Villemessant, Moïse Millaud*, Paris, Spes.

Morin E. (1990), « Sur l'interdisciplinarité » Actes du Colloque du Comité National de la Recherche Scientifique Interdisciplinarité, Introduction par François Kourilsky, Paris, Éditions du CNRS.

Moulier Boutang Y. (2007), *Le capitalisme cognitif : La Nouvelle Grande Transformation*, Editions Amsterdam, coll. « Multitude/Idées », 245 p.

Moulier Boutang Y. (2009), « Inescapable Google ? Organization of knowledge, Economic value in cognitive capitalism, and collective intelligence », Conférence *Society of Query, stop searching, start questioning*, Université d'Amsterdam.

Moulier Boutang Y. (2010), *L'abeille et l'économiste*, Paris, Carnets Nords, 255 p.

Moulier Boutang Y., Rebiscoul A. (2009), « Peut-on faire l'économie de Google ? », *Multitudes*, n°36, p. 83-93.

Mouriquand J. (2011), *L'écriture journalistique*, 4^e éd., Paris, PUF, coll. « Que sais-je ? », 128 p.

Mucchielli A. (2001), *Les sciences de l'information et de la communication*, Paris, Hachette, 160 p.

- Muhlmann G.** (2004), *Une histoire politique du journalisme, XIXe -XXe siècles*, Paris, PUF, coll. « Partage du savoir », 247 p.
- Muniesa F., Callon M.** (2009), « La performativité des sciences économiques », in : Steiner P., Vatin F. (dir.), *Traité de sociologie économique*, Paris, PUF, p. 289-324.
- Muniesa F., Millo Y., Callon M.** (2007), « An introduction to market devices », *Sociological Review*, 55 (2), p. 1-12.
- Najork M., Wiener J.** (2001), « Breadth-first search crawling yields high-quality pages. *Proceedings 10th International World Wide Web Conference*, p. 114-118.
- Negroponte N.** (1995), *Being Digital*, New York, Basic Books, 272 p.
- Niles R.** (2010), « Student Journalists Need to Learn SEO More Than They Need AP Style », *Online Journalism Review*.
- Nissenbaum H.** (2004), « Hackers and the contested ontology of cyberspace », *New Media & Society*, vol. 6, n°2, p. 195-217.
- Nora P.** (1972), « L'événement monstre », *Communications*, n°18, p. 162-172.
- Nourot P. M.** (1993) « Historical Perspectives on Early Childhood Education », in : Roopnarine J. L., Johnson J. E. (dir.), *Approaches to Early Childhood Education* (2ème édition), New York, Macmillan, p. 1-32.
- Olivesi S.** (2006), « Les anthropologies de la communication », in : Olivesi S. (dir.) *Sciences de l'information et de la communication. Objets, savoirs et discipline*, Grenoble, Presses Universitaires de Grenoble, p. 181-195.
- Ouakrat A.** (2011), *Publicité en ligne sur les sites de presse issus de l'imprimé. Construction du marché, logiques de fonctionnement et perspectives d'évolution*, Thèse en Sciences de l'information et de la communication, Université Panthéon-Assas, 564 p.
- Ouakrat A., Beuscart J.-S., Mellet K.** (2010), « Les régies publicitaires de la presse en ligne », *Réseaux*, n° 160-161, p. 135-161.
- Page L., Brin S., Motwani R., Winograd T.** (1999), « The PageRank Citation Ranking: Bringing Order to the Web », Technical Report, Stanford InfoLab.
- Pariser E.** (2011), *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, London, Penguin Books, 304 p.
- Peeters H., Charlier P.** (1999), « Introduction. Contribution à une théorie du dispositif », *Hermès*, n°25, p.15-24.

Peguera M. (2009), « When the Cached Link is the Weakest Link : Search Engine Caches Under the Digital Millenium Copyright Act », *Journal of the Copyright Society of the USA*, vol. 56.

Pélissier N., Diallo M. D. (2009), « Les sources cachées du journalisme », in : Simonnot B., Callezot G. (dir.), *L'entonnoir. Google sous la loupe des Sciences de l'Information et de la Communication*, Caen, C&F éditions.

Pélissier N. et alii, « L'information en-ligne : un nouveau paradigme pour le journalisme ? », in : Le Bœuf C. (dir.), *Pragmatique des communications instrumentées*, Paris, L'Harmattan, 2003, p. 21-63.

Perret J.-B. (2004), « Y a-t-il des objets plus communicationnels que d'autres ? », *Hermès*, n°38, p. 121-128.

Perrin R. (2013), « Google News : étude sur l'utilisation de la balise news_keywords en France », Blog *Resoneo*, disponible ici : http://blog.resoneo.com/2013/01/google-news-etude-sur-lutilisation-de-la-balise-news_keywords-en-france/

Plantelin N. (2010), « Optimiser son référencement dans Google News », *Intestable.org*, disponible ici : <http://www.intestable.org/2013/04/optimiser-son-referencement-dans-google-news-143372089>

Plantin J.-C. (2012), *Les pratiques de cartographie numérique en ligne. Expression, remédiation, circulation*, Thèse en Sciences de l'Information et de la Communication, UTC Compiègne, 404 p.

Plesner U. (2009), « An actor-network perspective on changing work practices: Communication technologies as actants in newswork », *Journalism*, 10(5), p. 604-26.

Pollock R. (2010) « Is Google the Next Microsoft: Competition, Welfare and Regulation in Online Search », *Review of Network Economics*, 9(4), p. 1-31.

Proulx S. (2005), « Penser la conception et l'usage des objets communicationnels », in : Saint-Charles J., Mongeau P. (dir.), *Communication. Horizon de recherches et de pratiques*, Québec, Presses de l'Université du Québec, p. 297-318.

Racine B. (2010), *Google et le nouveau monde*, Paris, Plon, coll. « Tribune Libre », 150 p.

Rafiei D., Bharat K., Shukla A. (2010), « Diversifying Web Search Results », *Proceedings of the 19th international conference on World Wide Web*, p. 781-790.

Rebillard F. (2006), « Du traitement de l'information à son retraitement », *Réseaux*, vol. 3, n°137, p. 29-68.

Rebillard F. (2007), *Le Web 2.0 en perspective: une analyse socio-économique de l'internet*, Paris, L'Harmattan, 162 p.

Rebillard F., Smyrnaioi N. (2010), « Les infomédiaires, au cœur de la filière de l'information en ligne », *Réseaux*, vol.2, n°160-161, p. 163-194.

Richard R. (2011), *Google décodé : les 300 critères de classement t. 1&2*, Webmarketing Academy.

Richmond S. (2008), « How SEO Is Changing Journalism », *British Journalism Review*, 19, p. 51-55.

Rieder B. (2005), « Networked Control: Search Engines and the Symmetry of Confidence », *International Review of Information Ethics*, vol. 3, p. 26-32.

Rieder B. (2006) *Métatechnologies et délégation. Pour un design orienté-société dans l'ère du Web 2.0*, Thèse en Sciences de l'Information et de la Communication, Université Paris 8 – Vincennes-Saint-Denis, 388 p.

Rieder B. (2012), « What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index », *Computational Culture*, 28 septembre 2012, disponible ici : http://computationalculture.net/article/what_is_in_pagerank

Rieder B., Sire G. (2013) « Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google's Tangled Position on the Web », *New Media & Society*, (publié en ligne).

Rieffel R. (2008), « La métamorphose de l'information : de sa production à sa coproduction », in : Greffe X., Sonnac N. (dir.), *Culture Web*, Paris, Dalloz.

Robinson S. (2007), « Someone's gotta be in control here. The institutionalization of online news and the creation of a shared journalistic authority », *Journalism practice*, vol. 1, n°3, p. 305-321.

Rogers R. (2010), « Mapping public web space with the Issuecrawler », in : Reber B., Brossard C. (dir.), *Digital cognitive technologies : epistemology and the knowledge economy*, London/ Hoboken, Wiley, p. 115-126.

Röhle T. (2009), « Dissecting the Gatekeepers. Relational Perspectives on the Power of Search Engines », in : Becker K, Felix S. (dir.) *Deep Search. The Politics of Search beyond Google*, Innsbruck, StudienVerlag, p. 117-132.

Röhls H. (1994), « Maria Montessori », *Perspectives : revue trimestrielle d'éducation comparée*, vol. XXIV, n° 1-2, 1994, p. 173-188.

Ruellan D., Thierry D. (1998), *Journal local et réseaux informatiques : travail coopératif, décentralisation et identité des journalistes*, L'Harmattan, 208 p.

Samuelson P. A. (1964), « Public Goods and Subscription TV : Correction of the Record », *Journal of Law and Economics*, 71, p. 81-84.

- Schmitz Weiss A., Domingo D.** (2010), « Innovation processes in online newsrooms as actor-networks and communities of practice », *New Media & Society*, 12(7), p. 1156-1171.
- Schwartzenberg E.** (2007), *Spéciale dernière : Qui veut la mort de la presse quotidienne française ?*, Paris, Calmann-Lévy, 281 p.
- Serres A.** (2002), « Posture critique et sociologie de la traduction », *Les Cahiers du CERCOR*, n° 1, p. 76-84.
- Serres M.** (1969), *La Communication, Hermès I*, Paris, Editions de Minuit, 248 p.
- Serres M.** (1974), *La Traduction, Hermès III*, Paris, Editions de Minuit, 272 p.
- Serres M.** (1980), *Le Parasite*, Paris, Grasset, 462 p.
- Serres M.** (1994), *Eclaircissements. Entretiens avec Bruno Latour*, Paris, Flammarion, coll. « Champs », 300 p.
- Serres M.** (2007), « Les nouvelles technologies : révolution culturelle et cognitive », Conférence à l'Institut national de la recherche en informatique et en automatique (INRIA) – Forum des 40 ans.
- Shkapenyuk V., Suel T.** (2002), « Design and implementation of a high performance distributed web crawler », *Proceedings of the 18th International Conference on Data Engineering* p. 357-368.
- Shoemaker P. J., Eicholz M., Kim E., Wrigley B.** (2001) « Individual and routine forces in gatekeeping », *Journalism & Mass Communication Quarterly*, 78 (2), p. 233-246.
- Shoemaker P. J., Reese S. D.** (1996), *Mediating the Message* (2nd éd.). White Plains, Longman, 314 p.
- Silverstein C., Henziger M., Hannes M., Moricz M.** (1999), « Analysis of a very large web search engine query log », *SIGIR Forum* 33, p. 6-12.
- Siméant J.** (1992), « Déontologie et crédibilité. Le réglage des relations professionnelles au CFJ », *Politix*, vol. 5, n°19, p. 37-55.
- Siritzky S., Roth F.** (1979), *Le roman de L'Express, 1953-1978*, Paris, Atelier Marcel Julian.
- Smyrnaiois N.** (2005), *L'industrie de l'éphémère. Emergence et consolidation de modèles diversifiés de production et de diffusion de l'information en ligne*, Thèse en Sciences de l'information et de la communication, Université Grenoble 3, 615 p.

- Smyrnaiois N., Boure R.** (2006), « L'infomédiation de l'information en ligne, le cas des filiales françaises de Google et Yahoo », *Document numérique et société*, Paris, ADBS Éditions, p. 43-55.
- Smyrnaiois N., Marty E., Rebillard F.** (2010), « Does the Long Tail apply to online news? A quantitative study of French-speaking news websites », *New Media & Society*, 12 (8), p. 1244 -1261.
- Smyrnaiois N., Rebillard F.** (2009), « L'actualité selon Google », *Communication & langages*, n° 160, p. 95-109.
- Smyrnaiois N., Rebillard F.** (2011), « Entre coopération et concurrence : Les relations entre infomédiaires et éditeurs de contenus d'actualité », *Concurrences*, n°3-2011, p. 7-18.
- Soloski J.** (1989), « News reporting and professionalism: Some constraints on the reporting of news », *Media, Culture and Society*, 11 (1), p. 207-228.
- Sookman B.** (2011), « Is Google News Legal ? », Blog de Barry Sookman, disponible ici : <http://www.barrysookman.com/2011/05/17/is-google-news-legal/>
- Sonnac N.** (2009), « L'économie de la presse : vers un nouveau modèle d'affaires », *Les Cahiers du Journalisme*, 20, p. 22-42.
- Sonnac N., Gabszewicz J.** (2013), *L'industrie des médias à l'ère du numérique*, Paris, La Découverte, coll. « Repères », 126 p.
- Sonnac N.** (2012), « Information, modèles d'affaires et concurrence », in : Degand A., Grevisse B. (dir.), *Journalisme en ligne. Pratiques et recherches*, Bruxelles, Deboeck, coll. « Info&Com », p. 179-207.
- Stallman R.** (2001) « Science Must Push Copyright Aside », disponible ici : <http://www.nature.com/nature/debates/e-access/Articles/stallman.htm>
- Steensen S.** (2011) « Online Journalism and the Promises of New Technology » *Journalism Studies*, 12(3), p. 311-327.
- Stigler G. J.** (1961), « The economics of information », *Journal of Political Economy*, 69(3), p. 213-225.
- Stray J.** (2010) « Why Link Out? Four journalistic purposes of the noble Nieman Journalism Lab », *Nieman Journalism Lab*, <http://www.niemanlab.org/2010/06/why-link-out-four-journalistic-purposes-of-the-noble-hyperlink>, accessed 14 June 2010.
- Strowel A.** (2011), *Quand Google défie le droit. Plaidoyer pour un Internet transparent et de qualité*, Bruxelles, Larcier, 240 p.

- Sullivan D.** (2007), « ACAP Launches, Robots.txt 2.0 For Blocking Search Engines? », *Search Engine Land*, disponible ici : <http://searchengineland.com/acap-launches-robotstxt-20-for-blocking-search-engines-12802>
- Sullivan D.** (2009a), « Dear WSJ: To Avoid Google Disease, Please Put A Condom On Your Content », *Daggle*, disponible ici : <http://daggle.com/dear-wsj-avoid-google-disease-put-condom-content-1451>
- Sullivan D.** (2009b), « Fights in the Google monopoly debate miss key points », *Search Engine Land*, disponible ici : <http://searchengineland.com/fights-over-the-google-monopoly-miss-key-points-18916>
- Sullivan D.** (2012), « Two Weeks In, Google Says “Search Plus Your World” Going Well, Critics Should Give It Time », *Search Engine Land*, disponible ici : <http://searchengineland.com/two-weeks-in-google-search-plus-your-world-109527>
- Talim J., Liu Z., Nain P., Coffman E.** (2001), « Controlling robots of web search engines », *Proceedings ACM Sigmetrics / Performance Evaluation Review*, vol. 29, n°1, p. 236-244.
- Tankard J. W., Ban H.** (1998), « Online Newspapers: living up to the potential? », *Annual conference of the AEJMC*, Baltimore, MD.
- Taveaux-Grandpierre K.** (2002), « De l’abonnement à la vente au numéro : le cas du Petit Journal (1863-1914) », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 159-183.
- Telang R., Mukhopadhyay T., Wilcox R.** (1999), « An empirical analysis of the antecedents of Internet search engine choice », *Workshop on Information Systems and Economics*, Charlotte, North Carolina.
- Telang R., Rajan U., Mukhopadhyay T.** (2004), « The Market Structure for Internet Search Engines », *Journal of Management Information Systems* / Fall 2004, vol. 21, n°2, p. 137-160.
- Toussaint Desmoulins N.** (2002) « Les évolutions des systèmes de distribution : un enjeu majeur pour la presse de demain », in : Feyel G. (dir.), *La distribution et la diffusion de la presse du XVIIIe siècle au IIIe millénaire*, Paris, Editions Panthéon Assas, p. 409-420.
- Tremayne M.** (2005), « News Websites as Gated Cybercommunities », *Convergence*, 11(3), p. 28-39.
- Tsui L.** (2008), « The Hyperlink in Newspapers and Blogs », in : Turow J., Tsui L. (dir.), *The Hyperlinked Society*, Ann Arbor, University of Michigan Press, p. 70-83.
- Ulken E.** (2005), « Non-Traditional Sources Cloud Google News Results », *Online Journalism Review*, disponible ici : <http://www.ojr.org/ojr/stories/050519ulken/>

- Vaidhyanathan S.** (2011), *The Googlization of everything: (and why we should worry)*, Berkeley / Los Angeles, University of California Press, 265 p.
- Van Asbroeck B., Cock M.** (2007), « Belgian newspapers v Google News: 2-0 », *Journal of Intellectual Property Law & Practice*, vol. 2, n°7, p. 463-466.
- Van Couvering E.** (2004), « New Media? The Political Economy of Internet Search Engines », *Conference of the International Association of Media & Communications Researchers (IAMCR)* Porto Alegre, Brazil, July 25-30.
- Van Couvering E.** (2008), « The History of the Internet Search Engine: Navigational Media and the Traffic Commodity », in : Zimmer M., Spink A. (dir), *Web Search. Multidisciplinary Perspectives*. Berlin / Heidelberg, Springer, p. 177-206.
- Van Couvering E.** (2009), « Search Engine Bias. The Structuration of Traffic on the World-Wide Web », Thèse en Economie Politique, London School of Economics.
- Van Couvering E.** (2011), « Navigational Media: The Political Economy of Online Traffic », in : Winseck D., Yin D. Y. (dir.), *Media Political Economies: Hierarchies, Markets and Finance in the Global Media Industries*, London, Bloomsbury.
- Van Der Wielen A.** (2010), « La transposition du test en trois étapes de la Convention de Berne en droit national français et anglais à travers la directive 2001/29/EC », *Société de l'information, droits et médias* (Université Paris Ouest-Nanterre-La Défense).
- Varian H. R.** (2007), « Position Auctions », *International Journal of Industrial Organization*, 25, p. 1163-1178.
- Varian H. R.** (2010), *Intermediate Microeconomics. A Modern Approach. Eighth Edition*, New York / London, Norton & Company, 740 p.
- Virilio P.** (1988), *La machine de vision*, Paris, Galilée, 158 p.
- Vise D. A.** (2005), *The Google Story: inside the hottest business, media and technology success of our time*, New York, Random House, 336 p.
- Wauters R.** (2010), « Report: 44% of Google News Visitors Scan Headlines, Don't Click Through », *TechCrunch*, <http://techcrunch.com/2010/01/19/outsell-googlenews/>
- Wauthy X.** (2008), « Concurrence et régulation sur les marchés de plate-forme: une introduction », *Reflets et perspectives de la vie économique*, 148(1), p. 39-54.
- Weber M.** (2012), « Newspapers and the Long-Term Implications of Hyperlinking », *Journal of Computer-Mediated Communication*, 17(2), p. 187-201.
- Weiss A., Domingo D.** (2010), « Innovation processes in online newsrooms as actor-networks and communities of practice », *New Media Society*, 12(7), p. 1156-1171.

White A. (2008) « Search Engines: Left Side Quality Versus Right Side Profits », Working Paper, disponible sur : <http://ssrn.com/abstract=1694869>

Wiggins R. W. (2001), « The Effects of September 11 on the Leading Search Engine », *First Monday*, vol. 7, n°10.

Winkler H. (1999) « Search Engines: Metamedia on the Internet? » in : Bosna, J. *et alii* (dir.), *Read me! Filtered by Nettime. ASCII Culture and the Revenge of Knowledge*, New York, p. 29-37.

Wojcicki S. (2007), « The Next Step in Google Advertising », disponible ici : <http://googleblog.blogspot.com/2007/04/next-step-in-google-advertising.html>

Woolgar S. (1991), « The Turn to Technology in Social Studies of Science », *Science Technology Human Values*, vol. 16, n°20, p. 20-50.

Wolton D. (2000), *Internet, et après ? Une théorie critique des nouveaux médias*, Paris, Flammarion, coll. « Champs essais », 240 p.

Wolton D. (2003), « Journalistes, une si fragile victoire... », *Hermès*, n°35, p. 9-21.

Yang S., Ghose A. (2010) « Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? », *Marketing Science*, 29(4), p. 602-623.

Yao S., Mela C. F. (2011), « A Dynamic Model of Sponsored Search Advertising », *Marketing Science*, vol. 30, n°3, p. 447-468.

Yoneyama J. (1997) « Computer Systems as Text and Space : Toward a Phenomenological Hermeneutics of Development and Use », in : Bowker G., Leigh Star S., Gasser L., Turner W. (1997), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, p. 105-120.

Zelizer B. (1992) *Covering the Body: The Kennedy Assassination, the Media and the Shaping of Collective Memory*, Chicago, University of Chicago Press, 308 p.

Sites à propos du SEO

Abondance : <http://www.abondance.com>

Café-référencement : <http://www.cafe-referencement.com>

Search Engine Land : <http://searchengineland.com>

SEOMoz : <http://moz.com>

SEOSphère : <http://www.seosphere.com>

Webrankinfo : <http://www.webrankinfo.com>

Etudes / Rapports

AT Internet (2013), *Baromètre des moteurs de recherche*,
<http://www.atinternet.com/ressources/ressources/etudes-publiques/barometre-des-moteurs/>

Idate (2010) « Le marché mondial de la publicité en ligne. L'essor des nouveaux marchés : Mobile, Réseaux sociaux & Vidéo en ligne », Market & Data Report, M10114, Octobre, 155p.

Optify (2011), *Study: How the New Face of SERPs Has Altered the CTR Curve*.
<http://www.optify.net/inbound-marketing-resources/new-study-how-the-new-face-of-serps-has-altered-the-ctr-curve>

SearchMetrics (2012), *Social Media, Backlinks and Classic SEO: U.S. and UK SEO Ranking Factors 2012*, <http://blog.searchmetrics.com/us/2012/06/07/us-and-uk-seo-ranking-factors-2012/>

Pew Research Center's Project for Excellence in Journalism (2013), *The State of The News Media*, <http://stateofthemedias.org>

ZenithOptimedia (2011) *Quadrennial Events to Help Ad Market Grow in 2012 Despite Economic Troubles*,
<http://zenithoptimedia.blogspot.com/2011/12/quadrennial-events-to-help-ad-market.html>

Sites de Google

Centre d'aide actualité : <https://support.google.com/news/publisher/?hl=fr>

Centre d'aide pour les webmasters : <https://support.google.com/webmasters/?hl=fr>

Blog politique : <http://googlepublicpolicy.blogspot.fr>

Google's SEO Starter Guide :
<http://googlewebmastercentral.blogspot.fr/2008/11/googles-seo-starter-guide.html>

Matt Cutts : <http://www.mattcutts.com/blog/>

Research at Google : <http://research.google.com>

Investor Relations : <http://investor.google.com>

Annexes

A. FIGURES, TABLEAUX ET CADRES

Figures

- Figure 1. Capture d'écran : liste de liens sur Google Search. (« UMP » - 20/11/12) /p. 149
- Figure 2. Capture d'écran : liste de liens sur Google Search. (« Les Echos » - 20/11/12)/p. 151
- Figure 3. Capture d'écran : lien vers un article du blog de Dan Sullivan /p. 152
- Figure 4. Capture d'écran : lien vers un article du site « Thaifood » sur Google Search /p. 153
- Figure 5. Capture d'écran : Transfert de requête vers Google Actualités / p. 155
- Figure 6. Capture d'écran : Transfert de requête vers Google vidéos / p. 156
- Figure 7. Capture d'écran : Proposition de reformulation / p. 156
- Figure 8. Sitemap Actualités : exemple / p. 164
- Figure 9. Capture d'écran : Author Rank Google Actualités / p. 171
- Figure 10. Capture d'écran : Le choix des rédactions / p. 175
- Figure 11. Schéma simplifié : Arbre de navigation / p. 186
- Figure 12. Marché mondial de la publicité en ligne (2010) / p. 188
- Figure 13. Capture d'écran : Interface annonceurs / p. 191
- Figure 14. Répartition des revenus Google/Editeur / p. 192
- Figure 15. Schéma : Diversification concentrique (1) / p. 195
- Figure 16. Schéma : Diversification concentrique (2) / p. 198
- Figure 17. Schéma : Modèle économique du dispositif d'infomédiation (1) / p. 201
- Figure 18. Schéma : Effets de réseau croisés du dispositif d'infomédiation (1) /p. 204
- Figure 19. Schéma : Modèle économique du dispositif d'infomédiation (2) / p. 205
- Figure 20. Schéma : Effets de réseau croisés du dispositif d'infomédiation (2) / p. 206
- Figure 21. Schéma : Modèle économique du dispositif d'infomédiation (3) / p. 206
- Figure 22. Schéma : Effets de réseau croisés du dispositif d'infomédiation (3) / p. 207
- Figure 23. Schéma : Modèle économique du dispositif d'infomédiation (4) / p. 208
- Figure 24. Schéma : Effets de réseau croisés du dispositif d'infomédiation (4) / p. 209

- Figure 25. Schéma : Modèle économique du dispositif d'infomédiation (Complet) /p. 209
- Figure 26. Schéma : Effets de réseau croisés du dispositif d'infomédiation (Complet) / p. 210
- Figure 27. Schéma : Arbre des opportunités / p. 213
- Figure 28. Capture d'écran : Requête « Carrefour » (12/10/2012) / p. 215
- Figure 29. Carte interactive du lobbying contre la loi allemande *Lex Google* / p. 249
- Figure 30. Le dispositif d'infomédiation schématisé / p. 286
- Figure 31. Graphe sociotechnique : modifications d'algorithme (2003-2012) / p. 312
- Figure 32. Structure d'audience approximative de *lexpress.fr* / p. 316
- Figure 33. Capture d'écran : Sitemap Article *Lemonde.fr* / p. 334
- Figure 34. Calcul du PageRank (1) / p. 336
- Figure 35. Calcul du PageRank (2) / p. 336
- Figure 36. Calcul du PageRank (3) / p. 337
- Figure 37. Capture d'écran : *Lefigaro.fr* (5 juin 2013) / p. 349
- Figure 38. Capture d'écran : Conjugueurs (« conjugaison verbe réfléchir » - 06/06/13) / p. 358
- Figure 39. Capture d'écran : Verbes thématiques (*Nouvelobs.com*) / p. 359
- Figure 40. Code HTML : *Lemonde.fr* (17/10/12) / p. 360
- Figure 41. Capture d'écran : *Leparisien.fr* (11/10/12) / p. 366
- Figure 42. Capture d'écran : *L'Entreprise* sur Google Actualités / p. 380
- Figure 43. Capture d'écran : *L'Expansion* sur Google Actualités / p. 380
- Figure 44. Capture d'écran : *Studio Ciné Live* sur Google Actualités / p. 381
- Figure 45. Capture d'écran : logo de *Studio Ciné Live* sur *Lexpress.fr* / p. 382
- Figure 46. Capture d'écran : le *Sitemap* à tiroirs de *Lexpress.fr* / p. 383
- Figure 47. Capture d'écran : Express Yourself (28/06/13) / p. 444
- Figure 48. Captures d'écran : Express Yourself sur Google Search et Google Actualités/p. 445
- Figure 49. SEO : échanges stratégiques / p. 469

Tableaux

- Tableau 1. Employés de Google interrogés / p. 64
- Tableau 2. Employés des SIPI interrogés / p. 68-69
- Tableau 3. Consultants interrogés / p. 70
- Tableau 4. Employés de RMF interrogés / p. 71-72
- Tableau 5. Champ de manœuvres des éditeurs (Etablir, Spécifier, Renforcer) / p. 282-283
- Tableau 6. Champ de manœuvres des éditeurs (Réduire, Interrompre, Rompre) / p. 283
- Tableau 7. Champ de manœuvres de Google (Pouvoir de faire faire) / p. 284
- Tableau 8. Champ de manœuvres de Google (Pouvoir sur) / p. 285
- Tableau 9. Optimisation des titres / p. 324
- Tableau 10. Usage des balises méta-tags (1^{er} octobre 2012) / p. 325-326
- Tableau 11. Liens entrants et sortants / p. 344
- Tableau 12. Moyennes du nombre de liens entrants et sortants / p. 344
- Tableau 13. Liens vers pages d'atterrissage / p. 350
- Tableau 14. Moyennes du nombre de liens vers pages d'atterrissage / p. 350
- Tableau 15. Liens visibles et invisibles / p. 354
- Tableau 16. Positions dans l'arbre de navigation / p. 366
- Tableau 17. Les 4 titres et la Coupole / p. 378
- Tableau 18. Les 4 sous-rubriques de la Coupole / p. 380

Cadres

- Cadre 1. Google, le pluralisme des sources et la diversité des sujets / p. 402-405
- Cadre 2. Quand Google couronne la faute journalistique / p. 409
- Cadre 3. Nombre et fonction des liens / p. 419-420
- Cadre 4. Les poissons-pilotes de Mediapart / p. 448-449

B. CARNETS D'ENTRETIENS

Dirigeants

Organisation

Comment votre entreprise est-elle organisée ? rédaction web ? rédaction papier ? combien de journalistes ? un référenceur ? Quelles relations entre les journalistes et le référenceur ?

Quelle est la structure d'audience ?

Que représente la part venant de Google Search ? de Google News ?

Vous sentez-vous dépendant de Google ?

Que faites-vous en terme d'acquisition d'audience ?

Combien coûte la stratégie de SEO ? Pour quels résultats espérés ?

En êtes vous satisfaits ?

La stratégie de SEO est elle justifiée autrement que pour des raisons financières ?

Journalistes

Les journalistes sont-ils formés à l'écriture pour Google ? Par qui ? comment ?

Que leur impose-t-on ? que leur recommande-t-on ? que refusent-ils ?

Dans quelle mesure demeurent-ils libres de choisir eux-mêmes les sujets qu'ils veulent traiter

Y-a-t-il des réfractaires ?

Avez-vous un ou plusieurs Front Page Editor ? Comment leur rôle s'articule-t-il avec ceux du référenceur et des journalistes ?

Relation avec Google

Google est-il un concurrent ou un allié ?

Êtes vous sur Google AdSense ? Doubleclick ? pourquoi ?

Si oui, pensez-vous que cela peut aider à la captation de trafic par Google ?

Dans quelle mesure les moteurs de recherche fournissent-ils un service concurrent du vôtre ?

Selon vous, cassent-ils les prix sur le marché de la publicité ?

Est-ce que vous pensez que les moteurs gagnent de l'argent grâce à vos contenus ?

Que pensez-vous de l'accord signé récemment en France ? le fonds de 60 millions d'euros ? le deuxième volet (accord publicitaire)

Que pensez vous des problèmes de droit de la propriété ? (presse belge)

Que pensez-vous du projet allemand « Lex Google » ?

Que pensez-vous d'ACAP ?

Que pensez vous de Google Pass Media ?

Comment voyez-vous l'avenir de votre relation avec Google ? Que reprocheriez-vous à Google ? Qu'est-ce qui vous semble au contraire louable dans son action ?

Concurrence

Que pensez-vous des stratégies SEO de vos principaux concurrents ?

Que pensez-vous des acteurs qui misent tout sur le référencement (fermes de contenus) ?

Référenceurs

Profil

Quelle est votre formation ? votre parcours professionnel ?
Pourquoi vous êtes vous intéressé au SEO ?
Comment avez-vous été formé ?
Comment et pourquoi avez-vous été embauché par un éditeur de presse ?
Combien gagnez-vous ? A expérience égale, est-ce plus ou moins que les journalistes ?
De qui dépendez-vous dans l'entreprise ?
Comment votre équipe est-elle organisée ?
Quels sont vos partenaires ?
Avez-vous un ascendant hiérarchique sur les journalistes, ou bien eux sur vous ?
Avez-vous des objectifs d'audience fixés au préalable ? comment et par qui sont-ils fixés ?
Avez-vous droit à un bonus si vous les dépassez ?
Comment sont mesurés les résultats de vos actions exactement ? par qui ?
Êtes-vous *white* ou *black hat* ? quelle est la différence ?
Comment vous tenez-vous informé des nouveautés en matière de SEO ?
Pouvez-vous me raconter une journée type ?
Traitez-vous plus avec la technique ou avec les journalistes ? ou les deux ? et le marketing ?

SEO

Quelles sont les particularités du SEO pour la presse ?
Combien cela coûte-t-il d'être bien référencé (notamment quand on est un site de presse) ?
Faut-il internaliser le SEO ? si oui, pourquoi ? comment ?
Peut-on, si on est un site de presse, se passer du référencement ? pourquoi ? comment ?
Qu'est-ce que c'est qu'un site bien construit pour le référencement (très précisément) ? et mal construit ? Quelles sont les erreurs qu'on retrouve dans les sites de presse ? dans les articles ?
Qu'est-ce qu'un article bien écrit pour le référencement ?
Quels sont les critères qui comptent le plus ?
Est-ce que ce sont les mêmes pour Google que pour Bing ?
A quel point le degré de fraîcheur compte ?
Les commentaires font-ils remonter une page ?
A quel point le buzz sur les réseaux sociaux est-il pris en compte ?
A quel point les liens qu'on fait vers l'extérieur comptent ? et les liens vers soi-même ?
Conseillez-vous de faire des liens entrants essentiellement ou non ?
Quelle est votre stratégie Sitemap ?
Quelle est votre stratégie pages d'atterrissage ?
Quel usage des méta-tags ? des microdonnées ?
Quelle est la fréquence du crawling sur un site de presse ? pourquoi change-t-elle ?
Qu'est-ce qu'écrire pour Google Search ? Google News ? (est-ce la même chose ?)
Comment est-on référencé sur Google News ? quels sont les critères très précisément ?
Comment optimiser la captation ? à quel prix ?
Comment sont référencés les autres supports (vidéo, radio) que le texte écrit ?
Vendez-vous/achetez-vous des liens ? Préconisez vous d'en vendre/acheter ?
Un site client d'AdSense ou de DoubleClick a-t-il des chances d'être mieux référencé ?

Ethique

Y a-t-il une éthique du SEO ?

Quelles pratiques seraient inacceptables à votre avis ?

Qui sont les adversaires des journalistes en matière de référencement ?

Le *black SEO* existe-t-il dans la presse embauchant des journalistes professionnels ?

Comment se passe la concurrence entre *black hat* et *white hat SEO* ? Quel est le rôle des journalistes dans cette concurrence ?

Les fautes et les rapports avec les journalistes

Quelles sont les mauvaises pratiques les plus usuelles ?

Où est-ce que ça « coince » avec les journalistes ? que font-ils mal ?

Quel genre de conflit entre vous et les journalistes ? des exemples précis ?

Comment leur signifiez-vous qu'ils ont fait une faute ?

Comment le SEO est-il conciliable avec la déontologie des journalistes ?

Comment collaborez-vous avec les *Front Page Editors* ?

Les journalistes se chargent-ils eux-mêmes des balises méta textes ? rentrent-ils dans le code ? comment ? pourquoi ?

Quel pouvoir d'opposition les journalistes ont vis-à-vis de vous ?

Que faites-vous pour concilier, ou réconcilier, leurs actions et les vôtres ?

Google

Que pensez-vous de la stratégie de diversification de Google ?

Que pensez-vous de la concentration sur le marché des moteurs ?

Quel est le pouvoir de Google, et votre pouvoir par rapport à la firme ?

Que pensez-vous de la communication, ou de l'absence de communication, au sujet de l'algorithme d'appariement ?

Journalistes

Profil

Quel est l'intitulé de votre poste ?

De quoi êtes-vous en charge ?

Depuis combien de temps êtes-vous employé ici ?

Que faisiez-vous avant ?

Quelle est votre formation initiale ?

Comment est organisée votre équipe ? la direction ? la technique ? le marketing ?

Comment est organisé le travail ?

Mesure d'audience

Quelle est la part de votre audience qui provient de Google Search ? Google News ?

Avez-vous accès aux mesures d'audience ? Si oui, les consultez-vous régulièrement ? pourquoi ? Si non, pourquoi ?

Qu'est-ce que cela représente pour vous ?

Y a-t-il des objectifs d'audience ? Si oui, que se passe-t-il lorsqu'ils ne sont pas atteints ?

SEO

Avez-vous été formé à « l'écriture Google » ?

Qu'est-ce qu'un article bien écrit pour le web ? pour Google ? pour vous ? est-ce la même chose ?

Qu'est-ce qu'un article mal écrit, mal optimisé ? Pourquoi ? Que se passe-t-il dans ce cas ?

Qu'est-ce qui justifie qu'on soit bien référencé sur Google Search et Google News ?

Que se passe-t-il en cas de faible audience / référencement ?

En plus d'écrire un article, que faites-vous au moment de la mise en ligne ? tags ? méta informations ? Quelles sont les bonnes pratiques ? les mauvaises ? que se passe-t-il en cas de faute ? Que pensez-vous de votre CMS ?

Comment les journalistes concilient-ils les prérogatives liées au référencement et leur déontologie ?

Quels sont les conflits possibles ?

Comment se passe votre relation avec les équipes techniques ? la direction ? la rédaction papier ?

Qui contrôle l'optimisation des articles pour Google ? comment exactement ?

Est-ce qu'un article peut être changé parce que son audience provenant de Google est jugée trop faible ?

Pertinence

Qu'est-ce que la pertinence d'une information ? Est-ce que vous trouvez que ce qu'on sait du calcul de pertinence effectué par Google serait une bonne mesure de ce que vous définissez vous-mêmes comme la pertinence ?

C'est quoi pour vous la déontologie du journaliste ? Est-ce que vous pensez que Google peut l'entraver ? la course au clic ?

Quel rôle le journaliste joue dans la société ? Quel rôle joue Google ? S'agit-il de rôles complémentaires ou concurrents ?

Moteurs de recherche

Que pensez-vous de la manière de s'informer qui consiste à formuler une requête à partir de ce que l'on sait déjà à propos d'un sujet d'actualité ?

Les moteurs de recherche sont-ils selon vous les nouveaux gatekeepers de l'espace public ?

Ce sont des alliés selon vous ?

Contribuent-ils à assurer votre légitimité ? L'entravent-ils ?

Quand des sites amateurs ou de communicants sont mis sur le même plan que des sites employant des journalistes professionnels, qu'en pensez-vous ?

Que pensez-vous de la concentration sur le marché des moteurs de recherche ?

Que pensez-vous du fait que ces firmes soient américaines

Que pensez-vous de Yahoo et Bing ?

Google est-il selon vous un média ? interprète-t-il l'actualité ?

Les moteurs ont-ils selon vous un rôle à jouer dans l'agenda-setting ?

La production journalistique et Google : chercher à ce que l'information soit trouvée

Nous cherchons dans ce travail à détricoter la relation à la fois compétitive et coopérative, indifféremment technique, économique, juridique, sociale, politique et résolument communicationnelle de Google et des éditeurs de presse. Pour cela, après avoir historicisé la rencontre de deux univers singuliers, nous décrivons ce que les éditeurs peuvent faire pour franchir le prisme du moteur de recherche et y optimiser la visibilité de leur production. Nous tâchons ensuite de décrypter ce que la firme Google peut faire faire aux éditeurs en analysant leurs relations de pouvoir, leurs incitations, leurs projets et leurs environnements informationnels respectifs. Enfin, nous rendons compte de ce que les éditeurs français issus de la presse imprimée font effectivement : ce qu'ils communiquent à Google, par quels moyens et à quel prix, pour quels résultats espérés, à l'issue de quelles concessions, quels détours, quelles contestations. Nous expliquons comment les conditions et les modalités de captation du trafic sont susceptibles d'influencer la valorisation du contenu, l'organisation de sa production, la structure du site, les pratiques journalistiques et les lignes éditoriales. Nous montrons qu'un aller-retour performatif se crée entre énoncés et conditions d'énonciation, agissant *par* et *sur* les textes, les architextes et les hypertextes. En somme, c'est à la compréhension de ce que deviennent l'actualité et ceux qui la mettent en récit, dès lors qu'ils cherchent à ce que l'information qu'ils produisent soit trouvée par les utilisateurs de Google, que notre thèse est consacrée.

Descripteurs : Google ; journalisme ; SEO ; Internet ; moteurs de recherche ; hyperliens ; HTML ; marchés multiversants

Google and journalistic production: Making news findable

In this thesis, we aim to disentangle the cooperative but also competitive relationship between Google and news publishers, which is at the same time technical, economic, legal, social, political and certainly communicational. In order to do so, we trace the historical development of two singular universes, describing what publishers can do to overcome the search engine and optimize their ranking. We then analyse how Google can influence publishers' conduct, by studying power relations, respective incentives, aims, and informational and socio-economic backgrounds. Finally, we report on actual practices of French traditional news publishers: what they communicate to Google, by which means and at what price, for which expected results, after which concessions, detours and controversies. Thus, we explain how search engine optimization is likely to affect the way content is valued, its production organisation, the website's structure, journalists' practice and editorial policy. We show a back and forth movement between performative utterances and performed circumstances, having an effect *on* and *by* texts, architexts and hypertexts. To sum up, this thesis is dedicated to understanding what happens to news and publishers once they strive for their information to be found by Google's users.

Keywords : Google ; journalism ; SEO ; Internet ; search engines ; hyperlinks ; HTML ; multi-sided markets