

DANS LES RAMURES DE L'ARBRE HYPERTEXTE
Analyse des incitations générées par l'opacité du moteur Google

Guillaume Sire

Maître de Conférences

Centre d'Analyse et de Recherche Interdisciplinaires sur les Médias

Institut Français de Presse / Université Paris 2 – Panthéon Assas

guillaumesire@gmail.com

Bernhard Rieder

Assistant Professor

Media Studies Department, University of Amsterdam

rieder@uva.nl

Résumé

Nous nous intéressons aux incitations au biais dont le traitement de l'information par le moteur de recherche Google peut faire l'objet. Nous montrons qu'il peut exister certaines distorsions concernant le marché de la publicité d'une part, et, d'autre part, l'information elle-même. Etant donné l'opacité du fonctionnement du moteur et la difficulté d'appréhender la relation économique et technique qui se noue entre Google, les éditeurs et les internautes, nous proposons un outil conceptuel nommé « arbre de navigation » basé sur trois éléments : les capacités, les probabilités et les opportunités. Grâce à cet outil, nous montrons pourquoi Google a intérêt à avantager certains sites dans les résultats de son moteur, et pourquoi les éditeurs peuvent penser qu'ils ont intérêt à partager avec Google une partie de leurs revenus. Nous ne prétendons pas que les incitations à privilégier certains sites influencent effectivement le résultat de l'action du moteur Google, mais plutôt qu'elles font partie de ce résultat. Dès lors, il convient d'interroger le rôle joué par la firme à la lumière du fait que les éditeurs, parce qu'ils ont besoin de recevoir du trafic sur leurs pages, agissent en fonction de ce qu'ils savent *avec certitude* au sujet de Google (des incitations poussent la firme à avantager ses partenaires économiques), et de ce qu'ils ignorent (les avantage-t-elle effectivement ?). Une observation en ligne de 20 sites de presse nous permet de montrer que les principaux éditeurs de presse français sont effectivement des partenaires économiques de Google, et d'expliquer comment Google peut être à la fois un concurrent, un partenaire et un client des éditeurs sur le marché de la publicité.

Abstract

Does Google have incentives to bias the treatment of information? With Google's numerous areas of activity, we intend to demonstrate that it would have the possibility to distort both the advertising and the information market. Google's way of working is largely opaque and it is difficult to get to the bottom of the economic and technical relationship between Google, publishers, and users. We have therefore developed a conceptual tool we call the "navigation tree" to show why it can be in Google's interest to favour some websites in the results it provides, and why publishers can think it is in their interest to share a part of their income with Google. The navigation tree is based on three elements: abilities, probabilities, and opportunities. We then claim that it is necessary to examine the role played

by the firm considering publishers' need to generate traffic to their pages and the fact that publishers may take action based on what they know with certainty about Google (incentives push the firm to favour its business partners) and on what they do not know (does it favour them?). Online observation of 20 news websites allows us to show that the major French news publishers are in fact Google's business partners. Google has therefore economic reasons to favour them in its results, whether it effectively does so or not. Eventually, we explain how Google can be all at once a competitor, a partner, and even a customer for publishers on the advertising market.

INTRODUCTION

« 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: 2)

En générant des liens qu'il hiérarchise sous forme de liste, le moteur Google produit un énoncé à propos de l'information disponible sur le web. La teneur de ce traitement dépend à la fois des actions des concepteurs du moteur, des webmasters et des internautes. Les ingénieurs de Google choisissent parmi les signaux émis d'en doter certains du rôle de critères algorithmiques. Assortis de pondérations, ces critères donnent lieu à un calcul de pertinence permettant de hiérarchiser les pages. Les webmasters et les internautes agissent quant à eux d'une façon qui, consciemment ou non, influencera le classement des documents. Si l'éditeur d'un site fait un jeu de mots dans le titre d'une page, celle-ci risque d'être moins bien référencée que s'il a choisi un titre informatif. S'il fait un lien vers une autre page, la pertinence supposée de cette dernière augmentera. L'internaute, quant à lui, formule une requête dont dépend le résultat final : « Sarko » ne donnera pas le même résultat que « Sarkozy ». L'énoncé produit par le moteur n'est donc imputable à aucun de ces acteurs en particulier, ni réductible à aucune de ces actions. Les acteurs, les actions, les algorithmes, l'interface, les logiciels de *crawling* et d'indexation, les serveurs, les câbles, les routeurs sont autant de *causes* dont la liste est un *résultat* situé dans l'espace (un pays, une région, une adresse IP, un profil utilisateur) et le temps (la même requête effectuée par le même internaute à deux moments différents est susceptible de générer deux listes différentes). Le dispositif déplace et interprète les cours d'action : il les *traduit* pour produire un énoncé.

Depuis la création du moteur de recherche en 1998, les dirigeants de Google sont progressivement passés d'une ambition descriptive — si un contenu est important alors il figure en tête du classement — à une ambition normative : si un contenu figure en tête du classement, c'est qu'il est important (Eisermann, 2009). En plus de produire un énoncé normatif, le moteur produit un espace de possibilité d'action : il ne se contente pas de hiérarchiser les références des documents, mais offre à ses utilisateurs la possibilité de se

rendre vers telle ou telle page dont le titre et la description lui auront semblé intéressants. Les actions rendues possibles sont hiérarchisées par Google qui en *recommande* certaines selon ce qu'il « juge » être un bon comportement à partir de ce qu'il « sait » de la pertinence des documents accessibles sur le web. On peut donc considérer qu'il existe, outre une ambition normative quant au calcul de pertinence de l'information, une ambition normative quant aux comportements de navigation.

Le moteur jouit d'une position dominante sur le web. Il est la première source de trafic pour de nombreux sites, qu'il inscrit à l'avant-scène d'un espace public conçu comme un continuum où certains propos sont « plus publics » que d'autres (Cardon, 2010). La survie économique de nombreux éditeurs dépend du trafic envoyé sur leurs pages par Google. En outre, il s'agit d'une question d'autorité et de légitimité dans l'espace public. Selon l'outil *Alexa*, le site « Wikipedia.org » recevrait environ 55% de son trafic depuis les moteurs de recherche, dont plus de 40% depuis Google. Le moteur permet donc à l'encyclopédie participative, et au modèle de savoir dont elle est le porte-étendard (Cardon, 2013), d'être légitime sur le web en contribuant à asseoir son autorité. Etant donné ce « pouvoir » qu'a Google, il y a lieu de s'interroger sur les dynamiques qui motivent les choix des concepteurs et influencent *in fine* la hiérarchisation des informations (l'énoncé) et la production de liens (les actions possibles). Les concepteurs ont leurs valeurs, leurs objectifs, leurs représentations, qui sont autant d'éléments à même d'influencer leurs choix. Par exemple, la décision d'intégrer aux critères algorithmiques la rapidité de chargement d'un site, faisant de la performance du contenant un critère de pertinence du contenu — décision marquée par une culture « orientée-ingénierie » (Auletta, 2010 : 19) — mérite d'être discutée. De telles interrogations nous semblent d'autant plus importantes dans le cas de la hiérarchisation opérée par Google vis-à-vis de l'information d'actualité, étant donné le rôle joué par celle-ci en démocratie.

Le traitement de l'information d'actualité opéré par le moteur est difficile à étudier empiriquement dès lors qu'on ne sait pas exactement comment le moteur fonctionne. Les algorithmes sont en effet considérés par les dirigeants de Google comme un secret industriel. Aucun utilisateur ni aucune autorité ne peut avoir accès aux détails de cette formule dont les auteurs prétendent « organiser les informations à l'échelle mondiale »¹. Quelques grandes lignes sont connues, et Google communique régulièrement pour annoncer que tel ou tel point est considéré comme un indicateur de pertinence, mais personne ne connaît l'algorithme dans le détail hormis une poignée d'ingénieurs basés à *Mountain View*. Qui plus est, l'algorithme est régulièrement modifié, sans que la teneur exacte des modifications ne soit annoncée. Certaines des modifications sont opérées manuellement, d'autres automatiquement : l'algorithme « réagit » aux comportements des internautes et se modifie ainsi *empiriquement*. Si les usagers cliquent systématiquement sur le deuxième lien de la liste, cette information sera prise en compte pour rectifier « l'erreur » d'appréciation. Ainsi, même ce qu'on croit savoir à propos de l'algorithme peut changer d'un instant à l'autre. D'autre part enfin, Google utilise des algorithmes de personnalisation, ce qui ajoute du brouillard au brouillard : deux internautes qui effectuent la même requête au même moment sont susceptibles d'obtenir des résultats différents (Pariser, 2011), sans qu'on puisse déterminer ni la teneur ni l'ampleur de la personnalisation, ce qui décourage toute tentative d'étude empirique.

¹ https://www.google.com/intl/fr_fr/about/company/

Que pouvons-nous identifier *avec certitude* au sujet du traitement de l'actualité opéré par Google ? Nous proposons de répondre à cette question en introduisant un outil conceptuel qui nous permettra de représenter le réseau hypertexte et de cartographier les forces qui s'exercent sur le fonctionnement du moteur. Dans une seconde partie, nous utiliserons cet outil pour dévoiler le traitement de l'actualité auquel Google a *intérêt* à procéder. Nous décrirons comment certains acteurs se positionnent étant donné les incitations dont ils savent que Google est l'objet. Enfin, grâce à une observation de vingt sites français, nous verrons quels éditeurs Google a effectivement intérêt à avantager dans ses résultats.

L'ARBRE DE NAVIGATION : CAPACITES, PROBABILITES, OPPORTUNITES

Qu'on soit éditeur de contenus, concepteur de service ou internaute, on « agit » sur le web : on met en ligne, on opère, on clique, on transfère, on télécharge. Le web peut ainsi être considéré comme un de ces « espaces de possibilité d'actions » dont Yoneyama nous dit qu'ils posent implicitement la question, non pas seulement de savoir « Quelle est la signification de... », mais aussi : « Quelle est la signification de la possibilité d'agir comme cela ? » (Yoneyama, 1997 : 114). Parmi les actions rendues possibles sur le web, toutes n'ont pas la même probabilité d'être réalisées. Pour chaque action, la probabilité de réalisation dépendra conjointement des intérêts et des compétences des acteurs concernés, ainsi que des caractéristiques de l'agencement qui prévaut à leur rencontre. Si on peut identifier avec précision la *capacité* d'action pour tel ou tel acteur, on ne peut en revanche qu'en estimer la *probabilité*, laquelle dépendra de l'*opportunité* représentée par cette action pour cet acteur, autant que de l'intérêt qu'auront d'autres acteurs à ce qu'il la réalise, ou au contraire à l'en empêcher, et de la capacité dont ces acteurs jouissent étant donné les instruments et les compétences dont ils disposent. S'ils jouissent d'une capacité forte, ils pourront contraindre directement le premier acteur. Sinon, ils devront le « conduire à » sans être en mesure de le forcer à quoi que ce soit. Un jeu de pouvoir, au sens foucauldien du terme (un pouvoir qui *vient de* et qui *est dans* les relations), se mettra inévitablement en place. Nous proposerons ici un outil conceptuel dont nous pensons qu'il peut aider à considérer les *capacités* et les *probabilités* d'action sur le web, ainsi que les *opportunités* que ces actions constituent, et à étudier ainsi des rapports de pouvoir parfois très subtiles qui ont des conséquences sur la communication et l'information.

Dès le départ, l'originalité du web a été d'offrir la possibilité de naviguer d'une page à l'autre par des liens hypertextes. Ces liens peuvent être produits manuellement par les éditeurs, ou générés automatiquement dans le cas par exemple d'un moteur de recherche, ou bien encore copiés, recopiés, échangés par les internautes. Tout en s'appuyant sur le réseau physique nommé Internet, le web est donc tissé de parcours sémantiques. Certains contenus occupent des positions centrales dans sa topographie, à la croisée des parcours, alors que d'autres se situent en périphérie et que d'autres enfin sont isolés, n'effectuant ni ne recevant

aucun lien, comme s'ils flottaient à côté du réseau². Ainsi, la capacité d'un internaute à naviguer n'est pas la même partout sur le web. N'étudier qu'une page serait omettre le répertoire dans lequel elle s'insère. C'est pourquoi il nous semble intéressant de penser le web comme un processus dynamique et topologique. Les liens donnent au web une « direction » : les documents, dès lors qu'ils sont en ligne, sont pris dans un mouvement qui « vient de » et « va vers ». Certains parcours sont plus attrayants que d'autres, certaines directions plus *engageantes*. Des opportunités se créent, dont dépendra *in fine* la probabilité qu'un internaute emprunte une voie plutôt qu'une autre. L'outil conceptuel que nous proposons ambitionne de rendre compte de ces dynamiques en déployant, pour chaque réseau de navigation, le réseau d'incitations afférent.

Nous calquons notre outil sur ce qu'en théorie des jeux les économistes appellent « un arbre de décision ». Il consiste à représenter les parcours de navigation depuis une page web, qui sert de racine à l'arbre, et de déterminer la probabilité afférente à chaque possibilité étant donné les opportunités qui se présenteront pour celui qui agit. Ainsi, on représente schématiquement l'ensemble des parcours qui s'ouvriront aux niveaux $n+1$, $n+2$, $n+3$, etc., pour un internaute dès lors qu'il se trouve sur une page « n ».

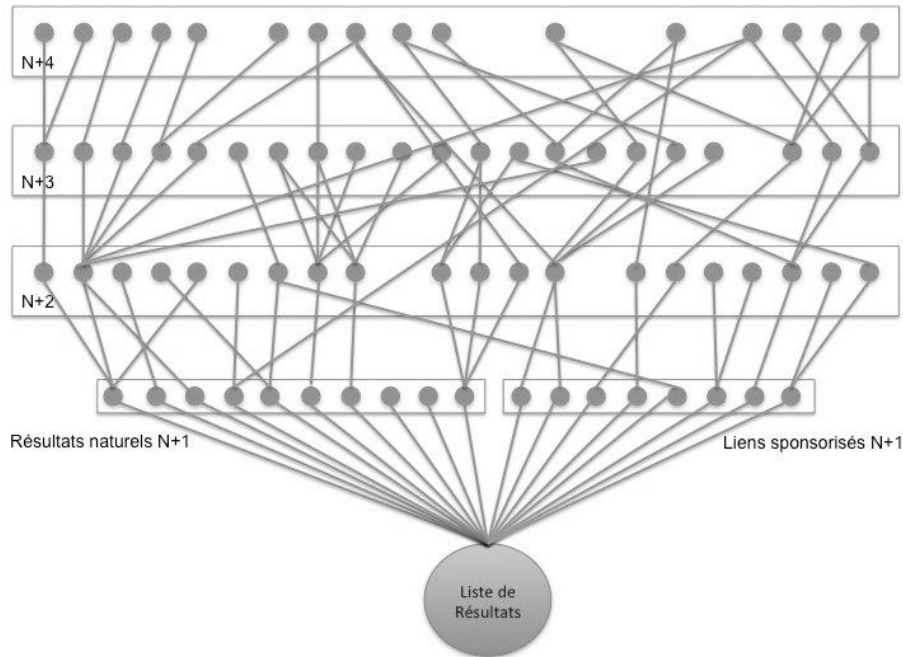
Considérons la liste produite par le moteur Google en réponse à la requête d'un internaute. Sur cette page figurent les liens dits « naturels »³, fruits du seul fonctionnement algorithmique, et, dans une colonne à droite ou dans un encart rosé au-dessus ou en-dessous des liens naturels, les liens dits « sponsorisés » sur lesquels un clic de la part de l'internaute générera un gain pour Google à qui l'auteur de la page vers laquelle le lien conduit reversera un montant calculé à partir des enchères effectuées par lui (Levy, 2009). Google prétend qu'il y a une division stricte entre l'activité qui prévaut à l'affichage des liens naturels et l'activité qui prévaut à l'affichage des liens sponsorisés. Cependant, nous n'avons pas d'autre choix que de croire sur parole Google, incapables de vérifier que le référencement sponsorisé n'influence pas le référencement naturel (ce qui est le cas, selon certains chercheurs qui se sont risqués aux spéculations (Yang et Ghose, 2010)).

Pour illustrer l'argument, représentons un arbre de navigation à quatre niveaux, auquel la liste de résultats générée par Google sert de racine (Figure 1). 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 parcours possibles pour ouvrir un nouvel ensemble de parcours possibles. Les parcours peuvent se rejoindre : certains documents du niveau « $n+3$ » effectuent des liens vers des documents du niveau « $n+1$ », accessibles par un chemin plus long que celui qui consistait à y accéder depuis la liste. Les ensembles générés par les liens naturels peuvent rejoindre les ensembles générés par les liens sponsorisés. Enfin, certaines pages ne pointent vers aucune autre page : les parcours navigationnels peuvent être limités ou se croiser.

² Le fameux *PageRank*, l'un des algorithmes utilisés par Google dans son ranking, peut d'ailleurs être interprété comme une mesure de *centralité* : la place d'un site dans le réseau des liens fonde donc son autorité.

³ Au sujet de ce terme, voir : (Farchy et Méadel, 2013)

Figure 1 : illustration à titre indicatif d'un arbre de navigation



Ce que nous avons représenté ici schématiquement constitue l'état du « jeu » lorsque la liste est à l'écran. Ainsi, la liste, qui est le résultat de la rencontre d'une multitude de cours d'action, est aussi la racine d'un arbre de navigation où chaque lien constitue un choix possible, afférent à une probabilité, conduisant à d'autres choix possibles, d'autres probabilités. Si on se place du point de vue de l'internaute qui utilise le moteur, on comprend qu'il ne s'agit plus seulement d'obtenir des informations pertinentes mais également des parcours satisfaisants. Chaque lien, de chaque niveau, a une probabilité d'être cliqué différente. Par exemple, sur la liste de résultats, selon une étude effectuée par Compete.com en 2012⁴, 85% des clics sont effectués sur les liens naturels et 15% sur les liens sponsorisés. Sur les liens naturels, selon cette même étude, 53% des clics sont effectués sur le premier résultat, 15% sur le second, 9% sur le troisième. Même chose pour les liens sponsorisés : les liens figurant en haut ont plus de probabilité d'être cliqués que les liens figurant en bas. Du point de vue de l'éditeur, la stratégie peut donc consister à chercher à apparaître le plus haut possible dans les résultats donnés par Google en optimisant son contenu (Sire, 2013), ou bien à chercher à être cité par un site référencé en bonne position sur la première page, 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 qui sont autant d'incitations susceptibles de peser sur le processus de communication.

⁴ <http://searchenginewatch.com/article/2215868/53-of-Organic-Search-Clicks-Go-to-First-Link-Study>

GOOGLE INCITEE A PRIVILEGIER CERTAINES PAGES DANS LES RESULTATS

Google, en plus de son moteur de recherche, a une activité publicitaire dont certains observateurs, dont l'ancien CEO Eric Schmidt, prétendent qu'il s'agit de son véritable cœur de métier (Vaidhyathan, 2011). Cette activité comprend les liens publicitaires affichés sur ses propres sites (*AdWords*), dont la firme Google est seule bénéficiaire des revenus. Ainsi, la firme, par rapport aux éditeurs qui financent leur production grâce à la publicité, est à la fois un partenaire qui leur apporte du trafic *via* son moteur de recherche et un concurrent puisqu'elle-même, comme les éditeurs, affiche des publicités sur ses pages pour financer un service qu'elle propose gratuitement. D'autre part, l'activité publicitaire de Google comprend également des annonces publicitaires distribuées sur les sites des éditeurs partenaires de son réseau de syndication, sous forme de liens ciblés (*AdSense*) et de bannières (*DoubleClick*). Ainsi, Google peut diffuser les publicités des annonceurs non pas seulement sur ses propres services, mais partout sur le web, dès lors que les éditeurs auront choisi d'en faire un partenaire économique. N'importe quel éditeur peut devenir un partenaire de *AdSense* et/ou de *DoubleClick*, et profiter du réseau d'annonceurs clients de Google et de différents services mis à sa disposition. Il partagera avec la firme les revenus générés par les publicités facturées au coût par clic ou au coût par affichage. Google se charge de la distribution technique des publicités, de la collection de données pour affiner le ciblage, et de la mise en disponibilité des interfaces que permettent aux annonceurs de placer leurs messages. Ainsi Google, en plus d'être un concurrent des éditeurs sur le marché de la publicité, leur propose d'être leur partenaire sur ce même marché. Cela rend extrêmement complexe la relation à la fois compétitive et coopérative, dite de « coo-pétition », qui caractérise la relation de Google avec les éditeurs, et en particulier les éditeurs de presse en ligne (Smyrnaio et Rebillard, 2009, 2011). Or, comme nous le verrons ci-après, le recours à l'arbre de navigation permet de saisir analytiquement certaines des dynamiques propres à cette relation, et d'en identifier les conséquences en termes de communication.

Nous avons récemment proposé une typologie des liens hypertextes générés par le moteur Google dans le but de décrire les incitations économiques attachées à chacun d'eux (Rieder et Sire, 2014). Cette typologie n'a pas pour but d'être exhaustive mais de révéler que les liens naturels peuvent être générateurs de profit et que, selon les parcours de navigation envisagés, les opportunités ne sont pas les mêmes pour Google. Nous pouvons ainsi déterminer les pages que Google a intérêt à avantager dans ses résultats organiques, soit parce que ce sont ses propres pages, soit parce que ce sont les pages d'éditeurs partenaires de *AdSense* ou *DoubleClick* ; et nous pouvons déterminer les pages que Google n'a pas intérêt à faire remonter dans les résultats, étant donné la probabilité de clic des premiers liens de la liste, parce que ce sont des services concurrents des siens, ou bien des clients d'autres réseaux publicitaires.

1. Liens propriétaires : pointent vers un site de Google où des publicités peuvent être affichées et des informations à propos du comportement de l'internaute collectées.
2. Liens concurrents : pointent vers des sites fournissant des services concurrents des services de Google.

3. Liens affiliés : pointent vers des sites partenaires du programme *AdWords*, directement depuis un des sites de Google.
4. Liens réseau : 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 : 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 : pointent vers des sites partenaires d'un service concurrent d'*AdSense* et/ou de *DoubleClick*.
7. Liens persistants : 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.

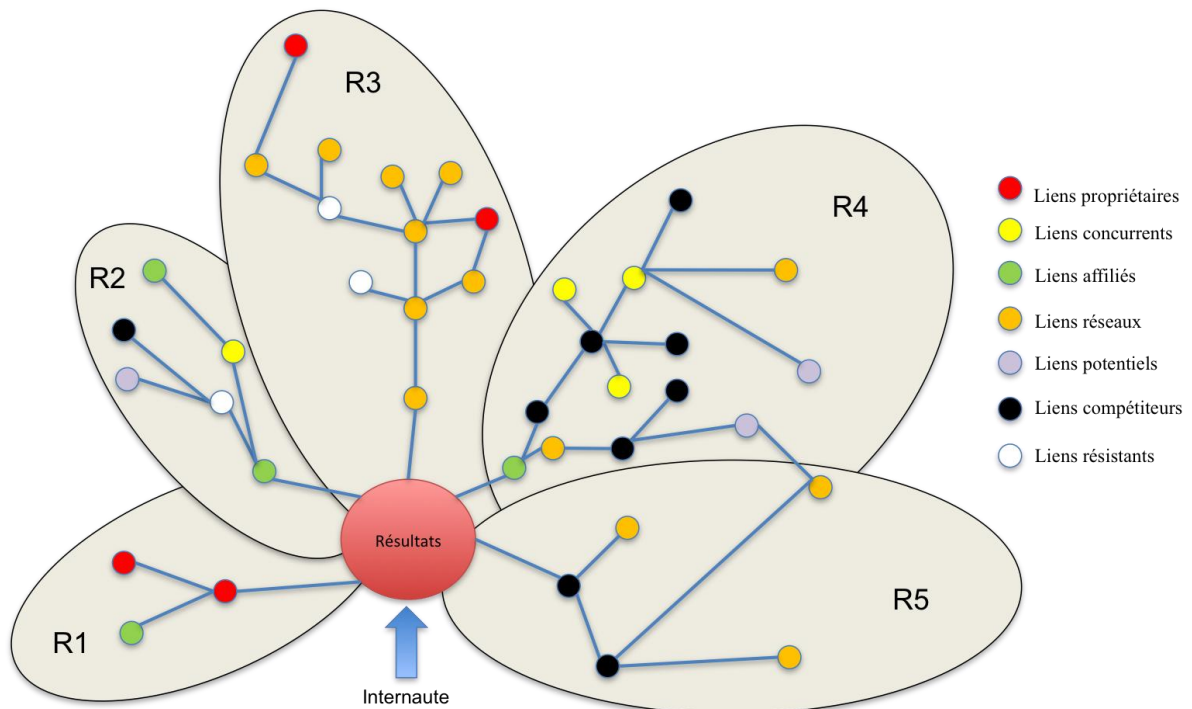
Notre typologie peut être complétée, discutée, et certains cas de figure peuvent correspondre à plusieurs des catégories décrites. Néanmoins, elle permet de dévoiler certains aspects systémiques du processus de communication. Cette typologie en effet, en plus de nous permettre d'identifier les pages en fonction des opportunités qu'elles constituent pour Google, permet d'analyser les parcours en fonction des incitations qu'ils représentent.

Etant donné sa double casquette, Google pourrait être tentée de privilégier les sites partenaires de ses réseaux *AdSense* et *DoubleClick* dans les résultats naturels parce que ces sites représentent une opportunité de gain directe. Par ailleurs, le concept d'arbre de navigation permet 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 et vers lesquelles l'internaute se dirigera *probablement*. Une page qui n'est partenaire ni d'*AdSense* ni de *DoubleClick* peut elle-même être intéressante à faire remonter dans les résultats dans la mesure où elle renvoie vers de nombreuses pages rémunératrices pour Google et dans la mesure où il existe une probabilité satisfaisante que l'internaute accède effectivement à l'une de ces pages.

Considérons un arbre de navigation schématique et simplifié ci-dessous (Figure 2). Chaque noeud 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 pour Google. Nous voyons sur cette représentation simplifiée (il y manque la probabilité de clic afférente à chaque lien hypertexte) à quel point les incitations se trouvant derrière la liste peuvent être complexes. Par exemple, le réseau R4, auquel l'internaute accède par un lien *AdWords* rémunérateur pour Google, est truffé de liens pointant vers des sites partenaires d'un réseau concurrent. Dès lors, R4 peut potentiellement générer moins de profits que le réseau R3, jonché 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 pour Google. Nous comprenons désormais qu'un réseau de navigation, du point de vue de Google, est aussi un réseau d'opportunités et donc, inévitablement, constitue un ensemble d'incitations qui justifieraient de biaiser les résultats du moteur. Nous ne prétendons pas que ces incitations influencent les résultats, mais plutôt qu'elles ne peuvent pas en être exclues *a priori*. Les dirigeants de Google peuvent choisir d'aller dans le sens de ces incitations — rien ne le leur interdit. Ils pourraient par exemple paramétrer leur algorithme de façon à ce que si

deux pages reçoivent un score de pertinence équivalent, celle des deux où se trouve une publicité de Google soit classée plus haut que l'autre. Ils pourraient même paramétrer le moteur de façon à ce qu'une page où se trouve une publicité de Google soit avantagée même si son score de pertinence est légèrement inférieur à une page où ne figurerait pas de publicité de Google. Cette observation mérite d'être considérée à l'aune d'un des axes de recherche principaux de l'économie politique des médias, qui consiste à interroger le danger d'un financement basé sur la publicité à l'égard de l'intégrité éditoriale qui prévaut au traitement de l'information (McChesney, 2008). Dans le cas du moteur, les implications éditoriales des choix opérés *a priori* par les ingénieurs qui paramètrent l'algorithme sont minimisées dans le discours des porte-parole de Google qui prétendent que le fonctionnement du moteur est « organique » ou « naturel ». Les ingénieurs, qui plus est, n'ont sans doute pas les mêmes états d'âme que les journalistes dès lors qu'il s'agit de ne pas être influencé par l'activité publicitaire. Finalement, le débat se heurte à un dispositif composé indifféremment de technique (dont on ne sait pas exactement comment elle opère) et de discours (dont on ne peut pas être sûr qu'il dit la vérité). Enfin, il est important de noter que si les dirigeants de Google décident de ne pas établir ce type de biais informationnel, comme c'est apparemment le cas, les incitations ne disparaissent pas pour autant, et ils pourront à tout moment revenir sur leur décision. C'est pourquoi ces incitations, même invisibles, font partie intégrante du processus de communication sur lequel nous disons qu'elles « pèsent », d'autant que, comme nous le verrons dans la partie suivante, certains acteurs agissent en fonction des incitations auxquelles ils savent *avec certitude* que Google est sujet.

Figure 2



L'arbre de navigation permet de déconstruire la séparation entre résultats naturels et liens sponsorisés. Pour qu'il soit complet, nous devrions y ajouter la probabilité de clic afférente à chaque lien hypertexte, dans le but de comparer les potentialités de profits propres à chaque parcours. Nous voyons apparaître ici le tissu d'incitations auquel les fondateurs de Google ont eux-mêmes fait référence lorsqu'ils ont expliqué dans l'article donnant naissance au moteur 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 logiciels de *crawling* enregistrent la structure hypertexte, et donc que le dispositif a accès aux informations permettant de tracer les ramifications de l'arbre de navigation, d'évaluer les probabilités de clic et de calculer les opportunités de profit attachées aux ensembles navigationnels.

Nous voudrions insister sur le fait que nous ne prétendons pas que l'arbre de navigation est considéré 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.

LA PRESSE EN LIGNE INCITÉE A DEVENIR PARTENAIRE DE GOOGLE

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 certains ont à l'esprit que cela peut être possible et qu'étant donné les incitations entremêlées, comme le disaient Sergey Brin et Larry Page eux-mêmes en 1998⁵, 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. Dès lors, les éditeurs préfèrent se positionner du côté des éditeurs que Google *aurait intérêt* à avantager dans le cas d'un biais visant à privilégier ses partenaires.

« 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 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, 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 ! »

Directeur de l'acquisition d'audience

⁵ « 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).

« [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. Moi je mets du AdSense sur mes pages, parce que ça rapporte, et parce que c'est pas plus mal : en quelque sorte, ça permet de se faire bien voir par Google. »

Référenceur

Nous voyons dans ces citations, extraites d'entretiens semi-directifs que nous avons menés à l'automne 2012 auprès d'employés d'entreprises de presse en ligne françaises, comment certaines actions peuvent être effectuées par les éditeurs non pas étant donné le fonctionnement du moteur, mais étant donné la multiplicité des métiers de Google et le réseau d'incitations qui entoure le fonctionnement du moteur. Dans la situation de coo-pétition qui caractérise leurs rapports avec Google, les éditeurs de presse peuvent donc estimer qu'il est dans leur intérêt d'adhérer à *AdSense* et/ou à *DoubleClick*, et donc de partager avec Google une partie des revenus générés par leur activité sur le marché de la publicité, et ce en dépit du fait que Google soit leur concurrent sur ce même marché. Par rapport à d'autres régies et d'autres réseaux de syndication, proposant des services comparables à *AdSense* ou à *DoubleClick*, on peut donc estimer que Google jouit d'un avantage lié à la fonction de prescription de son moteur et à l'incertitude technique qui en nimbe le fonctionnement.

OBSERVATION EN LIGNE

Nous rendons compte ci-après d'une observation en ligne effectuée en juillet 2014 auprès des sites de presse français. Cette observation vise à savoir quels éditeurs Google a effectivement intérêt à privilégier dans ses résultats. Nous avons limité notre observation aux 20 premiers sites de la catégorie « Actualités / Information Généraliste » du classement de l'OJD⁶, hiérarchisés en fonction du nombre moyen de visiteurs quotidiens en juin 2014.

Pour chacun des sites, nous avons observé la présence de modules *AdSense* ou *DoubleClick*. Nous avons ainsi pu identifier les éditeurs partenaires économiques de Google. Il se trouve que durant notre observation, certains sites effectuaient également des publicités pour les produits de Google, ce qui renforçait les incitations de la firme à faire en sorte que son moteur pointe vers ces sites. Ainsi, la firme peut être à la fois un concurrent, un partenaire et un client des éditeurs de presse sur le marché de la publicité. Lorsqu'un éditeur est partenaire de Google, nous avons renseigné la valeur « 1 » dans la case « Incitations » de notre tableau, et la valeur « 2 » lorsque Google est à la fois un partenaire et un client sur le marché de la publicité. Le chiffre « 0 » indique que Google n'est rien d'autre qu'un concurrent.

Nous avons utilisé l'outil *Alexa*⁷ pour nous faire une idée du pourcentage de visiteurs arrivés sur le site de presse par l'intermédiaire de Google. Nous avons additionné les chiffres donnés pour google.com et google.fr. Pour le trafic en provenance du moteur vertical Google

⁶ <http://www.ojd.com/Chiffres/Le-Numerique/Sites-Web/Sites-Web-GP>

⁷ <http://www.alexa.com>

Actualités⁸, nous avons repris les données générées par l’outil *Databees*⁹. Ce dernier donne la « part de voix », définie comme le rapport du nombre de liens du media affichés sur Google Actualités *versus* le nombre de liens affichés pour l'ensemble des medias.

Tableau 1

Nom du site	Moyenne de visiteur quotidien (chiffres OJD, juin 2014)	Partenaire Google	Publicité Google	Incitations pour Google	Part de voix totale sur Google News Databees (24 h, 23/07/2014)	Pourcentage de visiteurs en provenance de Google (.fr et .com, Alexa, 23/07/2014)
LeMonde.fr	1 610 456	0	0	0	7,42	22,50%
LeFigaro.fr	1 523 533	0	0	0	4,61	24,50%
NouvelObs.com	856 322	1	0	1	3,62	28,50%
LeParisien.fr	809 055	1	1	2	4,46	32%
Lexpress.fr	611 309	1	0	1	1,74	31,30%
20minutes.fr	575 317	1	0	1	1,6	25,30%
Bfmtv.com	551 907	1	1	2	3,36	24%
LePoint.fr	468 030	1	0	1	4,48	31,40%
Liberation.fr	422 968	0	0	0	3,56	23,10%
Ouestfrance.fr	417 703	1	1	2	1,83	33,40%
Francetvinfo.fr	387 511	1	1	2	1,19	24%
Huffingtonpost.fr	356 623	1	0	1	0,01	19%
Europe1.fr	347 036	1	0	1	5,3	36,30%
Sudouest.fr	324 666	1	0	1	0,02	29,10%
Ladepeche.fr	322 251	1	0	1	0,05	31,70%
Lavoixdunord.fr	258 006	1	0	1	NC	35,40%
Planet.fr	229 895	1	0	1	0,02	26%
Ledauphine.com	212 158	1	0	1	NC	25,50%
Rfi.fr	188 683	1	0	1	0,7	13,10%
Midilibre.fr	174 475	1	0	1	0,08	23,60%

⁸ Le moteur vertical Google Actualités fonctionne sur un champ d’indexation restreint aux seuls sites identifiés au préalable par Google comme étant des sites d’information.

⁹ <http://media.databees.net>

Alors que l'ensemble des 20 sites mentionnés sont des concurrents de Google sur le marché de la publicité, 17 sites sont également des partenaires économiques de Google sur ce même marché. 4 sites affichent des publicités pour les produits de Google, c'est-à-dire qu'en plus d'être leur concurrent et leur partenaire, Google est également leur client, ce qui renforce l'intérêt pour la firme de voir remonter ces sites dans les classements de son moteur.

Nous ne prétendons pas faire autre chose ici que de dévoiler les sites des éditeurs que Google a intérêt à avantager dans ses résultats sans prétendre que Google les avantage *effectivement*. L'exemple du site Rfi.fr montre d'ailleurs qu'un site peut être un partenaire économique de Google et recevoir assez peu de trafic en provenance de ses services.

Les sites de presse les plus visités — hormis lemonde.fr, lefigaro.fr et liberation.fr — sont des partenaires économiques de Google. Autrement dit, ils partagent une partie de leurs revenus avec la firme qui est aussi leur concurrent. S'il y avait eu une forte préférence pour un site partenaire de la part de Google, notre observation l'aurait sans doute révélé. C'est notre deuxième conclusion ici : rien ne permet de dire que la firme Google avantage dans son moteur les sites qu'elle a intérêt à avantager.

Notre observation, cependant, est incomplète. Nous ne nous situons ici qu'au niveau « n+1 » de l'arbre de navigation, et nous ne considérons pas les probabilités afférentes aux possibilités de navigation. L'enquête mériterait d'être poursuivie pour représenter l'arbre de navigation dans son ensemble.

CONCLUSION

Nous nous sommes intéressés dans cet article aux incitations au biais dont le traitement de la production journalistique par le moteur de recherche Google peut faire l'objet. Nous avons montré qu'étant donné les différents métiers de la firme, il peut exister certaines distorsions concernant le marché de la publicité d'une part, et, d'autre part, le traitement de l'information. Nous avons dévoilé pourquoi le référencement « naturel » sous-tend des opportunités de profit pour la firme Google, cela parce qu'elle agit, de manière dominante, à la fois sur le marché des moteurs de recherche et sur le marché de la publicité, et sous-tend donc nécessairement certaines incitations au biais informationnel. Nous avons introduit ces réflexions grâce à un outil conceptuel que nous avons nommé « arbre de navigation », qui permet d'étudier les incitations propres aux différents parcours de navigation hypertexte s'ouvrant depuis la page de résultats générée par un moteur.

Nous ne prétendons pas que les incitations à privilégier certains sites influencent le résultat de l'action du moteur Google, mais plutôt qu'elles font partie de ce résultat au même titre que les liens hypertextes et qu'elles ne peuvent par conséquent en être exclues *a priori* dès lors que nous n'aurions aucun autre choix pour cela que de croire en la bonne foi des porte-parole de Google. En publiant une liste de liens, le moteur ouvre des possibilités de navigation dont chacune a une probabilité d'être actualisée, et crée des opportunités, le résultat de son action n'étant réductible ni à la seule liste, ni aux seuls parcours, ni aux seules opportunités, ni au

discours des porte-parole quant à la neutralité qui a prévalu au tracé des parcours, mais à ces éléments *conjointement*.

Nous avons expliqué pourquoi un éditeur de presse peut être incité à partager ses revenus avec Google dans la mesure où le trafic apporté par le moteur est indispensable à sa survie, où le fonctionnement du moteur est opaque et où les intérêts de Google, eux, sont appréhendés. Le rôle joué par la firme doit selon nous être questionné à lumière du fait que les éditeurs, étant donné le besoin qu'ils ont de recevoir du trafic, ne peuvent pas ignorer ce qu'ils savent *avec certitude* au sujet de Google (des incitations poussent la firme à avantager ses partenaires) et ce qu'ils ignorent (les avantage-t-elle effectivement ?).

Nous avons montré comment une des questions classiques de l'économie politique des médias, concernant les pressions exercées par la publicité sur les logiques éditoriales, peut être renouvelée étant donné le fonctionnement automatisé des outils de traitement, et les relations économiques et techniques complexes qui se créent entre les éditeurs, dont le périmètre est localisé, et ces « infomédiaires » (Rebillard et Smyrnaio, 2010) œuvrant à l'échelle de la planète et pratiquant des stratégies intensives de diversification. Dans l'optique d'un tel renouvellement, nous pensons que le recours à l'arbre de navigation constitue une méthode heuristique permettant de comprendre les forces qui travaillent les choix des acteurs sur le web, et les effets communicationnels que ces forces sont susceptibles de générer. Pour être complet, cet arbre doit comprendre 1) *les capacités* de navigation hypertexte ; 2) *les probabilités* que ces capacités donnent effectivement lieu à des actions ; 3) *les opportunités* de profit constituées pour chaque acteur par chacune des capacités.

BIBLIOGRAPHIE

- Auletta K., 2010, *Googled. The End of the World as We Know It*. London : Penguin Books.
- Brin S. et Page L., 1998, « *The Anatomy of a Large-Scale Hypertextual Web Search Engine* », Computer Science Department, Stanford University.
- Cardon D., 2010, *La démocratie Internet. Promesses et limites*, Paris : Seuil.
- Cardon D., 2013, « Comment s'auto-organise la production des savoirs sur Wikipédia ? », In, Larqué L. et Pestre D. (dirs.), *Les sciences, ça nous regarde*, Paris, Les empêcheurs de penser en rond/La Découverte, p. 266-272.
- Dellarocas C., Katona Z. et 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.
- Eisermann M., 2009, *Comment fonctionne Google ?*, Université Joseph Fourier, Grenoble, <http://www.igt.uni-stuttgart.de/eiserm/enseignement/google.pdf>
- Farchy J. et Méadel C., 2013, « Moteurs de recherche et référencement : chassez le naturel... », *Hermès*, n°66.
- Freyssenet M., 1990, *Les formes sociales d'automatisation, Cahiers du GIP Mutations industrielles*, Paris, 1990, 47 p.
- Levy S., 2009, « Secret of Googlenomics: Data-Fueled Recipe Brews Profitability », *Wired* 16(7).
- McChesney R.W., 2008, *The Political Economy of Media: Enduring Issues, Emerging Dilemmas*, New York: Monthly Review Press.
- 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.
- Rebillard F., Smyrnaiois N., 2010, « Les infomédiaires, au cœur de la filière de l'information en ligne », *Réseaux*, vol.2, n°160-161, p. 163-194.
- Rieder B., Sire G., 2014, « Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google's Tangled Position on the Web », *New Media & Society*, vol. 16, n°2, p. 195-211.
- Smyrnaiois N. et Rebillard F., 2009, « L'actualité selon Google », *Communication & langages*, n° 160, p. 95-109.
- Smyrnaiois N. et Rebillard F., 2011, « Entre coopération et concurrence : Les relations entre infomédiaires et éditeurs de contenus d'actualité », *Concurrences*, n°3-2011, p. 7-18.

Sire G., 2013, « Google et les éditeurs de presse en ligne, une configuration négociée et négociable », *Sur le journalisme*, vol.2, n°1.

Vaidhyanathan S., 2011, *The Googlization of everything: (and why we should worry)*, Berkeley / Los Angeles, University of California Press, 265 p.

Yang S. et Ghose A., 2010, « Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? », *Marketing Science*, 29(4), p. 602-623.

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.