

SESSION 2016

**CAPES
CONCOURS INTERNE
ET CAER**

SECTION : DOCUMENTATION

ÉPREUVE À PARTIR D'UN DOSSIER THÉMATIQUE

Durée : 5 heures

L'usage de tout ouvrage de référence, de tout dictionnaire et de tout matériel électronique (y compris la calculatrice) est rigoureusement interdit.

Dans le cas où un(e) candidat(e) repère ce qui lui semble être une erreur d'énoncé, il (elle) le signale très lisiblement sur sa copie, propose la correction et poursuit l'épreuve en conséquence.

De même, si cela vous conduit à formuler une ou plusieurs hypothèses, il vous est demandé de la (ou les) mentionner explicitement.

NB : La copie que vous rendrez ne devra, conformément au principe d'anonymat, comporter aucun signe distinctif, tel que nom, signature, origine, etc. Si le travail qui vous est demandé comporte notamment la rédaction d'un projet ou d'une note, vous devrez impérativement vous abstenir de signer ou de l'identifier.

Tournez la page S.V.P.

A

Titre du dossier : Rechercher de l'information

A partir de ce dossier thématique comprenant 5 documents numérotés de 1 à 5, vous devez :

1. rédiger une note de synthèse dégageant une problématique d'ensemble (3 pages maximum) ;
2. développer une réflexion personnelle sur le rôle du professeur-documentaliste dans la formation à la recherche d'information ;
3. élaborer pour le document 3 la référence bibliographique suivie des éléments d'analyse.

La référence bibliographique de ce document sera établie, en application des normes en vigueur, en renseignant les zones pertinentes du bordereau de saisie.

Mots clés : indiquer exclusivement dans le champ du bordereau prévu à cet effet les mots clés qui permettront une recherche efficiente en langage naturel.

Le document 3 fera l'objet d'un résumé indicatif de 50 mots.

Règle de comptage des mots :

Les chiffres : une date = 1 mot (ex. 2007 = 1 mot) ; un pourcentage : 50% = 2 mots

Les signes : CNDP = 1 mot. Il est déconseillé d'utiliser des sigles peu connus dans l'éducation nationale

Les articles, même élidés = 1 mot

Les mots composés avec un trait d'union (ex. sino-soviétique = 1 mot), mais c'est-à-dire = 4 mots

Document 1

Clément Arsenault. La recherche d'information.

Document 2

Conseil national du numérique. Des méthodes, des approches, des compétences autant que des savoirs.

Document 3

Dominique Cardon. L'ordre des informations entre algorithmes et subjectivités.

Document 4

Rémi Sussan. A propos de la sérendipité.

Document 5

Nicole Boubée. Le rôle des copiés-collés dans l'activité de recherche d'information des élèves du secondaire.

Salaün, J.-M., Arsenault, C. *Introduction aux sciences de l'information*. Paris : La découverte, p. 101-105. (Grands repères)

3

La recherche d'information

Clément Arsenault, Lyne Da Sylva,
Dominic Forest, Yves Marcoux et Dominique Maurel*

La recherche d'information peut se définir sommairement comme un ensemble d'opérations effectuées dans le but de satisfaire un besoin d'information. Il s'agit plus précisément, dans la majorité des cas, d'un processus interactif entre, d'une part, un système informatisé, composé principalement d'une interface et d'une base de données, et, d'autre part, un utilisateur. Dans le cadre de ce processus, l'utilisateur tente de trouver, à l'aide de l'interface du système, des informations pertinentes contenues dans la base de données qui lui permettront de satisfaire, en totalité ou en partie, son besoin informationnel. Grâce à la prolifération des sources d'information, à la simplification des systèmes (en particulier des interfaces), à la mise en place de réseaux de télécommunication, et au développement du web, la recherche d'information est une activité humaine qui, depuis quelques années, a pris une importance capitale dans les sociétés modernes, notamment sur les plans économique, social et politique.

Les opérations découlant du processus de recherche d'information se trouvent au cœur même de la chaîne documentaire: en aval du traitement mais en amont de la diffusion. Avec la recherche d'information, on entre de plain-pied dans la sphère que l'on qualifie traditionnellement en bibliothèque

* Responsable du chapitre.

de «services publics». En effet, les opérations de recherche nécessitent soit l'intervention d'un intermédiaire expert (recherche déléguée à un bibliothécaire, spécialiste de l'information) qui effectuera les recherches pour un demandeur d'information, soit la conception de systèmes dans lesquels les demandeurs d'information pourront effectuer directement la recherche de façon autonome ou semi-autonome.

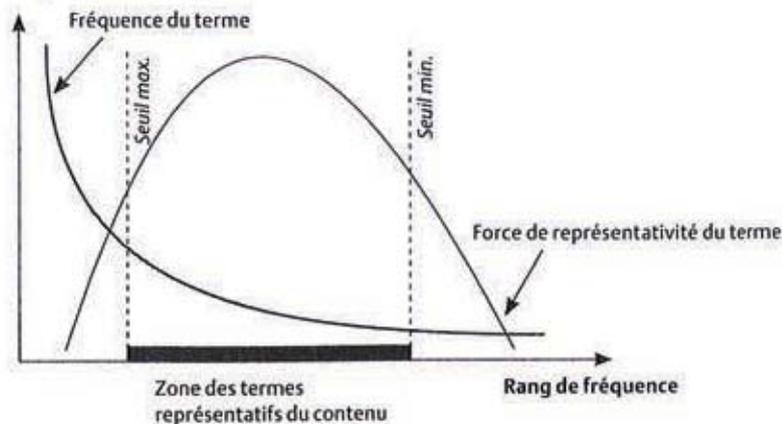
1. Les fondements, les concepts et les composantes

1.1 Les fondements historiques

Les systèmes informatisés de recherche d'information se sont d'abord développés grâce aux percées effectuées en recherche dans le domaine du repérage d'information au cours des années 1950 et 1960. L'abondante documentation accumulée durant la Deuxième Guerre mondiale, en particulier aux États-Unis, posait des problèmes de repérage sans précédent, et les systèmes manuels ou mécanisés mis en place jusqu'alors ne suffisaient plus à répondre à la demande. L'avènement des ordinateurs ouvrait de nouvelles avenues à explorer, non seulement pour le stockage d'information mais aussi pour la manipulation de texte et la création automatique d'index (cf. chap. 2, sect. 3.4). Parmi les premières recherches souvent citées dans le domaine, mentionnons celles de Hans Peter Luhn (chez IBM) qui démontra de façon concluante qu'il était notamment possible, avec de simples analyses de fréquence effectuées par exemple sur un texte de résumé, d'en déterminer et d'en extraire les termes significatifs (mots-clés). Pour identifier les termes représentatifs du contenu d'un document, l'idée générale avancée par Luhn était d'éliminer les termes peu fréquents, ainsi que les termes dont la fréquence d'apparition est très élevée comme les articles et les prépositions (Luhn 1958, 160-161). Nous reprenons son graphique de détermination de la force de représentativité des termes à la figure 3.1. Les travaux de Bar Hillel (1959) vinrent ajouter aux travaux de Luhn en suggérant que pour une sélection plus adéquate des mots-clés, on doit également tenir compte de la fréquence d'apparition des termes dans la langue dans son ensemble, car cet élément agit comme valeur de discrimination. Cette proposition est à la base de ce que l'on appelle la «fréquence inverse des documents» (*inverse document frequency — idf*), méthode formellement élaborée quelques années plus tard par Spärck Jones (1972) et qui est une des techniques les plus utilisées encore de nos jours pour identifier automatiquement les termes importants (i. e. porteurs de sens) dans un document.

FIGURE 3.1

Représentativité des termes



Dans les années 1960, l'élaboration d'importantes bases de données bibliographiques — notamment dans les domaines médical, juridique et journalistique —, jumelée à l'élaboration de systèmes interactifs de repérage d'information, a permis de mettre en place les premiers systèmes commerciaux de recherche d'information. C'est ainsi que naissaient, en 1972, Dialog et, en 1973, LexisNexis, deux des plus gros systèmes commerciaux de recherche d'information encore actifs de nos jours¹. Le rapide succès commercial de ces applications a favorisé l'émergence d'une véritable industrie de la recherche en ligne (Bourne et Hahn 2003, 279-352). Parallèlement, la recherche sur le repérage d'information s'est intensifiée avec notamment la mise en place de systèmes de bancs d'essai pour mesurer de façon empirique la performance des systèmes de repérage d'information.

Pendant ce temps, sur le terrain, on assiste au début des années 1980 à une véritable explosion du nombre de bases de données bibliographiques, factuelles et numériques, ainsi qu'à l'émergence de plusieurs autres systèmes commerciaux fournissant des services de recherche d'information en ligne (Gale 2008). La forte compétition résultant de cette rapide croissance a permis, compte tenu de la baisse des coûts d'utilisation, d'étendre significativement le bassin d'utilisateurs. De plus, vers le milieu des années 1980, divers facteurs ont contribué à la démocratisation de la recherche d'information. Les efforts investis dans la simplification des interfaces et des langages d'interrogation, l'apparition de nouveaux supports de diffusion tels que le CD-ROM

1. www.dialog.com/ et www.lexisnexis.com/.

et l'émergence des catalogues de bibliothèque informatisés ont fait en sorte que le demandeur d'information était maintenant en mesure, à la suite d'une formation initiale de base, d'effectuer lui-même ses recherches sans nécessairement avoir besoin de demander l'aide d'un intermédiaire expert.

Vers la fin des années 1980, la recherche se penche plus sérieusement sur un facteur jusqu'alors quelque peu délaissé par la communauté scientifique : l'utilisateur (Ellis 1992). Sont alors étudiés les besoins des usagers, leurs compétences et comportements informationnels, ainsi que les processus cognitifs impliqués dans la recherche d'information (mentionnons notamment les travaux de Belkin, Ingwersen, Kuhlthau, Spink et Wilson). Cet intérêt accru pour les utilisateurs de systèmes d'information correspond à la période de grande croissance dans le domaine de la recherche d'information stimulée par le développement d'Internet et plus particulièrement l'émergence du web. Dès la fin des années 1990, la grande majorité des systèmes commerciaux mettent au point des interfaces grand public qu'ils rendent accessibles via le web (moyennant, dans la majorité des cas, des frais d'accès ou d'abonnement). Parallèlement, les contenus publics disponibles sur le réseau Internet sont également reconnus comme d'importantes sources d'information. Dans la foulée sont créés des outils de recherche spécifiquement dédiés au repérage de l'information contenue sur le web. Certains de ces outils sont devenus les gigantesques moteurs de recherche que nous connaissons (Google, Yahoo!, Live Search [MSN], Ask) et qui indexent de substantielles portions du web en texte intégral. Étant donné la vaste quantité d'information à traiter ainsi que le manque flagrant de structure dans cet océan documentaire, ces systèmes misent grandement sur le développement d'algorithmes sophistiqués pour déterminer automatiquement la pertinence des résultats de recherche afin de les classer. Avec l'apparition de ce genre d'outils, de nombreuses techniques expérimentales sont pour la première fois mises en application.

L'accès direct à l'intégralité du texte en format numérique est vite devenu une réalité bien intégrée par les usagers et, de ce fait, un élément exigé qui ne peut maintenant plus être négligé dans les systèmes d'information. Au début des années 2000, les avancées technologiques ont permis de développer et de mettre en place des protocoles pour relier les données bibliographiques, contenues dans les bases de données et les catalogues, aux documents (textuels ou autres) décrits dans ces outils.

De nos jours, l'intégration de plus en plus poussée du web dans la vie quotidienne incite de nombreux utilisateurs à participer activement et

collectivement à l'élaboration et au développement d'importantes ressources d'information. Les blogues, les wikis et les sites de partage d'information qui sont apparus il y a quelques années constituent déjà une part importante de l'information accessible sur le web. Ce web « communautaire », qui s'inscrit dans la mouvance de ce que l'on nomme Web 2.0, contribue positivement au monde de la recherche d'information. Le rôle de l'utilisateur se transforme, car celui-ci contribue collectivement à l'élaboration des bases de connaissances en proposant des sources alternatives et de nouvelles avenues pour accomplir le processus de communication qu'est fondamentalement la recherche d'information, c'est-à-dire la mise en relation d'un demandeur et d'un fournisseur d'information (Meadow et al. 2007, 3).

En 1992, Robertson et Hancock-Beaulieu (1992, 458-459) identifiaient trois révolutions dans le milieu de la recherche d'information : la révolution de pertinence, la révolution cognitive et la révolution interactive (homme-machine). Avec le développement des outils de recherche collaboratifs, nous sommes fort probablement à l'aube d'une nouvelle révolution que l'on pourrait qualifier de révolution participative.

[...]

Des méthodes, des approches, des compétences autant que des savoirs

Faire partager aux apprenants une littératie numérique, c'est autant mobiliser des méthodes de travail, des attitudes, que l'acquisition de savoirs et de compétences. Nombre de ces méthodes et approches préexistent à la société numérique (ex : l'apprentissage de la recherche d'information, la construction d'un argumentaire, la méthode par projet ...) mais trouvent avec le numérique un support pour un usage renouvelé ou amplifié. D'autres sont inspirées du fonctionnement même du monde numérique (ex : les communautés coopératives du logiciel libre, les hackatons, barcamp, meetups...). Une des ruptures liées au numérique tient à l'abondance de contenus, à la fois réutilisables, appropriables, et accessibles par tout un chacun. Cette transformation crée à la fois une opportunité mais aussi le besoin d'apprentissages pour la réutilisation, la création de contenus, les modes de diffusion et la gestion des droits et des licences d'usage.

L'intérêt de ces postures et méthodes est triple. Elles permettent à l'apprenant d'acquérir les clés d'une compréhension du monde numérique qui pourront être convoquées quelles que soient les technologies auxquelles il ou elle sera confronté(e) tout au long de sa vie. Elles lui donnent les moyens d'une indépendance devenue indispensable dans un monde professionnel où il est de fait demandé à l'individu d'être toujours plus autonome. Des clés qui sont également indispensables au citoyen de demain, dans son activité sociale comme dans la gestion de sa vie privée.

Une nouvelle culture de l'apprentissage

Passer d'une culture de la compétition à une culture de la coopération

Le principe de compétition a structuré le développement économique des pays occidentaux au cours du vingtième siècle. Dans une vision libérale, c'est le moteur de la créativité individuelle et collective : compétition des entreprises bien sûr, mais aussi des villes et des territoires, des individus... Le système scolaire a participé de cette culture de la compétition, notamment par ses systèmes de notation des élèves et d'évaluation des enseignants. C'est aussi cette culture de la compétition qui a organisé le travail de l'enfant sur un modèle essentiellement individuel et « vertical ».

2 – Installer la littératie de l'âge du numérique

Nous nous heurtons maintenant brutalement aux limites de cette idéologie : économies en panne ou anémiées, inégalités sociales croissantes, ressources naturelles en voie d'épuisement... Et des logiques coopératives peuvent ouvrir des alternatives. Il ne s'agit pas ici de substituer un dogme à un autre – la coopération à la compétition – mais d'explorer en quoi des logiques coopératives peuvent nous amener vers d'autres modèles de développement, plus durables, et susceptibles d'accroître le bien-être des individus.

Le numérique, par son organisation intrinsèquement réticulaire, peut faciliter les logiques coopératives. On l'observe tous les jours, dans la manière dont travaillent les entreprises du numérique, dans les projets d'innovation ouverte inter-entreprises, dans l'émergence de services basés sur des échanges en pair à pair, dans la construction de communs de la connaissance, fruits d'un travail coopératif (des logiciels libres aux encyclopédies).

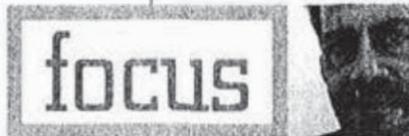
L'école, en introduisant dans son enseignement une culture numérique, peut également être vecteur de ce savoir-faire coopératif et y trouver des voies pour renouveler des pratiques d'évaluation qui génèrent des attitudes défensives et individualistes.

Sortir des logiques d'accumulation, faire l'apprentissage de la sérendipité, valoriser la curiosité

Le web, en ouvrant l'accès quasi instantané à des ressources documentaires en masse, tend à externaliser notre fonction mémorielle vers des ressources numérisées. Certains psychologues cognitivistes s'en alarment, et communiquent leur inquiétude aux enseignants, sans proposer nécessairement d'alternatives en phase avec des usages déjà bien installés. Une approche défensive ne nous paraît pas à la hauteur du problème. Il nous semble plus fécond de :

- acter que chaque grande révolution technologique interagissant avec le champ de la connaissance, à commencer par l'imprimerie, a indéniablement véhiculé une transformation cognitive, sans que cela ne soit nécessairement une régression ;
- abandonner plus radicalement les logiques d'accumulation et déplacer les processus de mémorisation vers d'autres objets (méthodes, mises en relations...). L'enseignement occidental a depuis longtemps réduit la part de mémorisation (comparé par exemple aux écoles des religions du livre, ou à l'enseignement en Chine), et il peut désormais passer à la vitesse supérieure.
- tirer parti de cet allègement cognitif pour mobiliser le temps et l'attention de l'enfant vers d'autres apprentissages, notamment celui de la sérendipité. Trop souvent caricaturé comme un vagabondage sur le Web, la sérendipité appelle une posture de curiosité, un sens critique sélectif et des méthodes spécifiques pour être fructueuse. Cette curiosité et cette exigence informationnelles peuvent désormais jouer un rôle essentiel dans le processus d'apprentissage.

Il est aujourd'hui tout aussi important de savoir chercher de l'information, de citer des sources, de construire un avis critique, de chercher par sérendipité, d'apprendre à interpréter que d'acquérir des savoirs et des connaissances disciplinaires. L'un n'est pas substituable à l'autre et inversement.



[classement] Avec la formation d'une sphère publique numérique s'est installé un système original de contournement du pouvoir que les « gardiens » (*gatekeepers*) de l'espace public traditionnel, journalistes, éditeurs, experts ou documentalistes exerçaient sur la sélection et la hiérarchisation des informations.

L'ordre des informations entre algorithmes et subjectivités

Sans doute l'une des promesses les plus subversives du Web fut-elle d'offrir une architecture d'accès à l'information si plastique qu'elle permettait d'interroger à nouveau un ensemble de questions que la clôture oligarchique de l'espace public ne permettait plus de penser dans les catégories rouillées des médias traditionnels. À partir de quel principe doit-on ordonner et classer les informations ? À quel procédé faut-il se fier pour donner de la visibilité à telle information plutôt qu'à telle autre ? Comment définir des voies d'accès et des circuits de hiérarchisation des contenus qui soient susceptibles de corriger les effets tant décriés du contrôle éditorial des *gatekeepers* : le mépris pour les périphéries, la connivence idéologique ou la soumission aux logiques économiques ? On voudrait revenir ici sur la façon dont différents procédés se sont mis en place pour organiser autrement - car il s'agit bien de suivre d'autres routes - l'information numérique. Pour cela, il nous faut identifier, au sein des principes sur lesquels reposent ces procédés, les réponses qu'ils essayent d'apporter aux critiques qui étaient faites aux *gatekeepers* traditionnels. Mais il importe aussi d'examiner les critiques qu'ils suscitent aujourd'hui en raison du rapport de plus en plus ambivalent que nous entretenons à l'égard des algorithmes du Web et de l'intensification de l'individualisme contemporain qui postule l'inaliénable

autonomie du sujet souverain dans la définition de ses centres d'intérêts.

DEUX CRITIQUES DES GATEKEEPERS

La critique, qui n'a cessé de soutenir toutes les initiatives introduites par les acteurs du Web pour classer les informations, s'est dirigée contre le pouvoir éditorial des *gatekeepers*. En raison du contrôle que ceux-ci exerçaient sur les moyens de diffusion, ils avaient pu faire de la rareté des informations publiées un moyen d'introduire des règles de sélection et de hiérarchisation propres à leurs mondes professionnels (journalisme, édition ou documentation). Gouverné dans chaque univers par des normes déontologiques historiquement construites autour des représentations dominantes de l'actualité, du publiable et de l'intérêt général, le travail des *gatekeepers* n'a cessé de se voir contester en raison des oublis, des exclusions et des déformations qu'il imprimait sur la représentation du monde livrée, sans retour possible, au public. Mais cette critique des *gatekeepers* a pris deux directions contradictoires : elle soutient d'abord que les informations présentées par les médias sont biaisées et qu'il importe de les redresser ; mais elle considère aussi que les informations que les médias poussent au centre de l'espace public n'intéressent pas les publics parce qu'elles ne prennent pas en compte leur diversité subjective. La première critique

s'inquiète des biais produits par la position de pouvoir des *gatekeepers* (critique anti-hégémonique) et la seconde voudrait élargir la sphère publique à de nouveaux formats de paroles et de contenus favorisant la participation des publics (critique expressiviste). Ces deux reproches adressés aux médias traditionnels, celui de ne pas être neutre et celui d'écraser la diversité, nourrissent contradictoirement la plupart des alternatives aux médias traditionnels qui se sont développées dans l'histoire des pratiques informationnelles¹. Les outils informationnels du Web ont cherché à répondre simultanément à ces deux types de critiques. Et, sans doute, est-ce parce que nous ne clarifions pas les attentes différentes, voire antagonistes, exprimées à travers ces deux types de critiques que nous tenons des discours aussi contradictoires sur les formes désirables de l'accès à l'information numérique.

LA NEUTRALITÉ PROCÉDURALE DES ALGORITHMES

À la première critique, celle des biais hégémoniques, le Web a répondu par des techniques procédurales en confiant la hiérarchisation des informations aux algorithmes des moteurs de recherche et aux compteurs des métriques d'audience. Face à l'arbitraire du procédé éditorial, les outils calculatoires du Web se sont construits sur la promesse de réparer les déformations des choix

Dominique Cordon est sociologue au Laboratoire des usages d'Orange Labs et professeur associé à l'Université de Marne-la-Vallée. Ses travaux portent sur les usages d'internet et les transformations de l'espace public numérique. Ses recherches récentes ont pour objet les réseaux sociaux de l'internet, les formes d'identité en ligne, l'autoproduction amateur et l'analyse des formes de coopération et de gouvernance dans les grands collectifs en ligne. Il conduit aujourd'hui une analyse sociologique des algorithmes permettant d'organiser l'information sur le Web. Il est par ailleurs l'auteur de nombreuses publications.

domi.cordon@orange.com

humains des oligarques de l'édition, du journalisme et de la documentation. Le principe calculatoire du PageRank, qui s'est aujourd'hui imposé à l'ensemble des moteurs de recherche du Web, a ainsi mis en œuvre d'une façon particulièrement originale le principe d'une « intelligence collective » des internautes. C'est parce que le Web a été conçu comme un tissu de sites qui s'adressent entre eux de la reconnaissance à travers les liens hypertextes qu'il est possible de le classer. En mesurant l'« autorité » au moyen d'un procédé classant les informations en fonction du nombre de liens qui pointent vers elles, pondéré par l'autorité attribuée à ces mêmes liens, le Page Rank est parvenu à généraliser à l'ensemble de l'Internet une mesure de la reconnaissance qui était jusqu'alors réservée au classement de la production scientifique¹. Le PageRank fait de tous les internautes qui publient des *gatekeepers* et propose, chose absolument inédite dans l'univers de l'information, de hiérarchiser les informations selon l'autorité qu'elles ont reçue des autres internautes publiants. À la différence de l'autorité, le procédé de classement par l'audience qui s'est généralisé sur le Web (YouTube par exemple) additionne le nombre de vues sur la base d'un calcul de popularité. Mesurant également le nombre d'internautes cliqueurs, il donne au classement de l'information une mesure immédiate et auto-réfléchissante de son impact sur le public. Aussi critiquée soit-elle, notamment en raison de sa proximité avec les intérêts du marché publicitaire, la mesure d'audience n'en constitue pas moins une manière légitime dans les sociétés démocratiques de faire du comportement du public l'ordonnateur de ses choix².

Ces procédés algorithmiques, la mesure de l'autorité et de l'audience, proposent deux alternatives pour hiérarchiser les informations à partir des comportements de ceux qui les citent ou les lisent, le lien hypertexte ou le clic. Ils postulent une extériorité du procédé d'ordonnement à l'égard de ceux qui l'utilisent. Hiérarchiser les informations selon l'autorité ou l'audience reste indépendant des désirs, des goûts et des choix de ceux qui y accèdent. Ces procédés se proposent simplement de « purifier » les choix humains des *gatekeepers* en demandant à un automate de résoudre un problème de choix

collectif d'une manière qui soit la plus « neutre » possible. Outre le fait que cette neutralité est de plus en plus mise à mal par le secret des algorithmes et un certain nombre d'effets mathématiques qui contribuent à transformer, au sommet de la hiérarchie, l'autorité en une mesure de popularité (c'est-à-dire d'audience), cette extériorité-univoque et centralisatrice fait désormais l'objet d'une critique de plus en plus forte de la part des internautes. Les algorithmes sont désormais soupçonnés d'enfermer les intérêts commerciaux des plateformes qui les conçoivent, relançant à nouveau la critique contre-hégémonique dans la recherche de biais déformant la « vraie » représentation des informations.

LA SUBJECTIVITÉ DES INTERNAUTES COMME PRINCIPE DE CLASSEMENT DE L'INFORMATION

Si le Web a donné une réponse algorithmique à la critique de partialité des médias, c'est en revanche avec de tout autres procédés qu'une solution a été apportée à la seconde critique. Cette fois, il ne s'agit plus de corriger des défauts dans la représentation des informations, mais d'élargir les formats expressifs de l'information en favorisant la diversité subjective des publics et en multipliant les parcours de navigation alternatifs. Pourquoi en effet produire un classement unique, écrasant et mathématique, quand les internautes témoignent de centres d'intérêt et de préoccupations si divers ? Comment ouvrir les formats de publication à des modes d'énonciation pluriels, moins cadenassés, appropriables par un cercle plus large de preneurs de parole ? C'est à cette seconde critique, celle de l'écrasement des singularités par l'hégémonie des calculs centraux, qu'est venu répondre un autre procédé de classement de l'information avec l'émergence du web participatif, des communautés et des réseaux sociaux. L'information n'est plus hiérarchisée de façon extérieure au public qui la consulte, mais à l'intérieur même des goûts, des centres d'intérêt et des échanges du public. L'ensemble des artefacts conçus par les plateformes du web social (Like, commentaires, Re-tweet, fils d'actualité, etc.) propose à l'internaute un classement de l'information personnalisé qui s'appuie sur la sociabilité des internautes (à travers les choix d'amis et de *followers*)

et sur ses actions de partage au sein du réseau. La hiérarchie des informations apparaît ainsi comme la conséquence des choix, des désirs et des goûts de l'internaute. Elle n'est plus extérieure à ses pratiques, mais se situe à l'intérieur même des activités par lesquelles il construit son identité numérique. Le succès du web des communautés et des réseaux sociaux s'enracine dans un ensemble d'attentes revendiquant le droit des individus de s'exprimer et de se confronter à la diversité sans avoir à se plier à l'étroit canal de diffusion des *gatekeepers* traditionnels. Relayant dans le domaine de l'information les effets de la crise de confiance dans toutes formes d'institutions, d'autorités ou de porte-parole, cette manière de circuler dans l'information numérique à partir de son réseau, de ses goûts et de ses partages témoigne de la volonté des individus de prendre eux-mêmes en charge ce dont ils déléguaient précédemment la responsabilité à d'autres : définir ce qui les intéresse ou ne les intéresse pas. En refusant leur allégeance à des agents chargés de faire des choix en leur nom, ils ouvraient ainsi une voie à une toute autre forme de hiérarchisation de l'information qui ne s'appuie ni sur les choix éditoriaux des *gatekeepers*, ni sur l'agrégation collective des jugements des internautes publiants ou cliqueurs, mais sur la prétention des individus à être à eux-mêmes leur propre *gatekeeper*.

NEUTRALITÉ ET SUBJECTIVITÉ

En critiquant les *gatekeepers* traditionnels, le Web a ainsi ouvert deux directions opposées au classement de l'information. La première confie à un calcul le soin de produire des choix collectifs, la seconde personnalise la fenêtre sur l'information des individus en la dessinant depuis leurs propres choix. La première prétend à la neutralité procédurale, la seconde assume et revendique la subjectivité des individus. La première est unique, centrale et homogène quand la seconde éclate les choix informationnels en une myriade

/////

¹ Dominique CARDON, Fabien GRANJON, *Mediactistes*, Presses de Sciences Po, 2013 (2^e éd.)

² Dominique CARDON, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Recherches*, 2013, vol. 31, n°177, pp. 63-95.

³ Cécile MÉADEL, *Quantifier le public. Histoire des mesures d'audience à la radio et à la télévision*, Economica, 2010

///// de niches communautaires et de nébuleuses individuelles. Si ces deux tendances s'observent dans les dispositifs numériques de classement de l'information, c'est parce qu'elles témoignent d'un système d'attentes qui associe deux attitudes apparemment contradictoires : il faudrait que les choix collectifs soient absolument « neutres » et débarrassés de toute sorte de biais pour que les individus puissent manifester plus librement la spécificité de leurs choix individuels. La hiérarchie de l'information venant de l'extérieur ne devrait être entachée d'aucun biais de sélection afin que, de l'intérieur, chaque sujet soit souverain dans ses choix.

Or, à ne pas clarifier ces deux dynamiques contradictoires, le débat sur la recherche d'information ne cesse de reprocher à l'une ce que fait l'autre et inversement. En effet, il est d'abord

demandé aux choix collectifs des algorithmes d'éviter les biais humains et économiques des plateformes et de tendre vers une très improbable « neutralité ». Mais, dans le même mouvement, il leur est reproché de produire des résultats centraux écrasant de conformisme qui dissimulent les périphéries et détruisent la diversité du Web.

Cependant, si ces algorithmes, dans l'esprit de la seconde critique, « personnalisent » leurs résultats, comme le font de plus en plus les moteurs de recherche, ils sont aussitôt accusés de sacrifier l'extériorité et la neutralité que réclame la première critique. Ces contradictions sont aussi apparentes à l'endroit de l'information classée par les métriques du web social. En se proclamant autonomes et souverains dans le choix des sources d'informations désirables, les

internautes se déclarent aussitôt inquiets du possible enfermement de certains dans la « bulle » de leurs propres centres d'intérêt⁴ - raisonnant en « troisième personne », les individus craignent toujours que les autres soient « enfermés » dans leurs choix, alors qu'eux-mêmes se montreraient capables de se libérer de ce risque aliénant. Livrés à eux-mêmes, les individus veulent choisir et refusent de déléguer leur confiance à d'autres pour faire ces choix, mais craignent alors que l'autonomie si ardemment revendiquée soit aussi une prison ! Sans doute serait-il utile d'examiner avec plus de clarté les représentations contradictoires qui sont aujourd'hui en jeu dans les débats sur le classement et la recherche d'information. »

⁴ EILI PARISER, *The Filter Bubble. What the Internet is Hiding from You*. New York : The Penguin Press, 2011

A propos de la sérendipité

Rémi Sussan 16/7/2009 @ 10:51 In Communication interpersonnelle, Interviews, Usages | <http://www.internetactu.net/2009/07/16/a-propos-de-la-serendipite/>

Il était une fois, nous dit un – [soi disant ?](#) ^[1] – conte persan, trois princes du royaume de Serendip qui, alors qu'ils étaient en voyage, découvrirent des traces du passage d'un chameau.. *“L'ainé observa que l'herbe à gauche de la trace était broutée, mais que l'herbe de l'autre côté ne l'était pas. Il en conclut que le chameau ne voyait pas de l'oeil droit. Le cadet remarqua sur le bord gauche du chemin des morceaux d'herbes mâchées de la taille d'une dent de chameau. Il réalisa alors que le chameau pouvait avoir perdu une dent. Du fait que les traces d'un pied de chameau étaient moins marquées dans le sol, le benjamin inféra que le chameau boitait.”* nous racontait déjà Pek van Andel, chercheur en sciences médicales à l'université de Groningue, dans un ancien article sur le sujet que l'on trouve sur le site [Automates intelligents](#) ^[2].

La sérendipité n'est pas le hasard

De ce conte vient l'expression anglo-saxonne *serendipity*, dont l'équivalent français (sérendipité) ne figure pas encore dans nos dictionnaires. Elle désigne l'action qui consiste à trouver quelque chose qui n'était pas initialement prévu par la recherche.

Depuis cet article, Pek van Andel s'est associé à Danièle Boursier, directrice de recherche en sciences sociales au CNRS, pour rédiger [De la sérendipité dans la science, la technique, l'art et le droit : Leçons de l'inattendu](#) ^[3], paru chez [L'Act Mem](#) ^[4] (collection libre science). Les auteurs y tentent d'esquisser un panorama de cet étrange phénomène.

La sérendipité est un phénomène complexe. Les deux auteurs ont pu ainsi en distinguer plusieurs types. Difficile d'ailleurs de repérer, dans les histoires de découvertes colportées par la culture populaire, les cas de “véritable sérendipité” : la pomme de Newton ? Trop beau pour être vrai ! Le savant aurait inventé ce récit dans les dernières années de sa vie, apprend-on. Les antibiotiques de [Flemming](#) ^[5], archétype de la découverte aléatoire ? C'est un cas de pseudo sérendipité, nous expliquent Boursier et van Andel. Flemming était déjà en train de chercher des substances antibactériennes, seulement, nous disent-ils, il les trouva par un chemin imprévu. Alors que la découverte était attendue, c'est la méthode qui a plutôt été originale . Si vous cherchez un bon exemple de sérendipité authentique, optez plutôt pour la découverte de l'Amérique.

Il importe tout d'abord d'éviter de confondre trop facilement la sérendipité et le “hasard”. D'ailleurs, le conte nous montre plutôt des ancêtres de Sherlock Holmes (dont les méthodes sont d'ailleurs largement analysées dans le livre) que des découvreurs guidés exclusivement par la chance. *“La sérendipité met en valeur la sagacité ou perspicacité de celui qui tombe dessus. Si on n'est pas préparé, on ne voit pas le fait”*, nous explique Danièle Boursier. *“Plus on connaît, plus on travaille, plus on possède une vision, plus on a de chance découvrir une fleur au bas côté de la rue”*, renchérit Pek van Andel.

Dans leur ouvrage, les deux auteurs multiplient les cas et les exemples, ce qui en fait un livre très agréable à lire, loin des lourds traités théoriques.

Pour Danièle Boursier, *“c'est un parti pris... Nous voulons redonner sa valeur au rôle du narratif, de la singularité. Trop généraliser, ça peut être intéressant, mais aussi asséchant. Il peut être bon d'étudier les individus cas par cas. On ne raconte pas seulement des histoires. On reprend ce que des gens ont raconté sur leurs expériences, que ce soit des scientifiques ou pas. Ce n'est pas la vérité qui nous intéresse, c'est ce que les gens pensent être vrai.”*

Mais les auteurs ne se contentent pas d'aligner les anecdotes. On trouve aussi des explications théoriques très précises de la sérendipité, qui se rattacherait à un mode de raisonnement peu connu : l'abduction. La plupart des textes d'épistémologie mentionnent avant tout la déduction et l'induction. L'induction consiste à déduire de grands principes à partir d'une série d'observations. La déduction est la démarche contraire : il s'agit, à partir d'un loi connue et considérée comme vraie, d'en tirer toutes les conséquences. L'abduction élabore des hypothèses à partir de faits étonnants. Pour le philosophe américain [Charles Peirce](#) ^[6], cité dans l'article d'Automates Intelligents : *“l'abduction est le processus de l'imagination d'une hypothèse explicative. C'est la seule opération logique qui introduit une idée neuve quelconque ; parce que l'induction détermine une valeur, et la déduction dérive seulement les conséquences inévitables d'une hypothèse pure. La déduction prouve que quelque chose doit être. L'induction montre que quelque chose marche de facto. L'abduction suggère seulement que cela serait possible”*. *“L'abduction est par excellence le mode de raisonnement du diagnostic médical ou de l'enquête judiciaire”*, est-il écrit dans *De la sérendipité*. *“Quand un médecin cherche à diagnostiquer une maladie, il raisonne par abduction en interprétant les symptômes d'un patient. De même, un détective, un policier, un juge instruisent l'affaire dont ils sont saisis en faisant des inférences qui sont souvent des abductions”*.

Appréhender les effets de la sérendipité

La sérendipité ne se limite pas à la science. Elle touche également le droit, la technologie, la société. Les deux auteurs reviennent souvent sur l'exemple de [Rosa Parks](#) ^[7], la femme noire qui, en refusant de s'asseoir au fond du bus, à la place réservée aux “gens de couleur”, comme on disait alors, avait déclenché le mouvement des droits civiques aux Etats-Unis. Un petit événement aux conséquences incalculables... Mais là aussi, n'imaginons pas qu'il s'agisse d'un pur hasard.

“Rosa Parks était une princesse de Serendip”, nous dit van Andel, *“car elle s'était renseignée sur les droits civiques ! Dix ans avant elle avait effectué le même acte sans obtenir le même résultat.”* Du reste, l'un des grands points forts de la démocratie est sa capacité à intégrer la sérendipité : *“la rationalité morale de la démocratie se résumerait à favoriser la disponibilité à accueillir l'inconnu, c'est-à-dire le nouveau fait, la nouvelle interprétation, le nouvel argument (...)”* écrivent les auteurs dans leur ouvrage. Mais il ne faut pas oublier non plus la sérendipité négative, les effets pervers qui se manifestent à la suite de décisions pourtant rationnelles : *“par exemple explique Danièle Bourcier, vous créez le RMI en pensant qu'il concernera 500 000 ayants droits, puis, lorsque le dispositif est mis en place dans la population, vous vous apercevez qu'il y en a le double”*.

Comment apprivoiser un tel phénomène, qui par définition échappe aux méthodes et ne s'accommode pas de mode d'emploi ? On ne peut bien sûr maîtriser la sérendipité, mais au moins on peut créer des contextes, des environnements, où elle a le plus de chances de se manifester. C'est ce qu'on appelle la “sérendipité institutionnalisée”.

Comment favoriser la sérendipité ?

Selon Pek van Andel, *“Aux Pays-Bas, il y a une tradition chez les chercheurs : le vendredi après-midi, ceux-ci disposent de la liberté d'accomplir des “recherches personnelles”*. *Chaque chercheur a officiellement une marge de manœuvre, sans être obligé de justifier ses actions”*.

“Dans certains instituts d'études avancées, on réunit des chercheurs venus de disciplines diverses et pas forcément en rapport, et on leur dit “et bien maintenant pensez !” jusqu'à ce qu'ils interfèrent les uns avec les autres. Ces instituts ont pour but de mettre des gens en relation et surtout de créer un bouillon de culture d'où sortira éventuellement un nouveau paradigme”, explique Danièle Bourcier, qui d'ailleurs a vécu l'expérience : *“dans un institut*

dépendant de l'académie des sciences aux Pays-Bas, après avoir été sélectionnés sur dossier, on s'est retrouvé à 40 chercheurs au bord de la mer du Nord, avec pas un chat aux alentours, on nous a donné 6 mois, et on nous a dit "just think".

Pour Danièle Boursier, "c'est une question de management, mais aussi une question personnelle, car si on donne un temps de liberté à des gens qui n'ont rien à trouver, ils ne trouveront rien. C'est donc la conjonction d'un type de management plus flexible, qui donne confiance aux gens pour qu'ils puissent poursuivre ce qu'ils ont envie de faire et la capacité de celui qui se pense impliqué dans un projet d'innovation ou de recherche à prendre sa chance à un moment donné pour aller plus loin ou à côté de ce qu'on lui demandait de faire. Une certaine forme de désobéissance productive."

Mais que faire si cet espace et ce temps libre ne sont pas disponibles ? On se heurte alors au phénomène de la recherche clandestine, qui ne se découvre que quand elle donne des résultats.

Évidemment, toutes les cultures n'ont pas autant de facilité à laisser autant le champ libre aux chercheurs. "Ce qui m'a intéressé dans dans mon séjour aux Pays-Bas, explique Danièle Bourcier, c'est cette façon d'inclure les innovateurs, de leur donner confiance et plus d'espace. C'est beaucoup plus difficile dans des pays comme l'Allemagne ou la France."

Si les collectivités ont parfois du mal à intégrer la sérendipité, peut-être faut-il l'enseigner aux individus dès leur plus jeune âge, à l'école ... ? "Dans le domaine de l'éducation nous avons suggéré qu'il faudrait, plutôt que demander à des élèves de répondre à la question qu'on leur pose et trouver la solution, les encourager à trouver d'autres solutions. Et que celles-ci aient la même valeur que celle indiquée par le maître.

Pourquoi ne pas aussi enseigner les erreurs des héros de la science et de la connaissance et tout ce qu'ils ont trouvé par essai et erreur".

"J'ai proposé aux Pays-Bas, et même en France de mettre dans les travaux pratiques des éléments surprenant sans prévenir les élèves, voir si quelqu'un les remarque et ce qu'il fait avec. On pourrait ainsi découvrir les nouveaux talents de la recherche", ajoute Pek van An del.

Finalement, conclut Danièle Bourcier, "Notre travail est une réponse à l'incertitude, à la rapidité des changements et à l'impossibilité de prévoir. Mais nous en montrons le côté positif. Nous disons : Mais pourquoi vous inquiétez-vous ! L'histoire du monde a aussi été faite sur des évènements imprévisibles. Et ce sera toujours comme ça !"¹

[1] soi disant ? : http://fr.wikipedia.org/wiki/Les_Trois_Princes_de_Serendip

[2] Automates intelligents: <http://www.automatesintelligents.com/echanges/2005/fev/serendipite.html>

[3] De la sérendipité dans la science, la technique, l'art et le droit : Leçons de l'inattendu:

<http://www.amazon.fr/s/%C3%A9rendipit%C3%A9-dans-science-technique-droit/dp/2355130183/internetnet-21>

[4] L'Act Mem: <http://www.lactmem.com/>

[5] Flemming: http://fr.wikipedia.org/wiki/Alexander_Fleming

[6] Charles Peirce: http://fr.wikipedia.org/wiki/Charles_Sanders_Peirce

[7] Rosa Parks: http://fr.wikipedia.org/wiki/Rosa_Parks

Communication au colloque « L'éducation à la culture informationnelle », Lille, 16-17-18 octobre 2008, <http://archivesic.ccsd.cnrs.fr/sic_00344161>

Le rôle des copiés-collés dans l'activité de recherche d'information des élèves du secondaire

Nicole Boubée

L'utilisation par les élèves du copier-coller, fonctionnalités anciennes et basiques des outils informatiques, amplement favorisée par l'accessibilité croissante de documents numériques ne cesse d'être stigmatisée du point de vue des apprentissages du domaine ou informationnels. Nous entrons dans l'analyse de ce phénomène en replaçant les opérations de copier-coller dans l'ensemble du processus de recherche d'information. Nous prenons appui sur notre étude empirique portant sur l'activité informationnelle d'élèves du secondaire, inscrite dans le cadre de la *Library and Information Science (LIS)* et menée à l'aide de la méthode de l'autoconfrontation croisée. Ainsi interrogeons-nous le(s) moment(s) du processus dans lesquels interviennent les copiés-collés, la forme et le contenu du document dans lequel sont collés les extraits et les verbalisations des élèves sur leurs copiés-collés. Après avoir retracé brièvement les investigations scientifiques décrivant le copier-coller comme prise de notes et plagiat, nous présenterons nos dispositif d'observation et résultats qui nous permettent de suggérer l'importance du rôle des copiés-collés dans le processus informationnel ainsi qu'une orientation pour l'encadrement pédagogique des pratiques documentaires.

1 Les copiés-collés comme prise de notes ou plagiat

Les rares investigations scientifiques de la pratique du copier-coller l'appréhendent au mieux en tant que prise de notes non reformulées ou par la négative en terme de plagiat.

1.1 Prises de notes

Des études psychologiques ont comparé les différents types de prises de notes dont le copier-coller fait partie. Igo *et al.* (2005) rappellent que, pour l'imprimé, le résumé, la paraphrase ou l'annotation des idées principales sont plus efficaces que la recopie mot à mot. Dans l'environnement numérique, ils montrent que le copier-coller restreint (par un outil informatique obligeant à écrire dans une matrice un nombre limité de mots) entraîne, chez les étudiants, un traitement plus en profondeur de l'information que le copier-coller non restreint, mesuré en terme de volume de restitution des idées contenues dans les documents. Mais les auteurs restent prudents: l'efficacité pourrait également dépendre de la profondeur d'analyse au moment de la prise de note ou de la révision ultérieure de ces notes.

L'intérêt de considérer le copier-coller comme prise de notes permet de revenir sur la phase de collecte dans le processus de recherche d'information tel que décrit par le modèle *ISP (Information Search Process)* de Kuhlthau (1991). Dans cette modélisation élaborée pour l'essentiel à partir d'observations d'élèves du secondaire, la collecte intervient une fois la

résolution du problème informationnel (*focus formulation*). Les actions décrites pour cette étape consistent dans le recueil de références bibliographiques et la prise de notes détaillées. Compte tenu de la place de la collecte dans ce modèle (avant-dernière étape), l'information collectée est plutôt spécifique que générale. De plus, dans cette phase, le sentiment de confiance continue d'augmenter. A partir de la modélisation *ISP*, il serait donc possible de considérer que les copiés-collés signalent un processus informationnel qui s'effectue de façon satisfaisante pour le chercheur d'information.

1.2 Plagiat

Le thème du plagiat lorsqu'il se rapporte aux pratiques documentaires peut fournir des vues plus précises des phases de collecte et d'usage de l'information. En 1995, Pitts note que les lycéens de 1ère et terminale trouvent acceptable de plagier une source et relève leur difficulté à organiser l'information au point que certains d'entre eux se contentent d'utiliser l'information dans l'ordre dans lequel ils l'ont trouvée. Chez les plus jeunes, Large et Beheshti (2000) constatent que le plagiat serait moins tentant avec les documents du Web malgré la facilité technique du copier-coller. Les chercheurs l'expliquent par le fait que les phrases dans les pages Web sont d'une qualité d'écriture moindre que celle des livres et seraient plus faciles à reformuler pour leurs jeunes participants de 6ème. McGregor et Streitenberger (1998) et MacGregor et Williamson (2005) examinent le copier-coller d'élèves de 1ère. Les lycéens les moins engagés dans l'activité informationnelle accordent de l'importance à la forme de leur production finale qui doit avoir une apparence qui « sonne juste » (« *look good* », « *sound right* »). Ceux-ci tendent à copier massivement les documents alors que les lycéens plus engagés dans l'activité de recherche et leurs apprentissages résumant ou paraphrasent les documents. Les auteurs indiquent également que la présence de consignes de l'enseignant sur la reformulation de l'information collectée limite l'étendue du plagiat sans pour autant l'éliminer entièrement. En outre, ils relèvent dans les productions de nombreuses erreurs et en concluent que les consignes déplacent simplement le problème. Recopies ou erreurs rendent compte des difficultés des élèves à se centrer sur le processus informationnel et à s'approprier leurs thèmes de recherche. Le thème du plagiat malgré son intérêt pour saisir les dimensions de la motivation et de la compréhension lors des activités documentaires a pour inconvénient de faire l'impasse sur la nature du processus de recherche d'information. La phase d'évaluation de l'information a fait l'objet de nombreuses investigations qui ont mis en évidence les aspects dynamiques et situés du processus de recherche d'information (Schamber, Eisenberg et Nilan, 1990). Il serait donc étonnant que la phase de collecte ne présente pas les mêmes caractéristiques. Ceci implique de considérer que le copier-coller n'est pas seulement le résultat de l'activité mais qu'il participe à son effectuation.

2 Contribution empirique

2.1 Cadre méthodologique

Notre étude s'appuie sur l'observation de quinze binômes d'élèves de la 6ème à la terminale. Nous n'avons pas donné de tâche de recherche, les élèves choisissant d'accomplir une tâche prescrite par un enseignant (tâche prescrite ou semi-prescrite lorsque le thème est laissé au libre choix des élèves) ou définissant leur thème de recherche à partir d'un intérêt personnel (tâche autogénérée). L'activité est filmée et des entretiens d'autoconfrontation sont effectués à partir du visionnage du film de l'activité environ une semaine plus tard. L'ensemble se déroule dans les établissements scolaires des élèves. Les observations ont eu lieu en 2005 et

2006. Les élèves décidant eux-mêmes d'arrêter leur tâche, la durée des sessions varie, 15 min pour la plus courte, 1h 04 pour la plus longue. La durée la plus commune oscille autour de 45 min. Les entretiens d'autoconfrontation durent pour la majorité d'entre eux environ 50 min.

Nous définissons le copier-coller comme un type d'extraction de l'information générant un « document de collecte » dans lequel sont placées ensemble des parties de documents copiées-collées dans le traitement de texte. Cette extraction se distingue de la prise de notes manuscrites ou de l'impression d'un document entier, autres extractions présentes dans nos sessions. 7 sessions de collégiens et lycéens ont élaboré un document de collecte. Tous les types de tâches, prescrites et semi- prescrites ou autogénérées ont donné lieu à un document de collecte ce qui n'est pas le cas des autres types d'extraction.

2.2 Résultats

2.2 1 Un processus de recherche d'information scandé par les collectes

Le volume de collectes est variable selon les sessions (11, 9, 8, 7, 5, 4, 3 collectes). Le nombre de collecte effectué par chaque binôme peut être considéré comme élevé pour ces recherches documentaires excédant rarement une heure. Le rythme des prélèvements est plutôt régulier. Quelle que soit la tâche, les collectes débutent tôt, voire très tôt dans l'activité (4 binômes entre moins de 4 min et 7 min). Une seule session montre un 1er prélèvement tardif (presque 30 min) qui rend compte d'une difficulté dans la définition du besoin d'information.

2.2 2 L'élaboration du document de collecte

Le document de collecte montre des traits communs dans la forme :

- *l'empilement des extraits collés les uns après les autres*. Cet empilement fait l'objet de commentaires identiques dans 2 sessions, les élèves de binômes lycéens s'assurant auprès de leurs camarades respectifs de l'empilement : « *Tu le mets à la suite de l'autre* » ; « *vas-y à la suite* ». Les verbalisations lors de l'activité indiquent que les jeunes attribuent à l'extrait une place future dans la production finale. Mais dans le cours de l'activité, l'extrait collé est simplement empilé. Cet empilement peut sembler dispenser du traitement conceptuel de l'information prélevée et les élèves en témoignent spontanément dans l'activité ou les entretiens. Une lycéenne précise : « (...) *on va sur Internet, on fait copier-coller des textes comme ça sans trop regarder dans quel ordre on les met puis après, une fois qu'on a à peu près tous nos éléments, on les reclasse, on les retravaille, on les réécrit, on précise* ». Par divers commentaires les élèves rendent compte de la profondeur de leur traitement de l'information et disent eux-mêmes ne pas avoir produit un traitement suffisant des extraits collectés.
- *la mise en forme du document de collecte différée*. Elle consiste en simple séparation par retour à la ligne, en modification de la police ou de la taille des caractères. Elle est le plus souvent différée au point que les retours à la ligne ne sont pas systématiquement réalisés, le but étant, semble-t-il, d'abord d'exécuter le processus informationnel : « *Tu colles tout comme ça après on mettra une mise en page* » dit une lycéenne après le prélèvement du 6e extrait collé ; « *ça ressemble à rien comme ça* »

constate un lycéen dans un autre binôme. « *Ouais, on figolera tout à l'heure* » lui répond sa binôme.

- *la longueur du document de collecte*. Les prélèvements sont généralement de taille modeste (quelques lignes, un ou deux paragraphes et contenant souvent des images). Les appréciations portées par les élèves sur la longueur montrent qu'au-delà d'environ 3 pages, ceux-ci commencent à trouver le document suffisamment rempli. Faisant défiler le document, les binômes vérifient la longueur du document pour conforter leur prise de décision sur l'arrêt ou la poursuite de la recherche. La quantité d'information contenue dans le document est la première mesure qui leur permet d'évaluer leur activité. Ce n'est pas la seule. Tous les binômes évaluent en plus le contenu du document de collecte.

2.2 3 Contenu des collectes

Les aspects formels du document de collecte et les commentaires associés incitent à penser que ce qui est empilé, ce sont aussi les jugements de pertinence affectés d'une valeur positive ou partiellement positive. Plusieurs éléments peuvent confirmer cette perspective. D'abord le maintien d'extraits mal collés ou difficiles à coller est constaté dans plusieurs sessions, généralement lors des premières collectes, phase du processus dans lequel on peut supposer que les critères Adéquation au thème et Nouveauté pèsent encore de façon significative. De plus les élèves « passent en revue » le document de collecte. Ils le font très régulièrement défiler dans le cours de l'activité en reprenant à voix haute les différentes pistes explorées ayant donné lieu à des prélèvements. Les binômes en effectuant cette vérification du contenu du document de collecte reprennent ainsi l'ensemble du parcours et des réflexions menées et ce dans le même ordre que la conduite de leur activité. Ils paraissent s'assurer de la progression de leurs travaux de recherche à partir d'un objet physique, leur document de collecte. En outre la concordance systématique entre le contenu du prélèvement et celui de la requête qui la précède semble confirmer que ce ne sont pas des collectes par défaut. Mais le lien le plus notable entre collecte et requête réside dans la reformulation qui suit la collecte. En effet une collecte entraîne fréquemment une reformulation de requête plutôt qu'une poursuite de la consultation du site dans lequel vient d'être prélevé l'extrait. L'examen des requêtes montre que ces reformulations contiennent souvent un nouveau concept.

2.2.4 Les fonctions du document de collecte explicitées par les élèves

Les élèves ont intégré les critiques qui leur sont faites vis-à-vis du copier-coller. Les autodépréciations courent en filigrane dans leurs verbalisations. En même temps, ils restituent une dizaine de motifs concernant majoritairement leur processus informationnel. Les copiés-collés servent à vérifier l'état de complétude du traitement du thème : « *Je vais rechercher si on a pas besoin d'autre chose* » dit ce lycéen en retournant dans le document de collecte après 35 min d'activité et faisant défiler le document : « *“la définition”, il descend dans le document, “qui est concerné”, il passe la souris sur une adresse URL, surtout ça, si ça concerne vraiment les jeunes, il continue à faire défiler la page, comparer, il faut relier tout ça.* » ; la qualité de la recherche : « *Pour voir si moi j'allais dire si c'est bien ou si c'est pas bien* » dit cette collégienne de 6ème qui explicite pourquoi sa binôme lui avait demandé dans l'activité de regarder son document de collecte, « *s'il faut rajouter des choses* » poursuit-elle ; stocker l'information pour continuer l'activité dans le Web : « *C'est pour ça en fait qu'on va*

faire une longue recherche parce qu'on va re y aller plusieurs fois » dit un lycéen ou, au contraire, poursuivre l'activité hors du Web : « *Comme ça on ne reste pas sur Internet* » dirait-il également ; réduire l'information pour ne pas se perdre : « *Ah, j'ai une idée ! Ce que tu vas faire, tu vas aller dans Office 2000. C'est ce que je fais pour ne pas me perdre* » dit cette élève de 4ème à sa binôme après 8 min d'activité ; pour éviter le multi-fenêtrage : « *Moi, c'est [le copier-coller] plus quand je fais une recherche sur plein de sites, enfin au lieu d'avoir plein de fenêtres ouvertes* » explicite cette lycéenne de terminale ; pour gérer le temps : « *Ce qui est bien avec le copier-coller, c'est qu'on peut sélectionner tout et ensuite vu qu'on a pas vraiment le temps sur l'ordinateur, on pourrait voir chez nous après* » dit une lycéenne de 1ère ; vérifier l'adéquation avec la tâche : « *On avait pas tout mais c'est vrai que, pour un exposé d'un quart d'heure, on avait déjà beaucoup de choses. D'ailleurs on avait 12 pages [plusieurs pages ne contiennent qu'une colonne étroite]* » explicite cette même lycéenne ; pour une bonne exploitation de l'information et pour lire : « *J'avais trouvé un truc intéressant pour mettre sur Word, après l'imprimer* » dit ce lycéen en BEP. Ainsi les motifs liés au document de collecte réfèrent-ils à de multiples aspects du processus informationnel bien plus qu'à la production finale qui dans le cas de tâche prescrite sera communiquée à l'enseignant.

3 Discussion et conclusion

Nous avons tenté de cerner les contours de la pratique du copier-coller souvent dénigrée du point de vue de l'apprentissage. Bien que nos sessions n'aient pas toutes produit un document de collecte, ce fait réduisant d'autant la taille de notre échantillon, des traits communs persistants d'un binôme à l'autre nous permettent de proposer une analyse de ce phénomène. Nous avons posé la question de sa fonction dans le processus de recherche d'information des jeunes. Une première indication de son importance dans l'activité est fournie par sa présence dans toutes les tâches, prescrites ou autogénérées. La pratique du copier-coller n'est donc pas strictement liée à la nécessité de communiquer à l'enseignant prescripteur une production finale. Un deuxième élément conforte notre vision du copier-coller comme composante du processus de recherche d'information. Ces collectes d'extraits de documents interviennent tôt dans l'activité et régulièrement tout au long de la recherche quel que soit le niveau et la tâche. Une modification de ces traits comme une collecte tardive ou irrégulière signale une difficulté dans l'activité. Ainsi le copier-coller est-il un « bon signe » dans le processus informationnel.

L'empilement des extraits dans l'ordre de « retrouvage », constant d'une session à l'autre, constitue une autre caractéristique majeure de ce mode de collecte. Il peut certes s'analyser comme une incapacité à organiser la documentation retrouvée (Pitts, *Ibid*). Mais divers éléments, verbalisations entre élèves s'assurant de l'empilement, remises à plus tard des opérations de mise en page pourtant jugées indispensables pour une lecture ultérieure, montrent la primauté accordée à la poursuite du processus informationnel et non au document en cours d'élaboration. Nous avons suggéré que ce qui est empilé, ce sont d'abord leurs jugements de pertinence. Les jugements étant établis de façon dynamique comme les études empiriques s'attachent à le montrer, il est probable que le maintien de l'empilement ne soit pas seulement dû aux contingences pratiques liées à la manipulation du traitement de texte. Déplacer un extrait serait rompre la logique de la sélection effectuée dans le processus. On peut comprendre que les élèves diffèrent ce réagencement, dans un deuxième temps, une fois le processus de recherche d'information clos. En maintenant l'ordre des collages, les élèves peuvent s'appuyer sur ce document de collecte en le consultant régulièrement — ce qu'ils font — pour établir le degré de satisfaction de leur besoin d'information. Le document de collecte permettrait une forme de surveillance de la démarche informationnelle. A l'aide de celui-ci, les jeunes du secondaire paraissent mettre en œuvre une forme de gestion de

l'activité de recherche d'information telle que décrite dans le modèle E.S.T. (Rouet, Tricot, 1998). La pratique du copier-coller suggérerait donc que les élèves du secondaire ne sont pas entièrement démunis de compétences métacognitives.

De même les verbalisations des élèves durant l'activité ou lors des entretiens mettent en avant l'importance du copier-coller pour accomplir la tâche de recherche même si les vertus trouvées coexistent avec un dénigrement de la pratique. Les fonctions que les élèves assignent à leurs copiés-collés multiples renvoient avant tout à la nécessité d'écarter l'information non pertinente dans la masse informationnelle du Web, dans un temps limité, en fonction d'une tâche, à partir de leurs connaissances du thème de recherche. Mais on peut esquisser une autre compréhension du phénomène du copier-coller. Ce moment d'extraction de l'information et de dépôt dans le traitement de texte pourrait être traduit comme l'action de « stabiliser » la situation telle que formulée dans la perspective phénoménologique par les sociologues Garfinkel et Goffman et en LIS par Dervin (1983, 2005). Nous avons supposé qu'après chaque collecte un changement de direction était visible dans les requêtes. Sans que cela soit systématique, il est à noter que les requêtes après un copié-collé contiennent le plus souvent un nouveau concept. Cette perspective trouve une théorisation approchante en psychologie cognitive, celle du test d'hypothèses proposée en 1975 par Levine. Elle s'oppose à la démarche moyens-fin au centre de la notion de résolution de problème (Tricot, 1998).. Les collectes dessineraient peu à peu le besoin d'information tel que défini par le chercheur d'information et en retour le document de collectes lui donnerait à voir le problème informationnel. Les copiés-collés signaleraient que les jeunes mettent en œuvre une construction particulière de résolution de leur problème d'information. Ainsi le copier-coller pour les élèves du secondaire pourrait-il avoir une double utilité, métacognitive et aide à la définition du besoin d'information.

Conclure sur l'utilité du copier-coller dans les pratiques documentaires peut paraître difficile à entendre pour les pédagogues. Dans l'enseignement secondaire, les réponses pédagogiques visent généralement à éviter le copier-coller et à souligner l'importance du déploiement de la culture informationnelle dans les cursus. Dans ce cadre, nos résultats sur la pratique du copier-coller pourraient fournir une nouvelle piste didactique. En effet, s'il convient de pas sous-estimer les opérations de copier-coller alors il est nécessaire que les médiateurs et tout particulièrement les professeurs-documentalistes ne les interdisent pas dans les consignes orales ou écrites données aux élèves. Ceci suppose de mieux distinguer deux tâches, rechercher de l'information et produire un document de synthèse, afin de dissocier plus fortement les consignes données pour chacune d'entre elles. Le copier-coller devrait être autorisé dans les tâches de formulation du problème informationnel et d'interactions avec les systèmes de recherche d'information. Ainsi le copier-coller nous paraît-il mettre à l'épreuve les conceptions des pédagogues à propos des activités informationnelles. Nous suivons Limberg et Sundin (2006) qui, à partir d'entretiens auprès d'enseignants et bibliothécaires suédois du secondaire, remarquent que ceux-ci s'attachent principalement à l'usage de l'information. En témoigne l'importance accordée aux habiletés en compréhension de texte, dans l'écriture correcte des références bibliographiques, l'évaluation critique de l'information et l'écriture d'un texte personnel. Les auteurs notent de la sorte des conceptions trop restreintes, fortement centrées sur la source, le document ou en d'autres termes sur la lecture et l'écriture.

Considérer le rôle du copier-coller est de nature à élargir les représentations enseignantes en offrant une meilleure visibilité du processus informationnel. Ceci pourrait contribuer à enrichir une didactique info-documentaire prenant appui sur ce que les élèves font.

Bibliographie

- Dervin B. « An overview of sense-making research : Concepts, methods, and results », Annual meeting of the International Communication Association, Dallas, 1983.
- Dervin B. « What methodology does to theory : Sense-making methodology as exemplar », In K. E. Fischer, S. Erdelez, L. McKechnie (dir.), *Theories of information behavior*. Medford : ASIS, 2005.
- Igo B. L., Bruning R., McCrudden M. T. « Exploring differences in students' copy-and-paste decision making and processing : A mixed-methods study », *Journal of Educational Psychology*, vol. 97, n° 1, 2005, p. 105-116.
- Kuhlthau, C.C.. « Inside the search process : Information seeking from the user's perspective », *Journal of the American Society for Information Science*, vol. 42, n°5, 1991, p.361-371.
- Large, A., Beheshti J. « The web as a classroom ressource : Reactions from the users », *Journal of the American Society for Information Science*, vol 51, n°12, 2000, p. 1069-1080.
- Limberg L., Sundin O. « Teaching information seeking : Relating information literacy education to theories of information behaviour », *Information research*, vol 12, n°1, oct. 2006. <<http://informationr.net/ir/12-1/paper280.html>>
- McGregor J. H., Williamson K. « Appropriate use of information at the secondary school level : Understanding and avoiding plagiarism », *Library and Information Science Research*, vol. 27, 2005, p.496-512.
- McGregor J. H., Streitenberger D. C. « Do scribes learn ? Copying and information use », *School Library Media Quaterly Online*, vol. 1, 1998. Reproduit in M.K. Chelton et C. Cool, *Youth information-seeking behavior : Theories, models and issues* (p. 95-118). Lanham : Scarecrow Press, 2004, p.95-118
- Pitts J. M. « Mental Models of Information: The 1993-94 AASL/Highsmith Research Award Study », *School Library Media Quaterly*, n° 23, 1995, p. 177- 184.
- Rouet, J.F., Tricot, A. Chercher de l'information dans un hypertexte : Vers un modèle des processus cognitifs. In A. Tricot & J.-F. Rouet (Eds.) *Hypertextes et Hypermédiats*. Paris : Hermès, 1998, p.57-74.
- Schamber, L., Eisenberg, M. B., Nilan, M. S. «A re-examination of relevance : Toward a dynamic, situational definition », *Information Processing and Management*, vol. 26, n° 6, 1990, p. 755-776.
- Tricot A. « Charge cognitive et apprentissage. Une présentation des travaux de John Sweller », *Revue de Psychologie de l'Éducation*, n°3, 1998, p. 37-64.

Bordereau de saisie - Références bibliographiques

Auteur(s)	
Auteur(s) du document hôte	
Titre	
Titre du document hôte	
Titre du périodique	
Type de support	
Numéro du périodique	
Edition	
Lieu de publication	
Editeur	
Date de publication	
Date de mise à jour	
Date de la référence	
Collation	
Collection	
Numéro de collection	
ISSN	
Disponibilité et accès	
ISBN	
Mots clés	

Résumé indicatif de 50 mots.

