

Outil hypergraphe d'exploration multimédia et pratiques de recherche-production

Sophie Barel, Florian Hémont, Florence Thiault

▶ **To cite this version:**

Sophie Barel, Florian Hémont, Florence Thiault. Outil hypergraphe d'exploration multimédia et pratiques de recherche-production. H2PTM'19 : De l'hypertexte aux humanités numériques, Oct 2019, Montbéliard, France. pp.03-21. hal-02324134

HAL Id: hal-02324134
https://hal.science/hal-02324134v1

Submitted on 21 Oct 2019

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

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

Outil hypergraphe d'exploration multimédia et pratiques de recherche-production

Sophie Barel* — Florian Hémont* — Florence Thiault*

* *Laboratoire PREFICS, Université Rennes 2*
Place du recteur Henri Le Moal, 35043 Rennes cedex

RÉSUMÉ. *Suite au développement de l'outil LIMAH, nous rendons compte ici des dynamiques de recherche-production qui se tissent avec celui-ci. Après avoir précisé le protocole d'observation et la méthodologie audiovisuelle utilisés nous présentons nos premiers résultats exploratoires. Nous aborderons alors la question de la construction de la crédibilité du Système de Recherche d'Information ainsi que le travail de recherche-production comme production de sens.*

ABSTRACT. *Following the development of LIMAH, we report here the research-production dynamics in which it participates. After having specified the analysis protocol and the audiovisual methodology used, we present our first exploratory results. We will then address the issue of building the credibility of information retrieval system as well as the work of research-production seen as a sense making process.*

MOTS-CLÉS : *hypergraphe, système de recherche d'information, production de sens, méthodologie audiovisuelle, exploration.*

KEYWORDS: *hypergraph, information retrieval system, sensemaking, audiovisual methodology, exploration.*

1. Introduction

Les professionnels de l'information (journalistes, *community managers*, bibliothécaires…) ont une activité professionnelle nécessitant une consultation rapide et efficace d'informations. Cette catégorie de professionnels s'appuie, dans leurs pratiques quotidiennes de recherche d'information, sur de grandes quantités de données, sous forme d'archives journalistiques, le plus souvent consultées par le biais de moteurs de recherche. Paradoxalement, ces professionnels utilisent sensiblement les mêmes types de services que le grand public.

Le projet dont est issu ce travail vise notamment à enrichir des collections multimédias et multisources d'actualités en créant des liens explicites entre documents. Cette recherche[1] est fondée sur une collaboration entre différentes entités et chercheurs de disciplines différentes notamment l'informatique, le Traitement Automatique des Langues, la psychologie expérimentale et les sciences de l'information et de la communication. Elle a abouti au développement d'un moteur d'exploration de corpus multimédia. Les chercheurs en sciences sociales s'intéressent plus particulièrement à la perception de l'utilité d'outils d'exploration d'actualités par des utilisateurs, ainsi qu'à la modification éventuelle des pratiques professionnelles liées à l'utilisation de ces nouveaux outils.

L'outil LIMAH[2] propose une manière alternative d'organiser des contenus, d'établir des liens entre eux, de manière à rendre leur exploration plus facile, sous forme d'une visualisation cartographique. La mise en page des contenus est simplifiée et normalisée. Il s'agit de comprendre comment ce dispositif intervient et participe à l'organisation d'une recherche à partir de médias de natures différentes. Il est composé d'articles de presse, d'émissions télévisuelles et radiophoniques. Cette contribution vise à présenter le processus de la phase de tests avec des utilisateurs ciblés, les analyses et recherches menées (de juin 2017 à février 2018). Est interrogée plus particulièrement la manière dont est appréhendé le logiciel LIMAH auprès du public qui est susceptible de l'utiliser et la façon dont ce dernier peut l'inscrire dans sa pratique professionnelle.

2. Présentation du Système de Recherche d'Information[3]

L'outil ici développé et testé propose de tisser des liens entre documents ou extraits de documents de nature variée : texte, audio, vidéo, voire des tweets. Les

---

[1]. Cette recherche est issue d'une collaboration entre différentes entités et chercheurs de disciplines différentes : TEXMEX (INRIA), LINA/TALN (Université de Nantes), IODE (Telecom Bretagne), CRPCC et PREFICS (Université Rennes 2) : http://limah.irisa.fr/. Ce travail a été financé par le laboratoire d'excellence CominLabs et l'ANR10-LABX-07-01.

[2]. LIMAH: Linking Media in Acceptable Hypergraphs. Pour la description du fonctionnement schématique du traitement des médias voir figure 1 (http://limahweb.irisa.fr/texmix/#/).

[3]. Désormais SRI.

pistes audio des vidéos ou émissions de radios sont également retranscrites à l'écrit. Dans notre étude de cas, nous avons travaillé à partir d'une base de données constituée de deux mois de recueil systématique de sources médiatiques (mai-juin 2015)[4], soit au total 4966 articles de presse, 1556 flux radio et 290 vidéos[5].

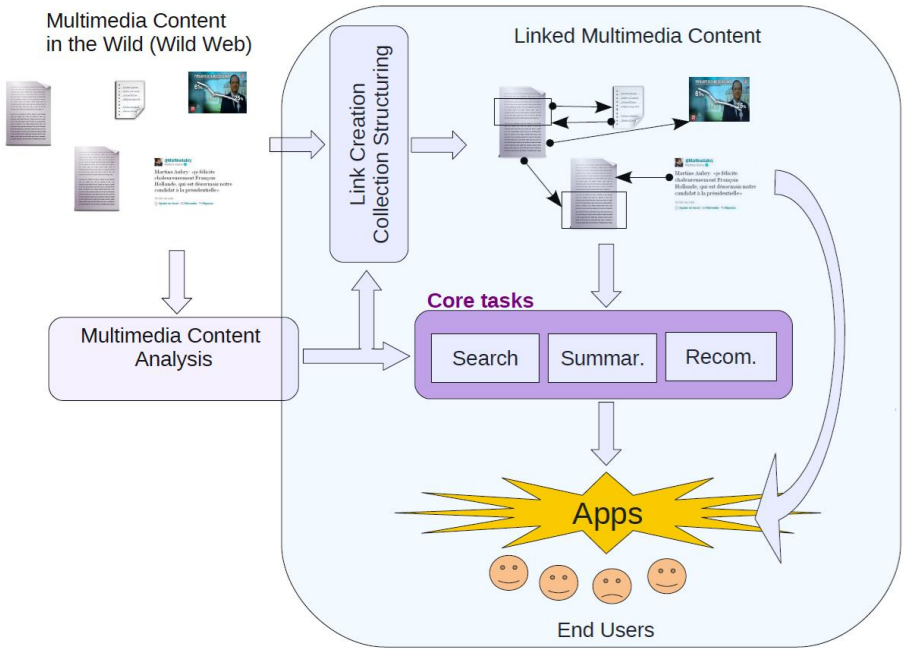


Figure 1. *Fonctionnement schématique du SRI LIMAH*

L'interface graphique se compose de deux parties principales : une interface de recherche par mots-clés[6] (Figure 2) et une interface de navigation permettant de passer d'un document à l'autre sans revenir à une recherche par mots-clés (Figure 3). En saisissant un ou plusieurs mots-clés, l'utilisateur se voit proposer une liste de documents répondant à sa requête. En sélectionnant l'un des documents qui lui sont soumis, l'utilisateur arrive sur une page présentant le document dans sa totalité, ainsi que divers métadonnées et moyens de navigation.

---

4. *C'dans l'air, Le Figaro, Libération, The Huffington Post, France 2, France 5, C à vous, France Inter, RMC, Le Point, Le Monde.fr*. Notons également que l'intégration de tweets a débuté mais n'était pas aboutie lors de nos tests.

5. Le corpus de documents collectés est présenté dans la publication suivante : http://limah.irisa.fr/docs/29_06_2016corpus2016.pdf

6. Qui spécifie notamment par item le nombre de documents lié par le SRI.

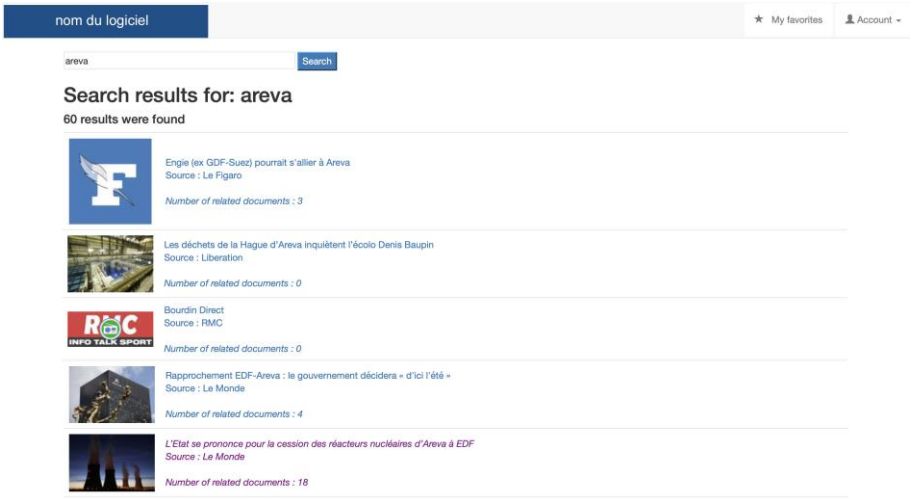


Figure 2. *Résultat d'une recherche par mots-clés*

Quand l'utilisateur clique sur un article, il voir apparaître le texte de celui-ci, ainsi qu'une carte des documents liés par le SRI. Les liens sont de différentes natures : *antérieur*, *postérieur* (il s'agit d'un document publié sur le même thème avant ou après celui sélectionné), *duplication* (le texte est repris sans pour autant ajouter d'information), *résumé* ou *développement*, *en réaction à* ou *qui réagit à*[7].

[7]. Les publications de Rémi Bois et al. (2015, 2017) et de Anca-Roxana Simon et al. (2015) relatent la manière dont les liens entre les différents documents sont opérés par le système.

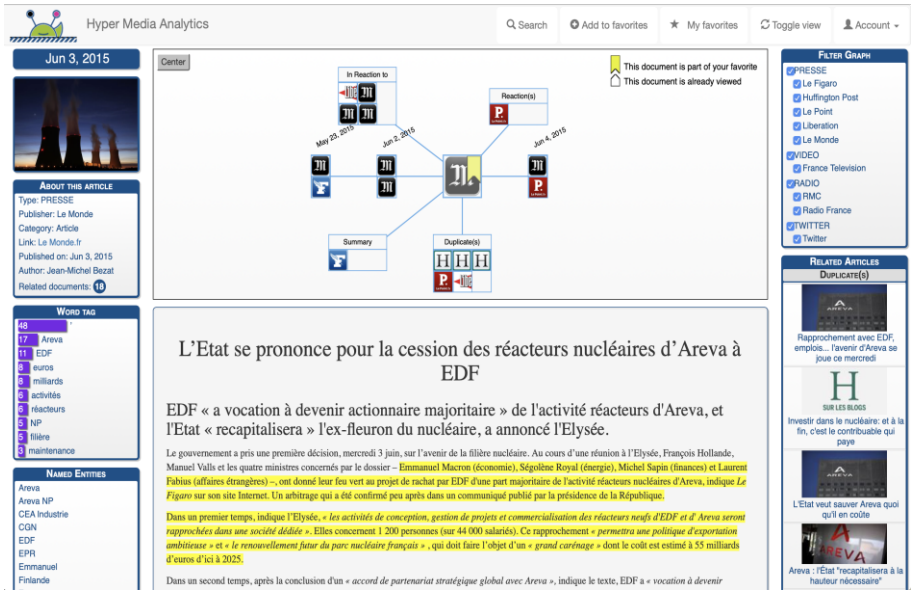


Figure 3. *Interface du moteur d'exploration*

L'utilisateur a également la possibilité de retenir sous forme de favoris des éléments, ainsi que de surligner des parties du document. Lorsqu'il consulte un document il voit apparaître des tags qui lient le document principal à celui survolé. Une partie des métadonnées est également mise en exergue, et l'utilisateur peut enfin sélectionner/désélectionner les sources de documents.

3. Analyser les dynamiques de l'activité

L'interrogation principale que nous avons est de comprendre comment les testeurs peuvent intégrer l'outil LIMAH à leurs pratiques professionnelles. Ainsi l'objectif de notre étude se situe moins dans la quantification de la performance des utilisateurs ou du système, que dans l'observation de la manière dont ils s'approprient le logiciel et comment ce dernier participe aux modifications de leurs pratiques de recherche et de production d'information.

3.1. *De l'analyse du SRI à l'analyse des pratiques d'un SRI*

« Harter (1996) utilise le terme de « schisme » pour désigner la coupure qui existe selon lui entre les chercheurs qui utilisent des méthodes de laboratoire pour évaluer la performance des SRI et ceux qui prennent en compte le comportement réel des usagers

dans l'évaluation des mêmes systèmes » (Chaudiron et al., 2002, p. 213). Dans le cadre de cette étude, nous nous intéressons plus particulièrement aux pratiques des utilisateurs dans un contexte d'évaluation des SRI.

Au-delà de l'étude de l'appropriation de l'outil, nous avons voulu observer son intégration dans un contexte de travail plus général ainsi que la façon dont le logiciel s'insère dans cet environnement. Selon Marcia Bates (1989) et David Ellis (1989) « Les individus emploient une multiplicité de techniques de recherche. Ils ne cherchent pas seulement en formulant des requêtes par mots-clés sur le sujet qui les intéresse, ils exploitent aussi les éléments qu'ils trouvent au fur et à mesure de leurs recherches. Ils suivent par exemple des références bibliographiques qui figurent dans certains documents consultés, ou bien les citations dont tel document a fait l'objet (chaînage), se focalisent sur un auteur particulier ou sur une revue qui traite du sujet qui les intéresse (forme de veille) et s'abonnent aux flux RSS de ces sites ». (Simmonot, 2012, p. 96). La conception de l'accès à l'information centrée sur un fonctionnement des applications (paradigme système) imposant au chercheur d'information un modèle rationnel et stable est contesté par les approches cognitives. En réalité, les usagers mettent en œuvre un agir informationnel complexe, « une multitude de tactiques que les seules requêtes qu'ils formulent ne permettent pas forcément de percevoir ou de comprendre. Ils peuvent aussi recourir à plusieurs sources simultanément. Si l'on étudie leurs comportements à partir d'une seule source de recueils de données, comme c'est souvent le cas dans les études centrées « systèmes », on ne peut pas percevoir l'ensemble ni la cohérence de leur activité » (*idem*.).

Dans cette perspective, nous considérons que l'activité de « recherche-production » est distribuée dans un ensemble d'outils comme des SRI, mais aussi des supports de prises de notes, de schématisation, qui bien souvent débordent d'un simple écran. Par ailleurs, cette activité complexe de recherche n'est en général pas autonome et entre en écho avec d'autres tâches faisant partie de la même activité. Aussi, nous proposons ici de recourir à l'expression « recherche-production » dans la mesure où les deux sont intriquées autour d'une même finalité. Recherche et production s'informent[8] mutuellement et composent le recto-verso d'une même activité.

Carol Kuhlthau (1991) a mis en avant un ensemble d'étapes constitutives des démarches de recherche-production : l'initiation, la sélection d'un thème, l'identification, l'exploration, la formulation, la collection, la présentation. Toutefois, l'activité de recherche-production se révèle bien moins séquentielle et contient davantage d'itérations et d'imbrications lors des différentes actions. Ainsi, la démarche de recherche-action pourrait s'étudier en terme de « *sensemaking* » selon l'approche de Karl E. Weick (2005). L'idée proposée est que le « (…) *sensemaking can be treated as reciprocal exchanges between actors (Enactment) and their environments (Ecological Change) that are made meaningful (Selection) and preserved (Retention)* » (2005, p. 414). L'activité de l'utilisateur est alors aussi bien

8. Au sens de mise en forme.

constitutive du contexte que ce contexte met en forme l'activité et le sens que l'acteur construit. Si on considère que le processus de production de sens est intimement associé à un processus d'*enactment,* alors la recherche (en tant que processus d'exploration complexe, itérative, soumise à de la sérendipité, etc.) est reliée à la production des utilisateurs.

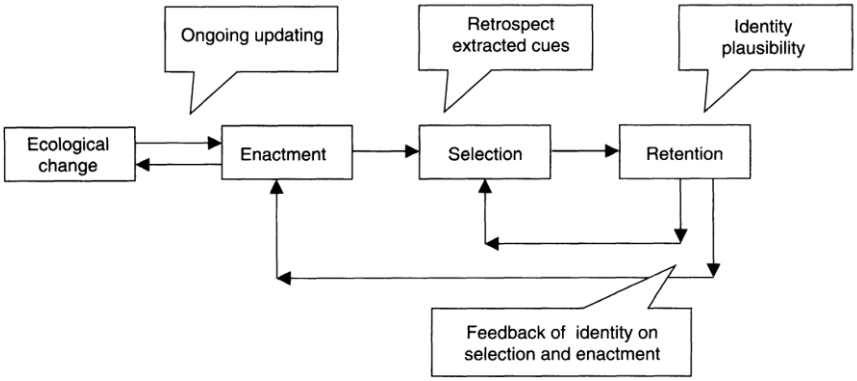


Figure 4. *« The Relationship Among Enactment, Organizing, and Sensemaking » (Weick* et al.*, 2005, p. 414)*

Nos interrogations portent sur la manière dont les utilisateurs construisent de façon dynamique du sens durant leur activité de recherche-production. Dans cette approche processuelle, nous envisageons cette activité telle une co-construction à laquelle participe un ensemble d'acteurs, de documents, d'outils techniques, etc. Il s'agit alors de tenter d'appréhender la manière dont le SRI s'intègre, dans des pratiques, à une écologie des artefacts (Hutchins et Klausen, 1996). Par exemple, la sélection-rétention est une action qui passe par le fait d'identifier des informations et d'en déposer une partie par copier-coller numérique ou par agrégation d'informations transcrites sur un papier brouillon (qui peut aussi bien être des données chiffrées que des schématisations).

Enfin, nous suivons McKenzie lorsqu'il invite à porter notre regard « (...) *on practices rather than on behavior shifts the analysis from cognitive to social and is consistent with the study of information seekers within their social context* » (2003, p. 24). Ainsi, demander à un utilisateur de rechercher tel ou tel élément à partir d'un SRI afin d'identifier soit l'efficacité du système soit la performance de l'utilisateur n'a pas la même visée que de lui demander d'utiliser un outil dans des conditions proches de celles de son activité (en termes de contraintes temporelles, de types de production et de recherche).

3.2. *Protocole de l'étude*

Le protocole d'observation élaboré se compose de trois phases. Les deux premières phases consistent en une analyse des différences entre les pratiques développées autour de l'outil LIMAH dans l'exploration des sources et celles déployées autour des moteurs de recherche habituellement utilisés par le testeur. La phase 3 correspond à l'observation et l'analyse des pratiques de recherche et production d'écrits de testeurs qui peuvent recourir comme ils l'entendent au logiciel développé et aux moteurs de recherche qu'ils utilisent d'ordinaire (l'idée étant d'appréhender la manière dont l'outil LIMAH s'insère en pratique dans une écologie de dispositifs techniques). Au préalable les testeurs ont visionné deux courtes vidéos tutoriels de découverte du logiciel. La première présente les fonctions de base de l'outil, la seconde explique les fonctions avancées (surlignage, gestion des favoris, etc.). Ensuite, les utilisateurs réalisent une tâche consistant à produire un travail sur un thème d'actualité défini. Une fois le test terminé, un retour d'expérience des utilisateurs est organisé.

Une méthodologie de captation audiovisuelle a été mise en œuvre afin de saisir l'activité sur écran mais également l'activité hors-champ de prises de notes. Cette méthode de captation est basée sur le composite méthodologique suivant : enregistrement de l'écran, duplication de l'écran pour l'observateur (poste d'observation 1), enregistrement vidéo (poste d'observation 2) du participant qui permet de filmer son activité hors écran (Figures 5 et 6).

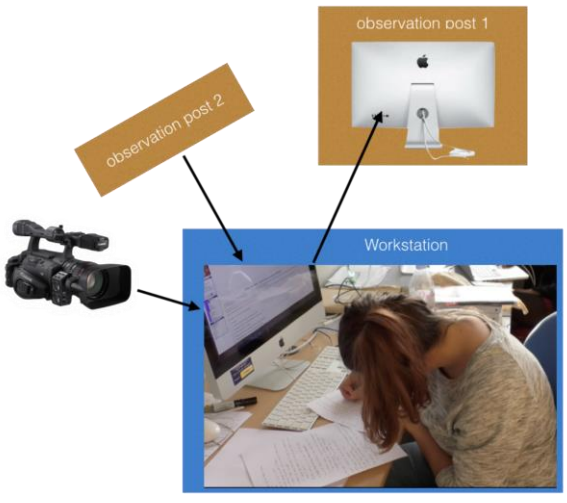


Figure 5. *Schéma du dispositif de captation*

	
Poste d'observation 1. Une caméra enregistre l'activité de l'utilisateur en train de conduire sa recherche-production. Nous pouvons observer comment s'opèrent les dynamiques entre le travail sur brouillon et l'élaboration de la recherche-production.	**Poste d'observation 2. Une duplication de l'écran de l'utilisateur permet de suivre en temps réel son processus de recherche.**

Figure 6. *Description du dispositif de captation*

Il s'agissait d'observer les stratégies de construction d'information. Une duplication d'écran permet de suivre en temps réel la façon dont l'utilisateur conduit son travail. Une caméra filme le testeur en train de réaliser sa recherche, afin d'examiner s'il prend des notes sur un papier brouillon et ainsi mieux percevoir la construction de sa recherche-production.

L'observation des pratiques ne saurait suffire pour comprendre le cheminement de recherche-production des utilisateurs. Ainsi, à l'issue de la production des participants, des entretiens compréhensifs ont été réalisés (Kaufmann, 2011)[9]. Ils portaient sur la construction de la production (les manières de problématiser, de s'approprier le sujet, de choisir un axe), les stratégies de recherche (démarche déductive, démarche inductive, etc.), les formes de recherche, la capacité à identifier des sources, la hiérarchisation de l'information lors de la construction de leurs écrits, les fonctionnalités utilisées, les modalités de présentation des résultats, la place projetée que le SRI pourrait avoir dans leur activité professionnelle.

Neuf professionnels ont été sélectionnés suivant un critère principal à savoir que la recherche informationnelle fait partie intégrante de leur activité : soit trois

9. Notons que les enregistrements vidéo ont également pu servir à reproduire lorsque cela était nécessaire une situation d'auto-confrontation (Theureau, 2010, Cahour et al., 2010).

journalistes, un bibliothécaire, un *community manager*/attaché de presse, un journaliste/webmaster/*community manager*, un chargé de projet numérique/formateur en bibliothèque, un agent de documentation, un doctorant en SHS. Il n'y a pas de catégorie de genre ou d'âge spécifique retenue pour constituer ce panel.

Les thématiques de recherche des tests ont été choisies en fonction des sujets pour lesquels il existait un certain nombre de documents relatifs disponibles dans la base de données. Au final, pour les deux premières phases d'exploration, les participants ont travaillé sur « l'eurovision 2015 » à partir de LIMAH, et « l'eurovision 2016 » leurs moteurs de recherche usuels. Certains débutaient par l'Eurovision 2015 et d'autres par l'Eurovision 2016 (afin de diminuer le biais de l'apprentissage et de l'habitude de traitement de ce type d'évènement). Durant la troisième phase de test, ils ont réalisé une exploration sur le sujet « Solar impulse d'Areva ». Les thématiques étaient énoncées de manière évasive afin de ne pas induire de traitement particulier du sujet dans la mesure où l'un de nos centres d'intérêts portait sur les cheminements des utilisateurs durant leur recherche-production. Il s'agissait moins ici de donner des contraintes précises aux participants de l'étude que de se rapprocher de ce qu'ils produisent dans leurs pratiques professionnelles quotidiennes (et ce, plus particulièrement lors de la 3ème phase). Ils étaient libres de choisir leur angle d'écriture et la forme du rendu final. Ainsi certains ont produit des brèves, des chroniques journalistiques, des résumés, d'autres des cartes heuristiques ou même des « *sketchnotes* » (prise de notes graphiques et visuelles), des communiqués et revues de presse, voire une réponse courte.

4. Premiers résultats

4.1. *La crédibilité : vecteur d'appropriation*

Le filtrage et la sélection des sources et des médias qui constituent la base de données posent la question de la crédibilité donnée à celles-ci. En effet, selon Fogg et Tseng (1999), il existe quatre sortes de crédibilité. La crédibilité présumée repose sur les stéréotypes, les représentations préalables et les hypothèses des usages vis-à-vis d'une source, d'une personne, d'un site web, etc. La crédibilité réputée, est liée quant à elle à l'affiliation d'un auteur ou d'un document à une institution, une source, ou un média reconnus, à la notoriété, etc... La crédibilité de surface, ou crédibilité « apparente », correspond à la représentation de la ressource ou de la personne. Et enfin, la crédibilité d'expérience, ou crédibilité éprouvée, est fondée sur l'expérience antérieure, essentielle dans tout jugement. Comme l'écrit Alexandre Serres : « Ainsi la crédibilité « réputée » correspond-elle à la source (éditeur, organisation, institution, auteur...), tandis que la crédibilité « de surface » concerne tout ce que nous appelons « dispositif » ; enfin, les crédibilités « présumées » et « d'expérience » renvoient davantage à l'usager lui-même, à ses représentations et à son expérience personnelle » (Serres, 2012, p. 123).

La question de la pertinence est soulignée dans les travaux qui portent sur l'évaluation des SRI. Dans ces études, « il est présupposé que le système suffit pour trouver les documents pertinents, c'est-à-dire que la pertinence est inhérente à la requête ; il n'est aucunement fait mention d'un quelconque besoin d'information de l'usager » (Chaudiron et Ihadjadène, 2002, p. 215). Les calculs de pertinence se fondent sur l'hypothèse selon laquelle le nombre de documents pertinents pourrait être déterminé de façon objective alors qu'il relève bien souvent de la subjectivité des utilisateurs (idem, p. 216). Ainsi, nous pensons que la pertinence s'évalue à l'aune des pratiques des utilisateurs et relève de leur expérience. Au-delà de la crédibilité des couples sources-médias, nous proposons ici de nous intéresser plus particulièrement à la manière dont les utilisateurs construisent un régime de crédibilité du SRI lui-même.

4.1.1. *Pertinence : construction de la crédibilité du SRI*

Pour la plupart, les utilisateurs ont testé la pertinence de l'outil, que ce soit par mots-clés ou en remontant le fil des pages, pour en comprendre le fonctionnement. Cette pertinence est très discutée. *« Il sort des articles, on sait pas trop pourquoi, alors que le terme Eurovision n'apparaît même pas dedans. La pertinence est étrange ». « Y'a un article avec des réacteurs nucléaires au Japon… C'est quoi le lien ? (...) C'est parce que j'ai ouvert un article du Monde et donc ça me ressort cet article du Monde là, même si ça n'a rien à voir ? », « Y'a des contenus sans rapport qui ressortent : sur la Grèce, etc… C'est parce qu'il y a le mot « euro » peut-être ? »* (User 33, journaliste, phase 1).

Ainsi, la pertinence développée dans le SRI peut parfois ne pas rencontrer la représentation des besoins informationnels que se construit l'utilisateur. Les testeurs essayaient d'ailleurs de comprendre le modèle de pertinence du système sans pour autant y parvenir. « *On ne sait pas si ça s'affiche par date, par lien... On ne connaît pas trop la pertinence d'apparition des liens, qu'est-ce qui fait qu'un lien s'affiche en premier ou non* » (User 36, bibliothécaire, phase 1). Il est probable que dans le cadre d'une base de donnée plus ample, ce problème serait moins sensible.

Durant leurs recherches, les utilisateurs ont tenté différents types de requêtes. Ils construisaient petit à petit leur compréhension du fonctionnement du dispositif via des requêtes simples (un mot) ou des requêtes plus développées (au moins 2 mots), voire contenant des opérateurs booléens. *« Il ne fait pas un très bon tri quand même… Quand tu mets plusieurs mots ensemble, je n'ai pas l'impression que ça analyse très bien pour rendre un contenu pertinent. » « Quand je tape des mots, j'ai l'impression qu'il Il n'y a pas une analyse des documents qui ressortent : on a Eurovision 2015, ok, mais dès qu'on rajoute un mot, y'a plus rien de précis.* » (User 38, doctorante, phase 2). « *Le - marche mais le + ne marche pas pour affiner la recherche* » (User 33, journaliste, phase 3).

Plusieurs utilisateurs commençaient par un « test » de l'outil avant de démarrer leurs recherches, afin de pouvoir moduler leurs mots-clés. Ce processus de découverte

des fonctions d'interrogation constitue une phase d'apprentissage du fonctionnement interne du SRI afin d'adapter leurs stratégies par la suite. Les utilisateurs se constituent petit à petit un modèle de pertinence d'usage. « *C'est une recherche sémantique. Ce qui a été un défaut pour moi y'a deux semaines*[10]*, je m'en suis finalement servie pour la fois d'après* » (User 32, webmaster/journaliste, phase 3).

Cependant, la mise en page standardisée est appréciée car elle laisse de côté les publicités, qui ralentissent et rendent floue la compréhension des contenus en ligne, et permet de faire gagner du temps en adaptation lorsque l'on passe d'un média à un autre. Cette harmonisation permet de se focaliser davantage sur le contenu que sur la forme. « *Où que t'ailles, tu n'as pas ces deux secondes d'adaptation vu que les mises en page sont différentes d'un site à un autre, ici c'est standardisé. Là on se n'attarde pas sur des choses pas importantes, genre la publicité* » (User 36, bibliothécaire, phase 1).

En fonction du choix des médias consultés au cours de leur recherche-production, la crédibilité n'est pas perçue de façon identique par les utilisateurs : « *Il y a tout un process pour évaluer une information. Avec le design, le vocabulaire, l'architecture de l'information, ... La crédibilité est quelque chose de complètement relatif* » (User 39, chargé de projet numérique/formateur, phase 1).

En tant que professionnels de l'information ou de la recherche, en s'appuyant sur une crédibilité d'expérience, ils sont habitués à faire le tri dans les contenus, mais apprécient que LIMAH réalise déjà une pré-sélection médiatique, ce qui leur a permis de « gagner du temps ». Les utilisateurs ont perçu le logiciel comme étant « d'apparence sérieuse » en voyant les sources mise à disposition, même s'ils pouvaient être en désaccord idéologique avec certaines. Le filtrage par source n'a été utilisé par aucun de ces testeurs, même par curiosité. Cette possibilité posait apparemment un problème déontologique aux journalistes, qui ne voulait pas manquer d'objectivité. Ce qui ne les empêchait pas d'être conscient des potentielles orientations politiques de certaines des sources journalistiques en les lisant tout de même. Ils jaugent le taux de crédibilité ou de confiance qu'ils accordent aux contenus pendant la recherche/production. « *J'ai privilégié les articles du Monde, j'ai un a priori qui fait que pour moi, Le Monde a moins de jugements de valeur, que Le Point ou Le Figaro* » … « *Si je filtre ce n'est pas par sources, mais par thématiques, ça me dérangerait d'enlever des sources, tous les médias sont biaisés, même si j'ai trouvé* Le Point *et L*e Figaro *très orienté en critiquant les socialistes »* (User 36, bibliothécaire, phase 3).

Nous avons sélectionné des testeurs dont la profession implique déjà une capacité à sélectionner et traiter des informations en ligne. Lorsqu'on les interroge sur les sources disponibles dans LIMAH, ils portent un regard critique sur elles, conscients notamment du contexte global sociopolitique des médias ou de l'environnement sociotechnique des documents (Serres, 2012, p. 10-11). Lors des entretiens, les questions au sujet de la pertinence d'une information, de l'autorité ou du contexte

[10]. Lors de la phase 1 ou 2.

sociopolitique des médias ont été soulevées par chaque utilisateur, que ce soit avec Google ou avec LIMAH. « *En fonction du média, ça va retenir l'opinion que j'ai sur lui. Le point est un peu biaisé politiquement, je garde un "point de vue d'opinion"*. » (User 36, bibliothécaire, phase 1). Le travail de filtrage est déjà, en quelque sorte, réalisé dans l'outil même. LIMAH n'a pas aspiré le même contenu que Google par exemple. Une pré-sélection de sources journalistiques a été effectuée pour ensuite réaliser l'aspiration des contenus, par thématique. Ce qui constitue déjà un cadrage en soi.

« J'ai lu des articles sur LIMAH, que j'ai utilisé en premier, pendant une demi-heure à peu près, mais je lis lentement. (...) J'ai utilisé LIMAH pour ne pas me faire parasiter par d'autre info, je voulais vraiment savoir sur une période donnée, et vraiment sans que des articles d'une autre période puissent apparaître, plus comme des archives » (User 38, doctorante en Sciences Humaines et sociales, phase 3). Le fait de réduire la masse d'information permet également de gagner du temps et de limiter l'impression d'infobésité.

4.1.2. *Entre gain de temps de travail et identification des sources.*

Dans leurs pratiques professionnelles, les testeurs sont amenés à se poser la question du temps qu'ils vont mettre pour réaliser une recherche-production. Précisons tout d'abord que le temps de l'activité n'était volontairement pas encadré afin que les utilisateurs ne soient pas dans l'urgence pour appréhender le dispositif et qu'ils puissent définir eux-mêmes des contraintes en cohérence avec leur activité professionnelle. « *Il y a une contrainte de temps au travail, du coup là j'en avais une »* (…) « *Il faut qu'en 15 minutes, j'ai le temps de balayer un sujet, ce qui peut être pertinent pour un étudiant par exemple. On est obligé de se mettre des barrières, de se baliser »* (User 36, bibliothécaire, phase 3). Si bien que plusieurs des testeurs se sont donnés un temps limité en début d'enquête à ne pas dépasser. Allant de 30 secondes au minimum pour le chargé de projet numérique à 2h pour un journaliste. Dans le cadre de cette étude, l'amplitude des recherche-production s'étend de 10min à 1h40.

Sur ce point, nous constatons une différenciation des durées de l'activité selon les outils utilisés. Nous avons calculé que le temps de travail sur LIMAH s'est avéré considérablement réduit par rapport à celui observé sur Google ou tout autre moteur de recherche choisi par le testeur. Ainsi, en moyenne sur LIMAH, les utilisateurs ont travaillé 30min, contre 55min sur les autres moteurs de recherche. Soit des durées quasiment divisées par deux. Plusieurs éléments peuvent expliquer l'origine de cette différence : le traitement proposé par l'outil et la limitation de la base de données.

Malgré les développements récents des moteurs de recherche, la question de l'infobésité informationnelle reste toujours d'actualité. « *Je me sens souvent perdue dans la masse d'informations disponible* » (User 37, *community manager*, phase 3) ; « *Avec internet il y a plus de moyen de se perdre, même si internet permet plein de choses* » (User 38, doctorante en Sciences Humaines et Sociales, phase 1). « *Je me*

suis fixé de comprendre l'historique de l'entreprise, et là je m'arrête. Pour ne pas perdre trop de temps, ou me perdre tout court » (*idem*, phase 3). Ces commentaires ont d'intéressant qu'ils indiquent que les utilisateurs rencontrent souvent un problème d'infobésité face à des ressources multiples. Dans le cas d'une phase de recherche durant laquelle un participant utilisait Google, nous avons d'ailleurs constaté son abandon suite à la confrontation à de multiples sources qui l'a conduit dans un état de confusion, après une lecture d'une trentaine de minutes.

Le choix d'une mise en page unique pour l'ensemble des contenus joue également un rôle. « *Il y a de la perte dans la mise en page des différents sites, puisque la mise en page est harmonisée, mais du coup on gagne du temps : en temps d'adaptation lorsque l'on passe d'un site à un autre, lorsqu'on a compris le système nous savons où trouver une information précise (date de publication, photo de Une, etc.), pas de publicité* » (User 36, bibliothécaire, phase 3).

Cependant cette mise en page unique peut troubler certains marqueurs usuels d'identification des institutions médiatiques : « *J'ai mis beaucoup de temps à comprendre que c'était le Huffington Post, la mise en page ne m'était pas familière* ». (User 32, webmaster, *community manager*, phase 1). Ainsi peut se trouver en tension d'un côté une forme de crédibilité allouée à l'outil pour ses capacités à réduire le champ des sources, et de l'autre, la crédibilité liée à l'expérience des sources qui sont identifiées en partie par l'intermédiaire de chartes graphiques.

4.2. *La recherche-production : l'exploration comme production de sens*

S'il s'agit moins ici d'un SRI au sens propre que d'un moteur d'exploration, notre approche vise à saisir les dynamiques du processus d'exploration.

4.2.1. *Les stratégies de recherche avec un outil d'exploration*

Lors des observations de la 3ème phase, le séquençage des pratiques a facilité la détermination des moments de problématisation, du temps consacré, du processus de recherche d'information, des aller-retour, des doutes constitutifs de la recherche-production. Nous avons identifié à la suite de Brigitte Simonnot (2012, p. 85-86) plusieurs étapes de la démarche documentaire : l'étape d'initialisation, de sélection, d'exploration, de formulation, de collecte, de présentation. Cette classification nous a permis de constituer des repères d'observation sur le cheminement des utilisateurs. Toutefois, si le processus est présenté ici de manière linéaire, nos observations donnent à voir qu'il ne l'est pas nécessairement. User 38 nous explique : « *Je mets plus de temps à trouver un angle qui m'intéresse ou à formuler un angle de recherche, même si j'ai des intuitions, que de faire la recherche derrière ou d'entamer la rédaction* ».

L'utilisateur peut passer d'une étape à une autre, revenir à une précédente, faire une « tentative d'évitement » (soit adapter son sujet à ce qu'il trouve et non pas

l'inverse), voire abandonner. La collecte peut aussi donner lieu à une sélection, ou inversement. Le sujet peut revenir sur une information pour la vérifier, faire du tri dans ce qu'il a récolté pour ne garder que ce qu'il juge nécessaire à la réalisation de sa production finale.

« *Tu prends des trucs, tu capitalises, pour la suite*. » « *Dans ma veille, je tague des choses sur Google Keep pour m'en servir plus tard (...) j'éparpille mes notes un peu partout*. » « *Je marche beaucoup par mot-clé, par noms... Je construis par personne, par personnage, dans une histoire, pour trouver qui a un intérêt dans l'histoire justement.* » « *Je suis à la recherche de la friction, des points de friction.* » (User 32, webmaster/*community manager*/formation journalistique, phase 3)

A l'instar de David Ellis (1989), ces observations confirment que le processus de recherche-production se compose de phases différenciées. L'auteur identifie deux types de démarches principales : les démarches de recherche inductives (qui s'appuient sur les données pour construire une orientation du sujet) et déductives (fondées principalement sur les connaissances antérieures de l'individu et la manière dont elles sont structurées pour y intégrer les données). Ces démarches s'entremêlent. Ainsi, les utilisateurs oscillent entre des explorations durant lesquelles les « équations de recherche » évoluent au fil de la recherche. Il s'agit de moments d'extraction d'éléments-clés (phase d'identification sélective de l'information pertinente dans une source d'information), de mots autour desquels l'équation de recherche se recompose. « *Vu les titres, je comprends qu'il y a une histoire financière* » (User 36, bibliothécaire, phase 3), « *je regarde les titres du sommaire pour donner une « ambiance », une tendance* » (User 38, doctorante en Sciences Humaines et Sociales, phase 1). Ces moments d'exploration sont complétés par la collecte d'éléments spécifiques (date de naissance ou origine d'une personne par exemple), de vérification, d'approfondissement du sujet.

Dans le cadre prescrit, les utilisateurs débutaient en testant la collection et le moteur de recherche. Ils éprouvaient les mots-clés et la pertinence des résultats fournis par le système, par l'intermédiaire de requêtes simples comme « eurovision 2015 », ou de requêtes plus complexes contenant des opérateurs booléens (guillemets d'expression, +, etc. comme « eurovision+palmarès »).

D'une manière générale, ce SRI relève davantage d'un outil d'exploration que de recherche d'éléments spécifiques. En l'état, ce n'est pas un outil que les utilisateurs arrivent à utiliser dans le cadre de recherche d'éléments précis. Par exemple, le testeur qui est chargé de projet numérique utilise habituellement les moteurs de recherche (ou de réponse) avec des questions précises (type : « Quel est le palmarès de l'Eurovision 2015 ? ») : « *Je cherche une information précise et si je ne la trouve pas sur un outil, je change l'outil* ». Il se donne alors « *30 secondes ou de 2 à 4 minutes maximum* » pour trouver les réponses à ses questions. Il ne développe pas d'attitude d'exploration. Si bien que le SRI proposé ne convient pas à ses pratiques professionnelles. Il explique alors qu'il s'agit d'un « *outil pas pour tout le monde, pas pour toutes les formes de productions* ». Un autre utilisateur nous confirme que pour lui, « *c'est un moteur*

d'exploration et non pas un moteur de recherche » (User 34, ingénieur informaticien, phase 1). Il est également à noter que durant la phase 3, une testeuse journaliste, n'a pas du tout utilisé l'outil car elle nous précise : « *Je savais déjà ce que je voulais faire* ». Ce qui confirme à nouveau le particularisme de l'outil testé en tant qu'assistant d'exploration.

4.2.2. *Les brouillons comme objet intermédiaire*

A la disposition des testeurs étaient proposés du papier et des stylos, pour qu'ils puissent prendre des notes durant leurs recherches. Tous les participants ont utilisé le papier fourni, même si certains usages se sont avérés très limités comme par exemple l'inscription du simple sujet de leur enquête (Eurovision 2015 ou 2016, Areva ou Solar Impulse). D'autres ont constitué des listes, noté des chiffres et des noms. Ils élaborent un agrégat d'informations au fur et à mesure de l'exploration, sous forme de cueillette d'information ou « *Berry-picking* » (Bates, 1989). Certains ont réalisé une représentation graphique de leur enquête (des cartes heuristiques ou des arbres perlés par exemple, voir figure 7).

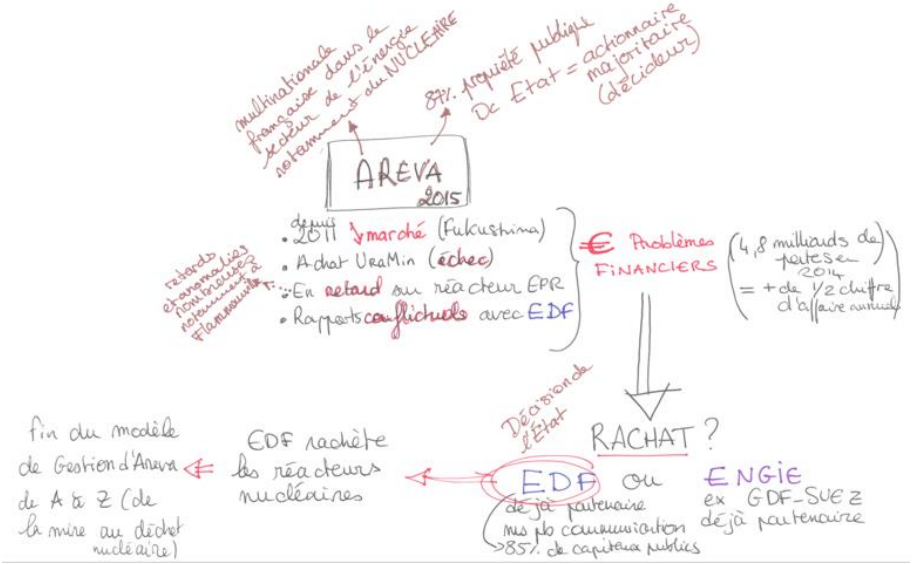


Figure 7*. Brouillon d'un testeur*

Les utilisateurs ont déjà dans leurs pratiques habituelles des manières de stocker les informations, que ça soit sur papier ou avec des outils en ligne. « *Dans ma veille, je tague des choses sur Google Keep pour m'en servir plus tard (…) j'éparpille mes notes un peu partout* » (User 32, webmaster/journaliste, phase 3).

Cependant aucun des testeurs n'a eu recours au module de favoris mis en place dans le système, sans doute car l'utilisation de celui-ci (un listing de documents surlignés) suppose une recherche au long cours là où les contraintes du projet ne permettaient pas ce type d'exercice. Les participants recouraient à des brouillons sur lesquels les différents éléments consignés étaient visibles en même temps. « *Ça serait vraiment bien de pouvoir centraliser toutes les phrases surlignées dans un seul document, pour pouvoir toutes les retrouver et toutes ensemble. À un moment, il y en a une, je savais que je l'avais surlignée mais je l'ai pas retrouvée même en cliquant sur plusieurs liens* » (User 41, journaliste et formateur, phase 1). Une liste d'éléments à la place ou en complément de la liste de documents faciliterait probablement le passage de ces listes d'éléments identifiés à une mise en récit nécessaire lors de la production écrite des utilisateurs. Notons d'ailleurs que si les utilisateurs n'exploitaient pas la liste de favoris, la stratégie des acteurs consistant à ouvrir/fermer/conserver des onglets s'y apparente largement. Une utilisatrice écrivait ses notes sur un brouillon papier, puis rédigeait toute sa production « au propre » sur une autre feuille encore. Elle nous a expliqué qu'elle avait besoin de garder une « trace » de ses erreurs, pour ne pas les refaire, au lieu de simplement les effacer sur un document texte numérisé. D'autres, en reprenant l'idée de « *Berry-picking* », reportent dans leur traitement de texte, l'ensemble des éléments qu'ils sélectionnent. Puis, ils les effacent, les complètent, les reformulent en fonction de la trame qui se construit. Le brouillon est dans ces cas également le lieu de la production finale. Il est métamorphosé petit à petit d'agrégat d'éléments épars en une production. Dans le cas d'une base de données plus étendue, les éléments retenus, pourraient également alimenter une fonction de recherche hypergraphe basée sur ces éléments : à la fois voir les connexions entre les documents retenus, mais aussi proposer de nouveaux documents en fonction de cet ensemble de documents favoris qui deviendrait ainsi la base du cadre de la recherche. Celle-ci serait dynamique et évoluerait en fonction du jeu de rétention des favoris.

5. Limites et prolongements

Du fait du caractère exploratoire de notre démarche, avec ses seuls neufs testeurs, il est en l'état difficile de prolonger plus en avant les perspectives posées ici traitant les dynamiques de recherche-production tel un processus de production de sens. De même, si la démarche permet de consigner par vidéo les manières dont un ensemble d'outils et d'objets techniques participent de l'activité, les matériaux recueillis sont encore sous-exploités et semblent autoriser davantage d'analyses sur les interactions à l'œuvre. Enfin, du fait de la durée limitée de l'implication des utilisateurs dans l'étude, évaluer la manière dont les règles/habitus professionnels peuvent participer aux formes d'appropriation de l'outil et réciproquement comment l'outil peut déplacer certaines règles/habitus professionnels s'avère complexe à identifier. Il y a un intérêt projeté des utilisateurs dans leurs pratiques professionnelles, qu'ils soient journaliste, chercheur, *community manager*, bibliothécaire ou encore chargé de projet numérique. Ils considèrent LIMAH comme un support où il serait possible de constituer sa propre

base de données, avec, pour certains, la sélection de médias déjà présente (Wikipédia, Cairn, Gallica, AFP), ou encore les réseaux sociaux, en fonction des besoins liés à leurs métiers. Il s'agit là d'une voie qu'emprunte l'équipe de chercheurs en informatique de ce projet.

6. Bibliographie

Bates M. J., « The design of browsing and berrypicking techniques for the online search interface », *Online Review*, vol. 13, n°5, 1989, p. 407-424.

Bois R., Gravier G., Sébillot P., Morin E., « Vers une typologie de liens entre contenus journalistiques ». *22e conférence Traitement automatique des langues naturelles, TALN 2015*, Caen, p. 515-521.

Bois R., Vukotić V., Simon A.-R., Sicre R., Raymond C., Sébillot P., Gravier G., « Exploiting Multimodality in Video Hyperlinking to Improve Target Diversity ». *MMM2017 - International Conference on Multimedia Modeling*, Reykyavik, Iceland : Springer.

Cahour B., Licoppe C., « Confrontations aux traces de son activité », *Revue d'anthropologie des connaissances*, vol. 4, n°2, 2010, p. 243-253.

Chaudiron S., Ihadjadène M., « Quelle place pour l'usager dans l'évaluation des SRI ? », Recherches *récentes en sciences de l'information : Convergences et dynamiques*, 2002, p. 211-231, Toulouse, ADBS Editions.

Ellis D., « A Behavioural Approach To Information Retrieval System Design », *Journal of Documentation*, vol. 4, n°3, 1989, p. 171-212.

Fogg B., Tseng H., « The elements of computer credibility », Proceedings *of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit*, ACM, New York, 1999, p. 80-87.

Hutchins E., Klausen T., « Distributed cognition in an airline cockpit », *Cognition and communication at work*, Cambridge University Press, Cambridge, 1996, p. 15-34.

Kaufmann J.-C., *L'entretien compréhensif,* Paris, Hachette, 2011.

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.

McKenzie P. J., « A Model of Information Practices in Accounts of Everyday Life Information Seeking », *Journal of Documentation,* vol. 59, n°1, 2003, p. 19-40.

Serres A., *Dans le labyrinthe : évaluer l'information sur internet,* Caen, C & F éd., 2012.

Simon A.-R., Bois R., Gravier G., Sébillot P., Morin E., Moens S., « Hierarchical Topic Models for Language-based Video Hyperlinking ». *Workshop on Speech, Language and Audio in Multimedia*, Brisbane, Australia.

Simonnot B., *L'accès à l'information en ligne : moteurs, dispositifs et médiations,* Cachan, Hermès-Lavoisier, 2012.

Theureau J., « Les entretiens d'autoconfrontation et de remise en situation par les traces matérielles et le programme de recherche "cours d'action" », *Revue d'anthropologie des connaissances*, vol. 4, n°2, 2010, p. 287-322.

Weick K. E., Sutcliffe K. M., Obstfeld D. « Organizing and the Process of Sensemaking », *Organization Science*, vol. 16, n°4, 2005, p. 409-421.