

CHAPITRE 2

RECHERCHE D'INFORMATION DANS LES DOCUMENTS ELECTRONIQUES

JÉROME DINET et ANDRÉ TRICOT

1. INTRODUCTION

Les êtres humains sont souvent dans des situations où ils manquent de connaissances pour comprendre et/ou pour agir. Une des solutions pour eux est de rechercher de l'information dans des documents, afin que l'information trouvée leur permette de mieux comprendre la situation et d'agir efficacement (Pirolli & Card, 1999). De plus en plus, ces documents sont électroniques. Ils permettent d'accéder à une quantité beaucoup plus grande d'informations, beaucoup plus rapidement, que les documents papiers. Mais ils impliquent aussi beaucoup plus de difficultés, par la quantité d'information et de fonctionnalités qu'ils présentent. Ces difficultés sont diverses : ne pas trouver l'information que l'on cherche ; trouver une information que l'on ne cherche pas ; ne pas comprendre l'information trouvée ; ne pas reconnaître l'information recherchée quand on l'a trouvée ; ne pas savoir utiliser les fonctionnalités proposées par le système ; se « perdre » dans le document ; oublier son but ; *etc.*

Face à l'augmentation considérable de la masse d'information, face à la diversification des supports d'informations (ouvrages papiers, CD-Roms, sites Web) et face à la nécessité d'avoir un accès à l'information de plus en plus rapide et moins assujetti à la distance géographique, le recours à l'outil informatique s'est vite révélé nécessaire pour effectuer des recherches d'informations (pour quelques repères historiques : Barès, 1984). Or, comme le fait remarquer Molard (1995), « le développement des Nouvelles Technologies implique la multiplication des types de documents

consultables pour le public [et] il en résulte une multiplication des possibilités de s'informer de façon facile et rapide. La documentation devient accessible à tous. Elle ne se limite plus aux seuls initiés fréquentant des lieux de documentation très spécialisés » (p.52). En d'autres termes, l'accès aux documents et aux informations jusqu'alors réservé à des professionnels est rendu possible pour le plus grand nombre, sans réelle formation. La numérisation des documents et leur mise sur supports informatiques a considérablement modifié la structure des informations, les pratiques qui y sont liées et notamment la recherche de ces informations (Jacobson & Ignacio, 1997). En d'autres termes, l'un des principaux paradoxes actuels est que des individus qui ne sont pas spécifiquement formés aux systèmes documentaires et à l'utilisation des outils de recherche d'information ont aujourd'hui accès très aisément à ces mêmes systèmes et outils initialement créés par et pour des spécialistes. Ce nouvel utilisateur qui a désormais accès à ces outils jusqu'alors « réservés » à des spécialistes et à l'environnement numérique se trouve vite démuné.

Sous l'impulsion de l'arrivée massive des technologies de l'information et de la communication au sein de tous nos espaces de travail et de loisir, nos comportements pour rechercher des informations ont profondément été modifiés (Marchionini, 1995). La numérisation des documents a eu des répercussions sur les aspects spatio-temporels et socio-économiques de l'activité : la quantité d'informations auxquelles nous avons désormais accès ne cesse d'augmenter ; une recherche d'information peut souvent être réalisée à toute heure de la journée et de la nuit (l'utilisateur n'est plus soumis aux horaires d'ouverture au public d'une bibliothèque) ; les formats d'informations ne cessent de se multiplier et d'associer diverses modalités sensorielles (par ex., avec les documents multimédias) ; nous ne sommes plus obligés de nous déplacer pour « butiner » dans des casiers de bibliothèques. Mais, la numérisation des documents a également eu des répercussions sur nos stratégies mentales : l'utilisateur peut choisir sa base de données, son point d'entrée dans la base choisie, *etc.* ; ce même utilisateur peut organiser différemment son travail ; nous pouvons utiliser des outils qui étaient inexistantes avec les systèmes « papier-crayon », tels que les aides en-ligne, les *thesaurii* ou la logique booléenne ; notre perception de l'activité de recherche d'information change dans le sens où nous devenons de plus en plus exigeants sur la qualité des résultats tout en développant notre impatience.

Ajoutons que les professionnels de la documentation, qui étaient les usagers quasi exclusifs des systèmes documentaires informatisés jusqu'à la fin des années 1980, ne réalisent pas les mêmes tâches que les non-spécialistes qui utilisent aujourd'hui ces systèmes. Les documentalistes recherchent de l'information pour autrui : leur but est donc nécessairement explicite ou du moins de rendre plus explicite la demande de l'utilisateur, souvent précis, leur démarche est méthodique, leur résultat ne bénéficie d'un retour qu'après que l'activité de recherche aient été terminée. C'est tout le contraire pour les non-spécialistes qui recherchent de l'information pour eux-mêmes. On comprend alors que des activités si différentes donnent lieu à des modélisations différentes.

L'objectif de ce chapitre est double : dans une première partie, seront présentés les modèles récents issus de la psychologie cognitive et des sciences de l'information et de la communication qui contribuent à mieux faire connaître cette activité humaine qu'est la recherche d'information dans les environnements numériques ; dans une seconde partie, seront discutés les facteurs expliquant les différences inter et intra-individuelles lors de recherches d'information réalisées dans des documents numériques.

2. DES MODÈLES COMPORTEMENTAUX GÉNÉRALISTES AUX MODÈLES DES PROCESSUS COGNITIFS

Ce sont les sciences de l'information et de la communication qui se sont les premières intéressées à l'activité de recherche d'information, en proposant des modèles des comportements des utilisateurs. Le principal objectif était d'une part, de fournir des outils pour permettre aux professionnels en contact avec ces utilisateurs de mieux comprendre leurs difficultés et d'autre part, de fournir des supports pour des formations éventuelles.

2.1. LES MODELES GÉNÉRALISTES

2.1.1. *Le modèle de Marchionini (1995)*

Lorsque l'on recherche des informations, l'objectif peut être extrêmement précis (par ex., trouver une date correspondant à un événement) ou relativement vague (par ex., préparer une exposition sur la vie au Moyen-âge). De plus, certaines recherches d'information ne visent pas forcément à trouver une information mais correspondent plutôt à du butinage, pour « passer le temps » ou encore pour s'amuser. C'est à partir de ce simple constat que Marchionini (1995) a proposé une typologie des comportements et stratégies de recherche d'information en fonction de la nature et du degré de précision de l'objectif :

- la recherche dirigée par un but : dans ce cas, la recherche vise à répondre à une question précise (« quelle a été la plus forte température relevée en France depuis 30 ans pour le mois de juillet ? ») ;
- la recherche semi-dirigée : ici, l'objectif de la recherche d'information est plus imprécis et moins bien défini (« devant préparer un exposé oral, je dois rechercher des informations sur la conquête de l'espace ») ;
- la recherche non dirigée : il n'y a pas réellement d'objectifs à l'activité, si ce n'est le simple plaisir de « surfer » et de butiner sans réel but informatif (« quels sont les thèmes qui font l'objet des forums électroniques ces temps-ci ? »).

C'est sur la base de cette distinction entre objectifs de recherche que, la même année, Catledge et Pitkow (1995) distinguent les chercheurs d'information (*searchers*) des simples butineurs (*browsers*), les premiers utilisant des stratégies analytiques et les seconds des stratégies par butinage. Marchionini (1995) décompose l'activité de recherche d'information en huit étapes :

- identification du besoin d'information par l'utilisateur ;
- définition et compréhension du problème lié au besoin d'information ;
- choix d'un système pertinent pour résoudre le problème : parmi les systèmes qu'un individu a à sa disposition (moteur, annuaire, métamoteur, base de données, banques de données, ouvrages papiers), il choisit celui qui lui semble le plus pertinent ;

- formulation d'une requête : une fois le système choisi, l'utilisateur doit l'interroger en produisant un ou des mots-clés ;
- validation et exécution de la recherche : durant cette phase, l'individu valide et lance la recherche effective généralement en cliquant sur le bouton approprié du système (par ex., « chercher », « OK », « valider », « lancer la recherche ») ;
- examen des résultats fournis par le système : le système répond à l'utilisateur en lui proposant des listes de résultats potentiellement pertinents, généralement sur la base d'une simple reconnaissance lexicale ;
- extraction des informations pertinentes : parmi les informations proposées par le système, l'utilisateur choisit celle qui lui semble la plus pertinente et la traite ;
- évaluation – itération : si l'utilisateur estime que les informations fournies par le système sont pertinentes et suffisantes, alors il met un terme à son activité. Sinon, il recommence depuis la première étape.

Le modèle proposé par Marchionini est donc un modèle séquentiel et linéaire. La recherche d'information est également vue comme un processus itératif dans le sens où les comportements et stratégies d'un individu peuvent évoluer au fil de l'activité. Par exemple, les résultats d'un premier cycle de recherche peuvent permettre d'affiner et/ou d'élargir la recherche en faisant apparaître des mots-clés auxquels l'individu n'avait pas spontanément pensé.

L'un des principaux apports des travaux de Marchionini et de ses collaborateurs a été d'introduire les aspects différentiels dans les modèles qui existaient alors. En effet, rapidement, Marchionini et ses collaborateurs se sont focalisés sur les différences de comportements et de stratégies de recherche d'information entre d'une part, les experts et les novices et d'autre part, entre différents types d'expertises (par ex., Marchionini, 1991 ; 1992 ; Marchionini *et al.*, 1993 ; pour une synthèse : Dinet & Rouet, 2002). Les résultats des études ont principalement montré que :

- deux types d'expertises relativement indépendantes existent : d'une part, l'expertise du domaine et d'autre part, l'expertise en recherche d'information. D'un point de vue cognitif, l'expertise du domaine renvoie aux savoirs ou connaissances déclaratives tandis que l'expertise en recherche d'information renvoie au savoir-faire ou connaissances procédurales liées à l'activité elle-même ;

- les experts du domaine emploient des stratégies centrées sur le thème et les informations contenues dans les documents. Concrètement, plus un individu a de connaissances sur le thème de la recherche et plus il est capable de produire un grand nombre de mots-clés associés à ce thème. De même, il consacre plus de temps à analyser le contenu des documents et évalue avec plus de précision et de rapidité la pertinence des informations qu'il trouve ;
- les experts en recherche d'information emploient des stratégies centrées sur l'activité elle-même. Concrètement, ces experts utilisent plus fréquemment et efficacement les outils spécifiques (opérateurs booléens, troncature, *etc.*), accordent beaucoup d'importance aux procédures et techniques pour mener à bien la recherche, et évaluent régulièrement si leur procédure est la plus pertinente en fonction des outils dont ils disposent.

En d'autres termes, le type d'expertise (du domaine ou en recherche d'information) semble jouer un rôle sur l'allocation des ressources des utilisateurs puisque la gestion cognitive de l'activité est différente. Les experts du domaine adoptent des comportements dirigés par le contenu (*content-driven*) tandis que les experts en recherche d'information adoptent des comportements dirigés par la tâche (*search-driven*). S'il adopte un comportement dirigé par le contenu, l'individu peut vouloir recueillir de manière exhaustive toutes les informations récentes relatives à un thème précis. S'il adopte un comportement dirigé par la tâche, l'individu peut décider de se focaliser sur la procédure et le respect des consignes, par exemple, en utilisant toutes les ressources documentaires dont il dispose. Bien évidemment, l'idéal est qu'un utilisateur combine les deux types d'expertises.

Les groupes de comportements identifiés par Marchionini comprennent différentes actions possibles. Par exemple, parmi les actions possibles liées aux comportements dirigés par la tâche, l'individu a le choix entre combiner des termes à l'aide des opérateurs booléens après avoir identifié les concepts pertinents dans le thème de la recherche d'information (stratégie dite de *building blocks*), partir de concepts généraux pour progressivement aller vers de plus spécifiques (stratégie de *successive fractions*) ou encore utiliser des termes contenus dans des résultats pour mener un nouveau cycle de recherche (stratégie de *interactive scanning*). Un individu donné n'utilise donc

pas un et un seul type de stratégies : de nombreux autres facteurs interviennent, dont le contexte de la recherche, le thème de la recherche d'information ou encore le système informatique utilisé.

Au sein des sciences de l'information, un second modèle connaît un vif succès, celui de Kuhlthau. Nous le présentons maintenant, en mettant en exergue les principales différences avec celui de Marchionini.

2.1.2. Le modèle Information Search Process (ISP) de Kuhlthau (1997)

L'une des principales caractéristiques du modèle proposé par Kuhlthau (1987) est d'intégrer les aspects cognitifs, affectifs et physiques liés à l'activité de recherche d'information. Mais, il ne s'agit pas d'un modèle basé sur des concepts définis avec précision puisque, pour Kuhlthau, les aspects cognitifs correspondent aux pensées et à l'intellect de manière générale, les facteurs affectifs aux affects, impressions et émotions et les facteurs physiques aux actions ou comportements.

Son modèle ISP est apparu pour la première fois dans une publication en 1985 mais a été depuis largement enrichi (Kuhlthau, 1988, 1991, 1994, 1997 ; Shannon, 2002). Le modèle ISP a rencontré et rencontre toujours un large succès auprès des spécialistes des activités documentaires (documentalistes, bibliothécaires, experts en bibliothéconomie), surtout s'ils exercent dans le milieu scolaire ou universitaire. Ce succès s'explique peut-être par le fait que le modèle intègre les facteurs affectifs et émotionnels dans la recherche d'information.

Selon Kuhlthau, la recherche d'information se compose de sept étapes successives, chacune faisant intervenir des composants issus de l'un ou de plusieurs des trois domaines (affects, intellects, actions) :

- l'initiation de la tâche : toute recherche d'information débute généralement par l'explicitation de l'objectif de la recherche d'information, cet objectif prenant parfois la forme d'une question imprécise. L'individu devant réaliser l'activité va alors, sur la base de ses connaissances antérieures relatives au thème, interpréter cette question ;
- la sélection : l'individu va planifier son activité en élaborant des sous-buts. C'est durant cette étape qu'est établi l'ordre dans

lequel les sous-buts doivent être atteints pour réaliser l'objectif final (répondre à la question) ;

- l'exploration : les différentes sources d'informations disponibles dans l'environnement sont recensées. Ces sources peuvent être des documents sur supports papier ou numérique, hors ou en-ligne, mais également une ou des personnes humaines (collègue, bibliothécaire, documentaliste, *etc.*) ;
- la formulation : une fois la source d'information la plus pertinente identifiée, l'individu va formuler des idées précises généralement traduites en mots-clés, pour pouvoir interroger cette source. Lors de cette reformulation, certaines idées précises peuvent alors apparaître subitement (*insights*) ;
- la collecte : durant cette étape, les informations permettant d'atteindre l'un des sous-buts (et donc le but général) sont prélevées et collectées ;
- la présentation : les informations collectées sont progressivement organisées et triées afin de répondre le plus adéquatement à la question initiale ;
- l'évaluation : après avoir organisé les informations collectées et les avoir retranscrites, l'individu contrôle et évalue le produit issu de son activité. C'est également durant cette phase d'évaluation que les révisions, vérifications et éventuelles corrections sont effectuées.

A partir des publications du début des années 1990 (Kuhlthau, 1993), les stratégies utilisables par l'utilisateur pour réaliser chacune des sept étapes ont été introduites au modèle. Le tableau 2.1 présente des exemples d'affects, pensées, actions et stratégies qui peuvent être alors associées à chacune des sept étapes de la recherche d'information.

Etape	Affects	Intellects	Actions	Stratégies
Initiation	Incertitude	Analyse de la demande Appel aux connaissances antérieures	Echange avec autrui Butinage dans la bibliothèque	Brainstorming Discussion
Sélection	Confusion Anxiété parfois	Elaboration des sous-buts Choix des mots-clés	Consultation de répertoires Recherches préliminaires	Discussion Exploration de thèmes liés
Exploration	Confusion Doute Incertitude	Identification des sources Précision de la demande	Localisation des sources Prise de notes	Lectures Listage des mots-clés
Formulation	Optimisme	Formulation de critères précis Production d'idées Apparition d'insights	Lectures de notes relatives aux sous-buts	Articulation des sous-buts
Collecte	Intérêt accru	Prélèvement d'informations	Prises de notes précises (par ex., citations)	Utilisation d'index Demande d'aide auprès du documentaliste
Présentation	Optimisme	Catégorisation et organisation des informations	Organisation des notes	Utilisations d'expériences antérieures
Evaluation	Satisfaction	Identification d'informations complémentaires	Contrôles et dernières relectures	Retour aux sources

Tableau 2.1. – *Exemples d'affects, pensées, actions et stratégies liés aux étapes de la recherche d'information selon le modèle ISP*

L'un des principaux apports du modèle ISP est donc d'intégrer les facteurs affectifs et émotionnels dans les comportements des usagers, ces facteurs n'étant pas considérés dans l'approche de Marchionini. Mais, la principale critique que l'on peut produire à l'encontre de ce modèle et de celui de Marchionini (1995) est que ce sont plus des modèles prescriptifs que des modèles explicatifs dans le sens où ils détaillent les stratégies optimales et les comportements tels qu'ils devraient être et non pas tels qu'ils sont réellement. Ces modèles sont donc extrêmement précieux pour les professionnels de l'information ou de la documentation qui peuvent voir en eux des supports de formation ou des référentiels de compétences à acquérir. Mais, ils renseignent très peu sur les comportements réels et les processus cognitifs sous-jacents des individus recherchant des informations dans les environnements documentaires complexes.

2.2. LES MODELES DES PROCESSUS COGNITIFS

Même si les modèles présentés précédemment intègrent certains aspects psychologiques, ils ne décrivent pas précisément les processus cognitifs impliqués lors de recherche d'information et ils ne reposent pas sur des fondements théoriques clairement définis. Depuis une vingtaine d'années, de nouveaux modèles centrés sur les processus cognitifs ont fait leur apparition.

2.2.1. Le modèle séquentiel de Guthrie (1988)

Guthrie (1988) est l'un des premiers auteurs à avoir proposé un modèle pour rendre compte de l'organisation des processus cognitifs d'accès à l'information dans des documents complexes. Selon cet auteur, la recherche d'information comporte cinq étapes successives ou constituants cognitifs :

- la formation d'un but : à partir d'une question et/ou d'une consigne, l'individu élabore une représentation mentale de l'objectif à atteindre (nature de l'information à chercher) ;
- la sélection d'une catégorie : l'individu choisit, parmi les sources d'informations disponibles, celle qui paraît la plus pertinente à partir des caractéristiques structurales du document (par ex., un paragraphe ou une colonne dans un tableau) ;
- l'extraction de l'information : dans la catégorie choisie, l'individu traite l'information de contenu (par ex., identification d'un mot ou d'une valeur numérique) ;
- l'intégration : l'individu intègre alors l'information traitée avec les informations déjà stockées en mémoire à long terme ou extraites depuis des informations tout juste traitées ;
- le recyclage : si l'objectif n'est pas atteint (par ex., si les informations sont incomplètes), alors l'individu répète les quatre premières étapes jusqu'à ce qu'une réponse satisfaisante puisse être produite.

Depuis la publication de ce modèle, les travaux s'intéressant aux processus cognitifs impliqués lors de la recherche d'informations ont connu une croissance rapide. Ceci tient pour une large part à la diffusion des technologies d'information et de communication, qui ont commencé à transformer radicalement les environnements documentaires tant dans les milieux scolaires que sur les lieux de

travail et, plus récemment, dans les foyers domestiques. Si le modèle de Guthrie (1988) reste pour l'essentiel valide, il est limité à des tâches simples de localisation. En effet, le traitement du contenu (identification, compréhension, évaluation) se réduit dans le modèle proposé par Guthrie (1988) à une simple localisation et extraction d'informations. On peut comprendre cette terminologie lorsqu'il s'agit de trouver le taux d'une taxe sur un bulletin de paie ou l'horaire de départ d'un train, moins lorsqu'il s'agit de documents et de questions complexes qui nécessitent le traitement de nombreuses pages. Même si ultérieurement Guthrie et ses collaborateurs ont appliqué leur modèle à des tâches de recherche dans des documents plus complexes (Dreher & Guthrie, 1990 ; Guthrie, Briten, & Barker, 1991), leur approche théorique reste marquée par le contexte des travaux originaux.

De plus, le modèle de Guthrie est un modèle linéaire, séquentiel, l'étape de recyclage n'intervenant qu'en fin de course pour assurer la reprise du processus en cas d'échec. Effectivement, un seul cycle est le plus souvent suffisant pour les simples tâches d'extraction de valeurs numériques ou d'informations telles que celles étudiées par Guthrie et ses collaborateurs. Mais si l'on s'intéresse au traitement de questions complexes à l'aide de documents numériques complexes, il est clair que le cas le plus fréquent est celui où la recherche demande plusieurs cycles, avec à chaque cycle une élaboration progressive du but, une sélection des parties du texte intéressantes, un examen plus ou moins approfondi, une relecture et corroboration des autres parties, et ainsi de suite.

2.2.2. Le modèle EST de Rouet et Tricot (1998)

D'après Rouet et Tricot (1998), il est possible de décrire l'activité de recherche d'information de façon plus complète que celle de Guthrie, en mettant en relation, d'une part, un cycle évaluation sélection traitement et, d'autre part, les activités de planification, contrôle métacognitif et régulation. Cette description, appelée EST (pour Evaluation, Sélection, Traitement), articule la compréhension du contenu des documents et le processus de recherche d'information comme résolution de problèmes. Cette modélisation est, à la fois, cyclique comme celle de Guthrie et fondée sur une distinction de trois grands types de traitements. Le modèle EST a une prétention générale. Il concerne la réalisation de tâches à but précis ou flou, l'extraction de

cibles uniques ou nombreuses, la sélection de références comme l'exploitation de documents. L'activité est particularisée en fonction de la tâche et de l'environnement.

De façon plus détaillée, la recherche d'information dans un document est considérée dans EST comme un cycle de traitement constitué de trois phases principales : la sélection de l'information, le traitement de l'information sélectionnée et l'évaluation de la pertinence de cette information, en fonction du but visé par l'individu. Dans ce modèle, l'activité de sélection et d'évaluation recouvre un processus de gestion de l'activité (planification de la recherche et évaluation de l'écart entre la situation actuelle et le but visé) et un processus de traitement des informations relationnelles (liens, menus, boutons). Ce double processus est conduit en relation avec la représentation que l'utilisateur se fait de la tâche. Celle-ci inclut une représentation du but et peut être modifiée dynamiquement au cours de l'activité.

Chaque phase du modèle EST est considérée comme un module de traitement cognitif.

Le module « évaluation » de EST a pour rôle principal de comparer la représentation du but, c'est-à-dire la représentation du besoin d'information, à la représentation du contenu traité. Cette comparaison donne lieu à un jugement de proximité entre ces deux représentations et se traduit en une prise de décision concernant la sélection des items suivants.

Etat 0 : à l'état initial de la recherche, le module évaluation est composé d'une représentation du but à atteindre. Ce but à atteindre correspond à l'état n du cycle. La représentation du but comporte un versant général (apprendre, concevoir, *etc.*) et un versant opérationnel (trouver telle information dans tel système). Cette représentation du but guide la sélection.

Etat 1 : une fois que la première action de sélection a été effectuée et que l'information trouvée a été traitée, le module d'évaluation compare l'information traitée et la représentation du but, pour évaluer en quoi cette information trouvée contribue à l'atteinte du but. Trois situations différentes sont possibles. À chacune de ces situations correspond une décision rationnelle. Si l'information trouvée correspond exactement au but (ou du moins, à la représentation que l'individu se fait du but), elle contribue totalement à l'atteinte du but et doit entraîner la décision d'arrêter la recherche. Si l'information trouvée ne correspond que partiellement à la représentation du but, le

but est partiellement atteint et cela doit entraîner une décision de continuer la stratégie de sélection, vers plus de précision ou au contraire plus de généralité. Enfin, si l'information trouvée ne correspond pas du tout à la représentation du but, le but n'est pas du tout atteint et doit entraîner une décision de changer de stratégie.

Toutefois, les individus peuvent agir de façon non rationnelle. Avec le modèle rationnel, le traitement opéré par le module « évaluation » n'entraîne pas de modification de la représentation du but mais seulement une décision concernant une modification de la « sélection ». Avec l'humain, la représentation du but peut changer :

- le but peut être totalement atteint, mais le module d'évaluation ne déclenche pas l'arrêt de la sélection : une nouvelle représentation du but prend la place de la représentation originale (plus vague, plus spécifique, *etc.*) ;
- le but est partiellement atteint, mais la représentation du but change ;
- l'information trouvée ne correspond pas du tout à la représentation du but et la représentation du but change, soit pour correspondre à l'information trouvée (biais de confirmation), soit pour évoluer vers une autre.

Il est à noter qu'en recherche d'information, un document peut ne pas correspondre au but mais s'en rapprocher. Le processus de jugement – décision est alors différent selon que la représentation du but s'exprime plutôt en termes de résultat – contenu ou en termes de procédure. En effet, une représentation en termes de résultat – contenu peut conduire à un jugement fin sans forcément être capable d'anticiper à long terme sur la pertinence d'une sélection. À l'inverse, une représentation en termes de procédure peut permettre d'anticiper à long terme la pertinence d'une sélection sans forcément fournir les éléments utiles pour un jugement fin sur le contenu. Ce type de représentation (en termes de résultat – contenu ou en termes de procédure) dépend généralement du degré d'expertise de l'individu dans le domaine traité et dans l'utilisation de ce type de système d'information. Une représentation trop profane en termes de procédure peut entraîner un phénomène de « poursuite sur la même route », procédé très inefficace ou très coûteux (Gray, 1990).

La difficulté principale de fonctionnement de l'évaluation est de maintenir la représentation du but dans un état stable. Cette difficulté a des conséquences négatives, car elle entraîne la désorientation de l'utilisateur (phénomène de noyade en digressions, Foss, 1988).

L'utilisateur peut, au bout d'un certain temps de consultation, perdre son but (c'est-à-dire ne plus savoir lui-même pourquoi il utilise le système). Mais le fait que la représentation du but puisse évoluer a aussi des conséquences positives. Dans les activités à but flou, la modification de la représentation du but peut consister à mieux le spécifier.

Dans le cycle EST, l'individu doit maintenir en mémoire de travail trois types de représentation : une représentation du but, une représentation de la stratégie de sélection, et une représentation du contenu traité. Or, plusieurs études ont montré que si certains mises en pages des documents électroniques semblaient profitables aux experts, ces mêmes mises en page perturbaient considérablement les novices (Gray, 1990 ; Rouet 1990), ces novices n'ayant pas développé de compétences spécifiques pour simultanément conserver actives ces trois types de représentations.

2.2.3. Le modèle CoLiDes (2000)

Kitajima, Blackmon et Polson (2000) ont élaboré un nouveau modèle des processus cognitifs impliqués dans les activités de recherche d'information, modèle qu'ils enrichissent depuis (Blackmon, Kitajima, & Polson, 2005 ; Blackmon, Kitajima, Polson, & Lewis, 2002). Ce nouveau modèle, dont l'acronyme est CoLiDes pour *Comprehension-based Linked model of Deliberate Search* s'applique spécifiquement à la navigation et la recherche d'information réalisée dans les sites Web surtout à vocation commerciale.

Selon ce modèle, les actions que peut réaliser un individu à partir d'une page Web font intervenir deux types de processus en parallèle : d'une part, des processus attentionnels et d'autre part, des processus de sélection d'action. Chacun de ces deux processus s'applique à plusieurs moments de l'activité :

- lors de la formation d'un but : comme pour la plupart des modèles présentés dans ce chapitre, à partir d'une question et/ou d'une consigne, l'individu élabore une représentation mentale de l'objectif à atteindre ;
- lors de l'analyse de la page Web : face à la page Web, l'individu procède à une analyse visuelle de la structure et du contenu de ladite page, en distinguant les catégories d'informations selon leur emplacement, leur typographie, etc. ;

- lors du découpage en zones de contenu : sur la base de ses connaissances acquises par apprentissage explicite ou par l'expérience, l'utilisateur segmente la page en zones signifiantes dans le sens où les informations présentes dans une même zone sont proches sémantiquement ;
- lors de l'analyse des zones de contenu : une fois une zone identifiée, le lecteur va focaliser son attention sur son contenu et décide de le traiter afin de le comprendre ;
- lors de la compréhension du contenu : sur la base des informations contenues dans la zone traitée, le lecteur élabore une nouvelle représentation mentale. Il compare alors cette représentation mentale du contenu de la zone avec sa représentation initiale du but : si la distance entre ces deux représentations mentales lui semble faible, alors l'utilisateur considère que les informations contenues dans cette zone vont lui permettre d'atteindre son objectif et il décide alors de sélectionner un lien présent dans cette zone ;
- lors de la compréhension du lien sélectionné : parmi les liens hypertextuels présents dans cette zone, l'utilisateur décide de sélectionner (et donc de cliquer sur) celui qui le semble le plus pertinent pour poursuivre son activité ;
- lors de la sélection et réalisation de l'action : cette dernière phase correspond à la réalisation effective du comportement, c'est-à-dire le plus souvent au clic sur le lien.

Comme pour les autres modèles exposés ci-dessus, CoLiDes postule que toute recherche d'information fait intervenir des processus cognitifs ascendants et descendants. En effet, le lecteur – chercheur d'information utilise d'une part, les caractéristiques perceptives prélevées depuis la page Web et d'autre part, ses connaissances stockées en mémoire à long terme relatives à ces caractéristiques et au Web en général pour naviguer.

Selon les auteurs de ce modèle, la perte de l'utilisateur dans les méandres des pages Web d'un site et l'abandon subséquent de l'activité peuvent provenir de deux types de problèmes selon le processus affecté : en ce qui concerne les processus attentionnels, des éléments perceptifs non pertinents peuvent détourner l'attention de l'utilisateur (par ex., l'apparition inopinée de bandeaux publicitaires) ; en ce qui concerne les processus de sélection d'action, une terminologie inadaptée ou trop complexe peut gêner la compréhension des contenus ou des liens hypertextuels. Dans tous les cas, l'abandon

provoque d'une part, une frustration chez l'utilisateur qui n'est pas parvenu à trouver ce qu'il cherchait et d'autre part, une frustration chez les propriétaires du site commercial qui ne réaliseront pas de vente. Sur la base de leur modèle, les concepteurs de CoLiDes mettent actuellement au point une méthode d'inspection de l'utilisabilité d'un site Web.

2.2.4. Les modèles adaptatifs issus de ACT-R

La théorie ACT-R (*Adaptive Control of Thought - Rational*) d'Anderson (Anderson *et al.*, 2004 ; Anderson & Lebière, 1998) a servi de base à deux modèles de la recherche d'information. C'est le cas du modèle ACT-IF de Pirolli et Card (1999), qui est présenté de façon relativement détaillée et critique dans Tricot (2006). C'est le cas aussi du modèle de Fu et Gray (2006). Rappelons que la théorie d'Anderson est un modèle rationnel de la cognition humaine : les activités sont envisagées comme relevant d'une adaptation de l'individu à son environnement fondée sur la recherche d'*optima*. Un *optimum* est simplement un bon rapport bénéfice / coût : les humains réalisent des activités dans des environnements quand le gain que représente potentiellement pour eux cette activité est supérieur au coût de la mise en œuvre de cette même activité. La rationalité est conçue comme limitée par des contraintes, notamment temporelles, ainsi que par les capacités et connaissances que l'individu a de l'environnement, donc de l'interprétation qu'il va faire de la situation, du but qu'il poursuit, de l'utilité espérée, du coût, *etc.* Les modélisations de l'activité issues de ACT-R sont précises, mises en œuvre informatiquement et donc comparables aux performances des humains. En bref, contrairement aux modèles présentés ci-dessus, elles sont véritablement réfutables.

Pour Fu et Gray (2006), comme précédemment pour Card et Pirolli (1999), la recherche d'information obéit au principe de rationalité : un individu va rechercher une information si l'utilité espérée de cette information est supérieure au coût que représente sa recherche. Les auteurs prennent l'exemple de la recherche d'un itinéraire rapide pour aller d'un endroit à un autre. L'utilité espérée est le gain de temps que représente un trajet rapide par rapport à un trajet moins rapide (par ex. celui au cours duquel on tâtonne). Le coût est le temps passé à chercher cet itinéraire. On va chercher cet itinéraire si l'on estime que le gain de temps sur le trajet est supérieur au temps passé à chercher.

La recherche d'information est donc conçue comme une alternative à la résolution de problème par l'action : l'évaluation du rapport bénéfice / coût de la recherche d'un itinéraire est comparée à l'évaluation du rapport bénéfice / coût de la réalisation de l'itinéraire par tâtonnement. Dans un environnement donné, avec un but donné, les individus qui manquent de connaissances pour atteindre le but vont rechercher de l'information si cela est plus rentable que d'essayer d'atteindre le but par l'action : essais et erreurs, raisonnements, tests d'hypothèses, ou encore tâtonnement.

Le modèle de Fu et Gray (2006) se focalise sur l'analyse des coûts de la recherche d'information, en particulier sur ce qui conduit les individus à décider l'arrêt de la recherche d'information. Il est fondé sur l'hypothèse générale selon laquelle face à un environnement nouveau les individus vont rechercher de l'information pour acquérir des connaissances à propos de cet environnement. Le modèle prédit que pour faire de bonnes évaluations du rapport bénéfice / coût, les individus doivent avoir suffisamment d'expérience (donc de connaissances) de l'environnement ; que dans les environnements changeants, les individus réagissent plus rapidement aux changements de coûts qu'aux changements de bénéfices ; que dans les environnements qui présentent un minimum local, c'est-à-dire une zone où continuer à rechercher de l'information ne rapporte rien pendant un certain laps de temps, les coûts élevés vont conduire à de mauvaises évaluations et à une exploration pauvre de l'espace problème. Les données empiriques obtenues par les auteurs sont largement cohérentes avec le modèle.

Tricot (2006) a déjà souligné les limites du modèle de Pirolli et Card (1999), dont certaines peuvent être aussi appliquées au modèle de Fu et Gray (2006) :

- la notion de besoin d'information est assimilée à celle d'utilité espérée de l'information, alors que nombreux sont les cas où une information utile et peu coûteuse, mais dont on ne connaît pas l'existence, ou surtout, qui pourrait contredire les croyances de l'individu, n'est pas recherchée ; nous allons aborder ce point ci-dessous ;
- les aspects émotionnels, notamment ceux liés au stress et à l'anxiété ne sont pas pris en compte, alors que, notamment dans le domaine médical, on montre que la recherche d'information est liée de façon très directe au stress et à

- l'anxiété, par exemple lors de l'annonce d'une maladie grave (par ex., Mesters, van den Borne, De Boer & Pruyn, 2001) ;
- les effets de l'expertise sont traités de façon très sommaire. L'expertise en recherche d'information est passée sous silence. Pourtant, de nombreux travaux empiriques montrent que la quantité d'information recherchée est liée de façon positive et quasi linéaire à l'expertise (plus on a de connaissance dans un domaine et plus on cherche d'informations dans ce domaine) : c'est le cas dans le domaine de l'orientation scolaire et professionnelle (par ex., Millar & Shevlin, 2003) ou dans celui de l'information sur les retraites (Joo & Grable, 2001).

Notons aussi que le modèle de Fu et Gray (2006) concerne la recherche d'information en général et non pas dans les documents. Leur matériel expérimental est un plan de réseau ferroviaire. L'intérêt du modèle dans le monde réel, en particulier en interaction humains – machines est discuté par les auteurs, mais de façon très prospective. Les auteurs suggèrent que leur modèle pourrait fournir une explication au « paradoxe de l'utilisateur actif » (Carroll & Rosson, 1987), selon lequel des utilisateurs expérimentés d'applications informatiques peuvent très bien adopter des stratégies d'utilisation non-optimales : par exemple, ils n'utilisent pas tel raccourci qui pourtant leur ferait gagner beaucoup de temps, ou bien ils persistent dans l'utilisation d'une application très mauvaise alors qu'apprendre à se servir d'une bonne application ne leur prendrait que peu de temps. Pour Fu et Gray (2006), leur modèle de l'effet du minimum local pourrait expliquer pourquoi les utilisateurs ne cherchent pas à faire autrement.

La modélisation de l'activité de recherche d'information dans les documents est donc un domaine récent, générant encore trop peu de travaux empiriques et théoriques. Il se divise, de façon assez commune, entre recherches à forte validité écologique et recherches à forte validité interne, sans que des ponts soient suffisamment établis entre ces deux approches.

3. LES FACTEURS IMPLIQUÉS ET LES CONTRAINTES

3.1. LES PARTICULARITÉS DE LA RECHERCHE D'INFORMATION DANS LES ENVIRONNEMENTS ÉLECTRONIQUES

Dans cette partie, nous allons décrire rapidement les effets liés à trois types de caractéristiques des documents électroniques : leur structure, leurs fonctionnalités et leur format de présentation. Il semble en effet que ces caractéristiques puissent affecter de façon importante l'activité de recherche d'information. On peut cependant regretter que, sur certains aspects, les résultats empiriques soient rares.

Tout d'abord, la structure des environnements documentaires électroniques présente la particularité de faire varier de façon importante sa profondeur (c'est-à-dire le nombre de pages ou de sous-menus que l'on peut ouvrir successivement et sans revenir en arrière) et sa largeur (c'est-à-dire le nombre de pages ou de sous-menus qui partent d'un même point). On sait que plus le menu est profond et plus les informations (fonctionnalités, contenus) auxquelles il donne accès sont difficiles à trouver (par ex., Miller, 1981). On a pu interpréter ce type de résultat comme étant dû à une surcharge cognitive. L'ergonome doit donc concevoir des artefacts peu coûteux à traiter cognitivement mais susceptibles d'aider l'utilisateur à se repérer dans le menu. Lemercier, Quaireau et Terrier (2003) ont fait l'hypothèse qu'un indigage perceptif coloré pourrait faciliter le traitement des menus, améliorer les performances de repérage et de mémorisation. Les résultats qu'ils ont obtenus vont, pour une bonne part, dans le sens de leur hypothèse.

Ensuite, les environnements électroniques présentent, comme nous l'avons dit plus haut, de nombreuses fonctionnalités. Un moteur de recherche (par ex., Google) permet spécifiquement une sélection parmi un ensemble important de documents, la puissance des moteurs actuels semblant sans véritable limite quantitative. Le problème principal des moteurs est la quantité de bruit (documents non pertinents) qu'ils génèrent, un trop grand bruit pouvant masquer le résultat. Leur utilisation implique une connaissance de leur syntaxe, voire de principes généraux, par l'utilisateur, certains novices pouvant

ne trouver absolument aucun document pertinent par l'intermédiaire d'un moteur (Koenemann & Belkin, 1996).

Les liens hypertextes permettent spécifiquement la gestion de buts flous, la lecture de l'hypertexte permettant au fur et à mesure à l'utilisateur d'élaborer une représentation plus structurée du contenu traité et de son but. Mais, permettant facilement au but informationnel d'évoluer, ils permettent aussi de le perdre. Les hypertextes imposent généralement un effort cognitif et un contrôle métacognitif très importants.

Malgré l'existence déjà ancienne des trois fonctionnalités précitées, on sait aujourd'hui que de nombreuses difficultés de navigation dans les documents électroniques existent. D'innombrables nouvelles fonctionnalités, dont celles qui permettent de retrouver ce que l'on a fait précédemment (les historiques) ou de cocher des sous-parties de la base, ont été inventées depuis une trentaine d'années pour tenter de pallier ces problèmes, donnant lieu à plusieurs centaines de publications par an dans les revues spécialisées. Nous ne traitons pas ici des techniques invisibles à l'utilisateur comme les techniques d'indexage de documents, le Web sémantique ou les ontologies. Ces techniques essaient toutes de décrire, de façon plus ou moins automatisée, le contenu d'un document, de façon plus « intelligente » que comme la somme des mots qui le composent. Mais ces techniques sont, à notre connaissance, très peu évaluées dans une perspective ergonomique.

Un effort particulièrement important a été fourni ces dernières années sur les techniques de retour des résultats : d'une part pour les outils de visualisation des résultats (qui peuvent par ex. être présentés de façon spatiale plutôt que comme une liste) et d'autre part, pour l'évaluation et la représentation de la pertinence des résultats (comme un nombre d'étoiles proportionnel à la pertinence). Le lecteur trouvera dans les articles rassemblés par Chen et Czerwinski (2000) une bonne présentation de ce domaine de recherche.

On peut être surpris de la rareté des études empiriques conduites sur les images non plus comme modalité de restitution des résultats mais comme documents : en quoi la recherche d'information parmi des textes diffère-t-elle de celle parmi les images ? Qu'est-ce qui est spécifique à la recherche d'images ? Quelques rares résultats sont en effet très intéressants : Choi et Rasmussen (2002) montrent que des critères de pertinence spécifiques (qualité et clarté de l'image) et non-spécifiques aux images (contenu thématique) sont conjointement

utilisés. Westman et Oittinen (2006) montrent l'influence du contexte de la tâche, c'est-à-dire de la façon dont va être utilisée l'image, sur les requêtes formées par des journalistes et des documentalistes dans une activité de rédaction d'articles journalistiques. Ces auteurs montrent aussi l'influence de facteurs affectifs dans les choix faits. Avant eux, Markkula et Sormunen (2000) avaient montré que les journalistes formulaient essentiellement leur requête en fonction d'objets contenus dans les images. Keister (1994) montre que la recherche d'images médicales mobilise deux grands types de descripteurs : la description de caractéristiques d'images (qui est spécifique à la recherche d'images), et le contenu thématique (qui n'est pas spécifique). La confrontation des études conduites dans les domaines journalistiques et médicaux semblent montrer encore une fois l'influence de l'exploitation visée du document recherché (par ex., chercher une image pour illustrer ou pour comprendre) sur la conduite de la recherche d'information. Ainsi, Jorgensen et Jorgensen (2005) montrent la grande fréquence de l'utilisation de stratégies exploratoires, par tâtonnements, dans la recherche d'image par des professionnels de l'image comme les publicitaires ou les infographistes. De même, les présentations cartographiques des résultats (Kartoo.fr étant l'un des outils les plus connus en la matière) semblent très coûteuses du point de vue cognitif (Ciaccia, Ihadjadene, & Martins, 2006).

Beaucoup de travaux se tournent plus récemment vers de nouvelles modalités d'interaction avec les systèmes d'information par le biais des techniques de réalité virtuelle ou de réalité augmentée, sans résultats ergonomiquement probants pour l'instant (Westerman, Cribbin & Wilson, 2001 ; Azuma, 2001). L'effet des modalités sensorielles mobilisées par l'interaction est aussi, curieusement, très peu étudiée : qu'est-ce qui différencie la recherche d'information avec un système dont l'interaction mobilise l'audition (par ex., serveurs vocaux) plutôt que la vision (bases de données textuelles classiques) ?

On sait par ailleurs, que l'ajout de fonctionnalités détériore souvent la performance de l'utilisateur, c'est une vieille « loi » de l'ergonomie (Eason, 1984). Lazonder, Biemans et Wopereis (2000) montrent que quand ces fonctionnalités sont apprises puis maîtrisées, elles deviennent efficaces.

En résumé, pour comprendre la particularité des documents électroniques, il faudrait connaître précisément les caractéristiques techniques de ces documents (structure, fonctionnalités, formats de

présentation) mais aussi les tâches pour lesquelles ils sont utilisés et les utilisateurs. Le triplet classiquement utilisé en ergonomie (individu – tâche – artefact) semble donc bien être un cadre d'analyse encore une fois pertinent.

3.2. LES FACTEURS ET CONTRAINTES LIÉS À LA RECHERCHE D'INFORMATION DANS LES ENVIRONNEMENTS ÉLECTRONIQUES

3.2.1. Les différences inter-individuelles

Bien que le domaine de l'étude de l'activité de recherche d'information dans les documents soit récent, des travaux ont été conduits dans une perspective différentielle.

Certains travaux concernent des différences entre individus non-spécifiques à la recherche d'information mais pouvant avoir un effet sur elle, comme la motivation ou le niveau en lecture. C'est le cas par exemple de David, Song, Hayes et Fredin (2007) qui montrent l'effet positif du sentiment d'auto-efficacité et de la motivation intrinsèque sur la réussite de la recherche d'information dans un hypertexte.

La lecture et la compréhension étant impliquées dans l'activité de recherche d'information, comme le soulignent en particulier les modèles ESP et CoLiDes évoqués plus haut, il semble évident que l'activité de recherche d'information soit profondément influencée par le niveau en lecture – compréhension. Le lecteur trouvera chez Rouet et Eme (2002) une bonne synthèse sur la question, dans une perspective développementale.

D'autres travaux concernent des aspects plus particuliers de la cognition humaine tels que la dissonance cognitive et leurs effets sur la recherche d'information. C'est le cas de ceux de Lin, Lee et Yen (2004), dont les résultats montrent que la recherche d'information concernant les lipides et cholestérol sur des étiquettes d'aliments est moins probable parmi les individus qui consomment le plus de lipides et de cholestérol. De même, des travaux en psychologie du raisonnement (par ex., Jonas, Schulz, Frey, & Thelen, 2001), montrent que plus on tient à une décision, à une opinion, à une position, moins on ressent le besoin d'information (potentiellement contradictoire) à son propos. Plus on est soumis à des informations contradictoires que l'on refuse et plus on tient à sa décision. Comprendre pourquoi les

différences entre individus les conduisent à chercher ou pas telle information semble donc pouvoir être éclairé par la psychologie de l'engagement. Il n'y a en revanche pas de recherche à notre connaissance sur l'effet des capacités attentionnelles ou métacognitives sur la recherche d'information.

D'autres recherches tentent de rendre compte de différences inter-individuelles spécifiques à la recherche d'information. Par exemple, Tricot, Drot-Delange, Foucault et El Boussarghini (2000), ont tenté de rendre compte de l'expertise en recherche d'information. Pour cela, ils ont présenté une étude de cas à 16 enseignants, dont certains étaient des utilisateurs réguliers du Web et d'autres des non-utilisateurs du Web. L'étude de cas fictif présentait un adolescent en proie à des difficultés personnelles et scolaires. A la suite de la présentation écrite du cas, 20 propositions étaient faites : des conclusions, des indications de conduite à tenir. Les participants devaient évaluer la pertinence de chacune pour permettre de mieux comprendre le cas. Ils étaient contraints en temps, et avaient à leur disposition un site Web d'enseignement de la psychologie de l'adolescent, comprenant approximativement 600 pages.

Les résultats montrent que les deux groupes ont sensiblement les mêmes performances au pré-test en psychologie de l'adolescent et à l'étude de cas. Les meilleures compétences à utiliser le Web des jeunes enseignants ne leur permettent pas d'obtenir une meilleure performance à l'étude de cas. L'étude de l'utilisation du site Web apporte des informations intéressantes. Il semblerait que les faibles compétences à utiliser le Web conduisent les participants à ne pas utiliser le site Web et à répondre par eux-mêmes aux questions. Réciproquement, les bonnes compétences à naviguer sur le Web conduisent ceux qui les possèdent à utiliser (mais très inefficacement) le site Web. Tout se passe comme si la majorité des participants avait eu un comportement essentiellement imprécis : ouvrir beaucoup de nœuds, quel que soit leur contenu. Les savoir-faire des utilisateurs réguliers du Web les ont conduits à utiliser cet instrument... mais pas à être efficaces. Le site Web avait été choisi parce qu'il est difficile à utiliser comme une encyclopédie : c'est un site Web d'enseignement, correspondant à un semestre de cours, et non pas à une recherche d'information limitée à 35 minutes. Il semblerait donc que la présence d'un instrument correspondant aux savoir-faire des individus puisse déclencher une utilisation, voire une sur-utilisation (une utilisation poussée mais inefficace) de l'instrument, quand bien même cet

instrument est faiblement utile. Quand l'instrument inadéquat ne correspond pas aux savoir-faire des individus, ces derniers décident de ne pas l'utiliser, ce qui est au bout du compte un comportement plus rationnel que celui des individus « compétents ». Ainsi, l'expertise dans le registre technique, contrairement à ce que suggère Marchionini (1995), ne permet pas, dans certains cas, d'obtenir de meilleures performances dans une tâche de recherche d'information. Des différences entre individus sont sans doute à chercher dans les registres techniques et de contenu, mais aussi dans d'autres, par exemple ceux liés au traitement de l'information et à la métacognition.

Dans le même ordre d'idée, Chevalier et Kicka (2006) montrent que des concepteurs de site Web peuvent obtenir une performance égale à celle de novices, voire une moins bonne performance que des utilisateurs familiers dans une tâche de recherche d'information. Mais, dans cette étude, l'analyse de l'activité des participants a révélé que si les concepteurs de site Web obtiennent une performance équivalente à celle des novices, leur démarche en revanche est très différente.

3.2.2. Les différences intra-individuelles

Un même individu peut tantôt être performant lors d'une recherche d'information et tantôt rencontrer de nombreuses difficultés, se perdre dans les documents et informations, voire abandonner son activité. Même si la question de ces différences individuelles lors de recherche d'information dans les environnements électroniques n'a pas reçu une grande attention de la part des psychologues, certaines études révèlent et expliquent l'existence de ces différences intra-individuelles.

3.2.2.1. Le rôle des contraintes et des consignes

Toute recherche d'information naît d'un besoin d'informations. L'individu doit donc comprendre la situation, c'est-à-dire construire une représentation mentale de la situation et des consignes, les contraintes spatio-temporelles étant déterminantes dans cette construction (par ex., temps disponible pour réaliser l'activité, type de système utilisé).

La représentation mentale élaborée par l'individu dépend étroitement de la question qui initie l'activité. Certains auteurs se sont spécifiquement intéressés à l'impact des questions qui initient les recherches d'information sur les comportements et stratégies des

individus (par ex., Dinet & Rouet, 2002 ; Wagner & Sternberg, 1987). Globalement, les résultats montrent que le temps de recherche d'information augmente quand augmente la difficulté de la question. Ainsi, une question qui exige la recherche et la comparaison d'informations issues de plusieurs sites Web s'avère plus complexe que la simple recherche visant à répondre à une question factuelle.

Bien évidemment, les connaissances initiales sur le domaine jouent un rôle primordial sur les stratégies de recherche d'information. Comme cela a été décrit plus haut dans la partie relative au modèle de Marchionini (*cf.* §2.1.1.), plus un individu possède de connaissances sur le domaine, c'est-à-dire le sujet de la recherche d'information, plus il emploie des stratégies centrées sur le thème et les informations contenues dans les documents (*content-driven*). Concrètement, plus un individu a de connaissances sur le thème de la recherche d'information et plus il est capable de produire un grand nombre de mots-clés associés à ce thème, plus il consacre de temps à analyser le contenu des documents et plus il évalue avec précision et rapidité la pertinence des informations qu'il trouve. Chez les enfants, les connaissances sur le domaine se traduisent par un vocabulaire plus riche pour mener les recherches d'information, moins de fautes d'orthographe dans leurs requêtes et une meilleure maîtrise des dictionnaires intégrés aux systèmes de recherche d'information (Borgman & Siegfried, 1992 ; Borgman *et al.*, 1999 ; Edmonds, Moore & Balcom, 1990 ; Solomon, 1993).

Selon les contraintes et consignes, un individu adapte donc ses comportements. Halttunen (2003) a ainsi identifié cinq stratégies selon le degré de précision des consignes et contraintes :

- l'identificateur du processus (*process identifier*), principalement centré sur les premières étapes de la démarche de recherche d'information telles que l'identification du besoin d'information, l'identification des sources d'information dont il dispose et le choix de la démarche ;
- l'identificateur des sources (*source identifier*), davantage centré sur l'identification et l'évaluation des sources d'informations à sa disposition ;
- le chercheur (*searcher*), surtout occupé à réaliser la recherche effective et à manipuler les moteurs de recherche ;
- le formalisateur de problème (*problem formulator*), principalement centré sur l'identification et la définition du besoin d'information ;

- l'évaluateur (*assessor*), surtout concentré sur l'évaluation des informations trouvées et dans une moindre mesure sur l'évaluation des sources d'information.

Aucun individu n'utilise qu'une et une seule stratégie : tout dépend de la consigne initiale. Comme le montre le tableau 2.2, certaines phases font l'objet d'une attention complète (cases en noire), modérée (cases en gris foncé), faible (cases en gris clair) ou quasi-nulle (cases blanches). A titre d'exemple, les individus identifiés comme « identificateurs du processus » se focalisent essentiellement sur les premières étapes de la recherche d'information (identification du besoin d'information, sélection des sources d'informations, recherche effective), mais prêtent tout de même une certaine attention aux trois étapes suivantes (acquisition d'information, évaluation de l'information et utilisation de l'information).

Profil	Phase					
	Besoin d'information	Sources d'information	Démarche de recherche d'information	Acquisition d'information	Evaluation de l'information	Utilisation de l'information
Identificateur du processus						
Identificateur des sources						
Chercheur d'information						
Formalisateur de problème						
Evaluateur de pertinence						

Tableau 2.2. – *Profils des chercheurs d'information selon Halttunen (2003)*

3.2.2.2. Les facteurs matériels et ergonomiques

Les matériels (périphériques et systèmes) disponibles peuvent expliquer en partie les difficultés ou du moins les différentes stratégies des individus. Quel que soit le système de recherche d'information utilisé, l'individu interagit généralement avec ce système via un clavier alphanumérique et une souris électronique. Or, ces outils sont parfois difficiles à utiliser surtout pour les jeunes (Dinet, 2003 ; Hooten, 1989 ; Solomon, 1993). L'implication des habiletés sensori-motrices est trop souvent écartée des considérations des chercheurs en psychologie cognitive dès lors qu'il est question de nouvelles

technologies. Pourtant, les capacités sensorielles (auditives mais surtout visuelles) et motrices jouent un rôle déterminant dans la navigation dans les documents numériques. Ainsi, certains auteurs (Donker & Reitsma, sous presse a, b ; Joiner et al., 1998) ont montré que des tâches simples (déplacer des objets à l'écran à l'aide de la souris, pointer le plus rapidement possible un objet à l'écran avec la souris, *etc.*) sont très difficiles à réaliser pour les jeunes utilisateurs. Le développement de ces habiletés est sous l'influence d'une part, d'aspects développementaux propres aux enfants et d'autre part, d'aspects extérieurs tels que l'entraînement. D'autres études tendent à montrer que l'utilisation du clavier et de la souris représente aussi des obstacles pour les personnes âgées (Battisti & Dinet, 2006 ; Rouet, 2003).

Plusieurs paramètres internes au document numérique peuvent perturber l'activité de recherche d'information, surtout chez les jeunes utilisateurs ou utilisateurs novices (Dinet, 2006a ; Hirsh, 1996, 2000) :

- les fonctionnalités offertes et permises et plus particulièrement leur nombre, leur hétérogénéité et la connaissance qu'en ont les utilisateurs ;
- l'utilisabilité des interfaces, le terme « utilisabilité » renvoyant ici surtout à des aspects de surface tels que la lisibilité, les couleurs utilisées et les icônes présents ;
- la structure du document, les plus courantes étant les structures linéaire et en réseau, le nombre de liens étant également important ;
- le contenu du document. Ainsi, le vocabulaire présent dans le document détermine en grande partie son utilisabilité.

Pour évaluer ces paramètres internes au document numérique, différentes grilles des critères ergonomiques sont disponibles (*cf.* chapitre 11 dans cet ouvrage). Une étude récente (Dinet, 2006b) a permis de montrer que de jeunes adolescents pouvaient être sensibilisés à ces critères ergonomiques en leur faisant procéder à l'examen de divers sites Web.

4. CONCLUSION

L'étude psychologique de la recherche d'information dans les documents est récente, encore peu développée, mais en plein essor.

Dans les sociétés de l'information et de la connaissance, les humains sont constamment en train de rechercher de l'information au sein de ressources chaque jour plus riches et plus nombreuses. Les difficultés qu'ils y rencontrent mettent en exergue l'importance de l'amélioration de ces ressources, amélioration qui passe par une meilleure connaissance de ces difficultés. Ainsi, la psychologie dans ce domaine a une composante ergonomique très forte.

Les réflexions dans le domaine ont été longtemps focalisées sur la formation avec le *leitmotiv* suivant : il faut former les individus à la recherche d'information. Mais, les formations proposées aux non documentalistes consistaient essentiellement en l'apprentissage de procédures utilisées par des experts. Or, les activités de ces experts, leurs connaissances déclaratives et surtout procédurales sont très différentes des personnes visées par ces formations. Cette impasse est aujourd'hui dépassée, avec l'ambition de développer une *information literacy* (une formation aux compétences informationnelles) dans la plupart des pays où l'industrie de l'information est importante.

La recherche en ergonomie des documents doit donc, selon nous, à présent chercher à répondre aux questions suivantes :

- comment aider l'utilisateur à gérer l'évolution de son but pendant la recherche d'information ?
- comment représenter les résultats d'une requête pour que l'utilisateur traite et comprenne ceux qui sont pertinents et non pas seulement les deux ou trois premiers ?
- comment permettre à l'utilisateur de choisir l'outil adapté à son but de recherche ?
- quel est l'utilité de chaque grande fonctionnalité (moteur, index, liens hypertextes, *etc.*) relativement aux tâches de recherche d'information ?
- comment adapter les ressources aux divers contextes professionnels dans lesquelles elles sont utilisées ?

Du côté de la psychologie de la recherche d'information, de nombreuses questions et zones d'ombres demeurent, comme celles liées au contexte collectif de la recherche d'information. En effet, même si certaines études s'intéressent à la dimension collective / collaborative de la recherche d'information dans les documents électroniques (Boubée, Tricot & Couzinet, 2005 ; Fidel, 1999 ; Hyldegard, 2006), les effets sont encore largement méconnus.

De même, la modélisation des tâches de recherche d'information est balbutiante. Or, quand on ne sait pas décrire des tâches dans un

domaine de recherche en psychologie ergonomique, on a forcément des difficultés à comparer les résultats empiriques obtenus. Récemment, Byström et Hansen (2005) ont proposé une taxonomie simple et pertinente d'un point de vue ergonomique, pour décrire la tâche de recherche d'information selon trois niveaux :

- le contexte de travail de la tâche (*work task*), qui est décrit en termes d'organisation, de ressources, de contraintes, de culture ;
- le contexte de la recherche d'information (*information seeking context*), qui correspond au besoin initial d'information ;
- le contexte de la recherche d'information (*information retrieval context*), qui correspond au traitement cognitif des documents, et notamment des résultats de la recherche.

C'est maintenant au sein de chacun des trois niveaux qu'il faut développer des outils de description suffisamment génériques et explicites pour pouvoir être utilisés par différents chercheurs.

Enfin, concernant les compétences et l'expertise en recherche d'information, la perspective proposée par Aula et Nordhausen (2006) semble prometteuse. Ces auteurs tentent en effet de caractériser l'expertise en recherche d'information sur le Web. Ils montrent que l'augmentation en années d'utilisation du Web est liée de façon linéaire à l'amélioration des performances en recherche d'information, que le but et thème soient très précis ou au contraire plus flous et relativement imprécis. Ils montrent que d'autres variables, telles que la fréquence hebdomadaire d'utilisation du Web, ont des effets significatifs sur différents aspects de la performance en recherche d'information : nombre de requêtes par minute, nombre de requêtes par questions, nombre de mots par requêtes ou temps passé à évaluer la qualité des résultats. Ainsi, ces auteurs tentent d'étudier la recherche d'information en caractérisant l'expertise, les tâches réalisées et les mesures de la performance. C'est sans doute ce type de recherche qui pourra apporter des connaissances plus directement exploitables en ergonomie des documents et des systèmes documentaires électroniques.

5. BIBLIOGRAPHIE

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ : Erlbaum.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036-1060.
- Aula, A., & Nordhausen, K. (2006). Modeling successful performance in Web searching. *Journal of the American Society for Information Science Technology*, *57*, 1678-1693
- Azuma, R.T. (2001). Augmented reality : Approaches and technical challenges. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 27-63). Mahwah, NJ : Erlbaum.
- Barès, M. (1984). *La recherche documentaire dans le contexte télématique : modalités d'automatisation et utilisation des bases de données*. Paris : Hachette
- Battisti, S. & Dinet, J. (2006). *La sélection de sites Web par les personnes âgées : Impact des compétences métamnésiques*. Communication à *JETCSIC'2006*, Université Paris-X Nanterre, 30 juin.
- Blackmon, M. H., Kitajima, M., & Polson, P. G. (2005). Tool for accurately predicting Website navigation problems, non-problems, problem severity, and effectiveness of repairs. *Proceedings of Conference on Human Factors in Computing Systems (CHI'2005)*, ACM Press, 31-40.
- Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the Web. *Proceedings of Conference on Human Factors in Computing Systems (CHI'2002)*, ACM Press, 463-470.
- Borgman, C. L., Hirsh, S. G., Walter, V. A., & Gallagher, A. L. (1999). Children's searching behavior on browsing and keyword online catalogs : The science library catalog project. *Journal of the American Society for Information Science*, *46*, 663-684.
- Borgman, C. L., & Siegfried, S. L. (1992). Getty's synonyme and its cousins : a survey of applications of personal name matching algorithms. *Journal of the American Society for Information Science*, *43*, 459-476.
- Boubée, N., Tricot, A., & Couzinet, V. (2005). *L'invention de savoirs documentaires : les activités de recherche d'information d'usagers dits « novice »*. Communication présentée au *Colloque « Enjeux et usages des TIC : aspects sociaux et culturels »*, Bordeaux, 22-24 septembre
- Byström, K. & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science Technology*, *56*, 1050-1061.

- Carroll, J. M., & Rosson, M. B. (1987). The paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought : cognitive aspects of human-computer interaction* (pp. 80-111). Cambridge, Mass : MIT Press.
- Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World Wide Web. *Computers Networks and ISDN Systems*, 27, 1065-1073.
- Chen, C. & Czerwinski, M., (Eds.), (2000). Empirical evaluation of information visualization. *International Journal of Human-Computer Studies*, 53(5).
- Chevalier, A., & Kicka, M. (2006). Web designers and Web users : Influence of the ergonomic quality of the Web site on the information search. *International Journal of Human-Computer Studies*, 64, 1031-1048.
- Choi, Y., & Rasmussen, E. M. (2002). User's relevance criteria in image retrieval in American history. *Information Processing and Management*, 38, 695-726.
- Ciaccia, A., Ihadjadene, M., & Martins, D. (2006). Évaluation des moteurs de recherche cartographiques. Analyse exploratoire. *Document Numérique*, 9, 111-125.
- David, P., Song, M., Hayes, A.F., & Fredin, E.S. (2007). A cyclical model of browsing: The dynamics of motivation, goals, and self-efficacy. *International Journal of Human-Computer Studies*, 65, 170-182.
- Dinet, J. & Rouet, J.-F. (2002). La recherche d'information : processus cognitifs, facteurs de difficultés et dimension de l'expertise. In C. Paganelli (Ed.), *Interaction homme – machine et recherche d'information* (pp. 133-161). Paris : Hermès.
- Dinet, J. (2003). La recherche documentaire informatisée à l'école : désactivation en mémoire et difficultés de sélection de références pertinentes. *Psychologie Française*, 48, 3-17.
- Dinet J. (2005). *La sélection collaborative de pages Web pertinentes*. Communication présentée à EIAH'2005, 25 - 27 mai, Montpellier.
- Dinet, J. (2006a). Le choix des sites Web par les enfants et adolescents : impact de la typographie. In A. Piolat (Ed.), *Lire, écrire, comprendre, apprendre avec Internet* (pp.135-150). Marseille : Éditions Solal.
- Dinet, J. (2006b). L'ergonomie des sites Web pour et par les jeunes. *AC-TICE*, 44, 34-36.
- Donker, A., & Reitsma, P. (sous presse, a). Aiming and clicking in young children's use of the computer mouse. *Computers in Human Behavior*.
- Donker, A., & Reitsma, P. (sous presse, b). Drag-and-drop errors in young children's use of the mouse. *Interacting with Computers*.
- Dreher, M. J., & Guthrie, J. T. (1990). Cognitive processes in textbook chapter search tasks. *Reading Research Quarterly*, 25, 323-339.
- Eason, K. D. (1984). Towards the experimental study of usability. *Behavior & Information Technology*, 3, 133-143.

- Edmons, L., Moore, P., & Balcom, K.M. (1990). The effectiveness of an online catalog. *School Library Journal*, 10, 28-32.
- Ericsson, K. A., & Kintsch, W. (1995). Long term working memory. *Psychological Review*, 102, 211-245.
- Fidel, R. (1999). A visit to the information mall : web searching behavior of high school students. *Journal of the American Society for Information Science*, 50, 24-37.
- Foss, C. (1988). Effective browsing in hypertext systems. *Proceedings of RIAO'88 Conference, User oriented context-based text and image handling* (pp.82-98). Cambridge : MA Press.
- Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, 52, 195-242.
- Gray S.H. (1990). Using protocol analyses and drawing to study mental model construction during hypertext navigation. *International Journal of Human-Computer Interaction*, 2, 359-377.
- Guthrie, J. T. (1988). Locating information in documents : examination of a cognitive model. *Reading Research Quarterly*, 23, 178-199.
- Guthrie, J. T., Briten, T., & Barker, K. G. (1991). Roles of document structure, cognitive strategy and awareness in searching for information. *Reading Research Quarterly*, 26, 300-321.
- Halttunen, K. (2003). Students' conceptions of information retrieval : Implications for the design of learning environments. *Library & Information Science Research*, 25, 307-332.
- Hirsh, S. G. (1996). Complexity of search tasks and children's information retrieval. *Proceedings of the 59th American Society for Information Science Annual Meeting*, 33, 47-51.
- Hirsh, S. G. (2000). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, 50, 1265-1283.
- Hooten, P. A. (1989). Online catalogs : Will they improve children's access ? *Journal of Youth Services in Libraries*, 2, 267-272.
- Hyldegard, J. (2006). Collaborative information behaviour: exploring Kuhlthau's information search process model in a group-based educational setting. *Information Processing Management*, 42, 276-298.
- Jacobson, F. F., & Ignacio, E. N. (1997). Teaching reflection : information seeking and evaluation in a digital library environment. *Library Trends*, 45, 771-802.
- Joiner, R., Messer, D., Light, P., & Littleton, K. (1998). It is best to point for young children : a comparison of children's pointing and dragging. *Computers in Human Behaviors*, 14, 513-529.

- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions : An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80, 557-571.
- Joo, S., & Grable, J. E. (2001). Factors associated with seeking and using professional retirement-planning help. *Family and Consumer Sciences Research Journal*, 30, 37-63.
- Jorgensen, C., & Jorgensen, P. (2005). Image querying by image professionals. *Journal of the American Society for Information Science & Technology*, 56, 1346-1359.
- Keister, L. H. (1994). User types and queries: Impact on image access systems. In R. Fidel et al. (Eds.) *Challenges in Indexing Electronic Text and Images*. (pp. 7-22). Medford, NJ: Learned Information.
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A comprehension-based model of Web navigation and its application to Web usability. In S. McDonald, Y. Waern, & G. Cockton (Eds.), *Proceedings of HCI 2000, People and Computers XIV – Usability of Else* (pp.357-373). New-York : Springer.
- Koenemann, J., & Belkin, N. J. (1996) A case for interaction: A study of interactive information retrieval behavior and effectiveness. *Proceedings of CHI'96* (pp. 205-212). New York : ACM Press.
- Kuhlthau, C. C. (1985). *Teaching the library research process*. West Nyack, New-York : Scarecrow Press.
- Kuhlthau, C. C. (1988). Longitudinal case studies of the information search process of users in libraries. *Library and Information Science Research*, 10, 19-25.
- Kuhlthau, C. C. (1991). Inside the search process : information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42, 361-371.
- Kuhlthau, C. C. (1993). *Seeking meaning : a process approach to library and information services*. Norwood, NJ : Ablex Publishing Corp.
- Kuhlthau, C. C. (1994). Impact of the Information Search Process model on library services. *Research Quarterly*, 34, 21-26.
- Kuhlthau, C. C. (1997). Learning in digital libraries : an information search process approach. *Library Trends*, 45, 708-725.
- Lazonder, A.W., Biemans, H.J.A., & Wopereis, I.G. (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science*, 51, 576-581.
- Lemercier, C., Quaireau, C., & Terrier, P. (2003). Effet de la couleur sur la récupération d'informations dans un menu hiérarchique complexe. *Psychologie Française*, 48, 43-56.

- Lin, C. T., Lee, J. Y., & Yen, S. T. (2004). Do dietary intakes affect search for nutrient information on food labels ? *Social Sciences and Medicine*, 59, 1955-1967.
- Marchionini, G. (1991). Information seeking in electronic encyclopedias. *Machine-Mediated Learning*, 3, 211-226.
- Marchionini, G. (1992). Interfaces for end-user information seeking. *Journal of the American Society for Information Science*, 43, 156-163.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge, MA : Cambridge University Press.
- Markkula, M., & Sormunen, E. (2000). End-User searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 4, 259-285.
- Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in fulltext end-user-oriented search systems : the role of domain and search expertise. *Library and Information Science Research*, 15, 35-69.
- Mesters, I., Van den Borne, B., De Boer, M., & Pruyn, J. (2001). Measuring information needs among cancer patients. *Patient Education and Counseling*, 43, 255-264
- Millar, R., & Shevlin, M. (2003). Predicting career information-seeking behavior of school pupils using the theory of planned behavior. *Journal of Vocational Behavior*, 62, 26-42
- Miller, D. P. (1981). The depth/breadth tradeoff in hierarchical computer menus. In *Proceedings of the Human Factors Society* (pp. 296-300). Santa Monica : Human factors society.
- Molard, C. (1995). *Le langage de la documentation*. Collection Connaitre et Parler. Poitiers : Editions Méthodes et Stratégies.
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106, 643-675.
- Rouet, J.-F. (1990). Interactive text processing in inexperienced (hyper-) readers. In A. Rizk, N. Streitz & J. André (Eds.), *Hypertext: concepts, systems and applications* (pp. 250- 260). Cambridge : Cambridge University Press.
- Rouet, J.-F. (2003). *100 fenêtres sur Internet : rapport de fin de contrat*. Poitiers : Maison des Sciences de l'Homme et de la Société.
- Rouet, J.-F., & Eme, E. (2002). The role of metatextual knowledge in text comprehension : Some issues in development and individual differences. In P. Chambres, M. Izaute, & P. J. Maresco, (Eds.), *Metacognition: Process, Function and Use* (pp. 121-134). Kluwer Academic Press.
- Rouet, J.-F., & Tricot, A. (1998). Chercher de l'information dans un hypertexte. In A. Tricot, & J.-F. Rouet (Eds.), *Les hypermédias : approches cognitives et ergonomiques* (pp. 57-74). Paris : Hermès.

- Shannon, D. (2002). Kuhlthau's Information Search Process. *School Library Media Activities Monthly*, 19, 19-23.
- Solomon, P. (1993). Children's Information Retrieval Behavior : A Case Analysis of an OPAC. *Journal of the American Society for Information Science*, 44, 245-264.
- Tricot, A., (2006). Quelques préalables à un modèle cognitif de la recherche d'information et de l'apprentissage avec documents électroniques. In A. Piolat, (Ed.), *Lire, écrire, communiquer, apprendre avec Internet*. (pp. 441-462). Marseille : Solal.
- Tricot, A., Drot-Delange, B., Foucault, B., & El Boussarghini, R., (2000). Quels savoir-faire les utilisateurs réguliers du Web acquièrent-ils ? *Journal d'Intelligence Artificielle*, 14, 93-112.
- Wagner, R. K., & Sternberg, R. J. (1987). Executive control in reading comprehension. In B.K. Britton & S.M. Glynn (Eds.), *Executive control processes in reading* (pp. 1-2). Hillsdale, NJ : Lawrence Erlbaum Associates.
- Westerman, S. J., Cribbin, T., & Wilson, R. (2001). Virtual information space navigation: Evaluating the use of head tracking. *Behaviour & Information Technology*, 20, 419-426.
- Westman, S., & Oittinen, P. (2006). Image retrieval by end-users and intermediaries in a journalistic work context. In *Proceedings of the 1st international Conference on interaction in Context* (pp. 102-110). New York : ACM Press.
-