

Recherche d'information et apprentissage avec documents électroniques

André Tricot

Hypermédiat & Apprentissages, ERT 34, IUFM de Midi-Pyrénées
Laboratoire Travail & Cognition, UMR 5551 CNRS, EPHE, Université de Toulouse 2
56, avenue de l'URSS – 31 078 Toulouse cedex
andre.tricot@toulouse.iufm.fr

Résumé

Les analyses de la littérature empirique dans le domaine de la recherche d'information et de l'apprentissage avec des documents électroniques mettent essentiellement à jour des résultats incohérents. Il semble donc nécessaire de disposer d'un cadre théorique pour planifier de façon plus cohérente des expériences et interpréter les résultats obtenus. Plusieurs modèles cognitifs de la recherche d'information existent et présentent généralement une articulation faible avec l'apprentissage. Le modèle ACT-IF (Pirolli & Card, 1999) semble à l'heure actuelle le plus complet et le mieux formalisé. ACT-IF décrit les systèmes cognitifs engagés dans de la recherche d'information comme évoluant vers des états stables qui maximisent les gains de l'information, gains qu'il est possible d'exprimer selon un rapport bénéfice / coût. Pour autant, ACT-IF présente un certain nombre de limites, notamment en ce qui concerne le besoin d'information, l'anxiété et le stress, la planification de la recherche d'information et enfin, les différences experts-novices. L'analyse du modèle et de ses limites conduit à définir quelques préalables à un modèle cognitif de la recherche d'information et de l'apprentissage avec documents électroniques. La recherche d'information dans les documents électroniques serait, d'une part, une alternative à l'apprentissage par l'action, et d'autre part, un moyen d'apprendre.

1. Introduction

Il est possible de rechercher et de trouver des documents électroniques sur Internet, de traiter et de comprendre les textes qu'ils contiennent, pour en apprendre le contenu. Cette activité complexe est mise en œuvre dans différents contextes et étudiée avec diverses méthodes. Si bien que les analyses de la littérature empirique dans ce domaine mettent à jour des résultats incohérents (Dillon & Gabbard, 1998 ; Amadiou & Tricot, soumis). Pour tenter de comprendre et d'organiser les résultats empiriques dans ce domaine, il faut disposer d'un cadre théorique. Un tel cadre est ébauché dans ce chapitre. Il s'agit en quelques sortes d'un cahier des charges pour un modèle cognitif de la recherche d'information et l'apprentissage avec les documents électroniques.

Dans une première partie, quelques modèles cognitifs de la recherche d'information sont présentés, ainsi que leur articulation avec l'apprentissage. Ces modèles, pour nombreux et intéressants qu'ils soient, ne sont pas suffisamment formalisés pour être implémentés. Leur réfutabilité est faible.

Au contraire, le modèle ACT-IF (Pirolli & Card, 1999), qui est présenté dans la seconde partie du chapitre, est formalisé, implémenté et réfutable. Selon ce modèle, lui-même fondé sur l'analyse rationnelle d'Anderson (1990), les systèmes cognitifs engagés dans de la recherche d'information évoluent vers des états stables qui maximisent les gains de l'information, gains qu'il est possible d'exprimer selon un rapport bénéfice / coût. Une façon de maximiser le rapport bénéfice / coût est l'apprentissage.

Mais des résultats empiriques, qui sont rapportés dans la troisième partie du chapitre, montrent quatre limites importantes de ACT-IF. Le besoin d'information y est conçu de façon

trop sommaire. L'anxiété et le stress ne sont pas pris en compte alors qu'ils ont un effet important sur le déclenchement d'une recherche d'information. La planification de la recherche d'information et les erreurs dont elle peut faire l'objet ne sont pas correctement décrites. Enfin, les effets des différences experts-novices sur la manipulation du système d'information sont ignorées.

Après l'exposé de ces limites, une ébauche de modèle cognitif de la recherche d'information et de l'apprentissage est proposée dans la quatrième partie. La recherche d'information dans les documents électroniques est conçue comme un moyen d'apprendre, alternatif à l'apprentissage par l'action. L'apprentissage est lui-même conçu comme, d'une part, une solution adaptative pour optimiser la relation entre un individu et son environnement, d'autre part comme une alternative au stockage d'information dans des documents.

2. Les modèles cognitifs de la recherche d'information (RI) et leur articulation avec l'apprentissage

De nombreuses modélisations de l'activité humaine de RI sont aujourd'hui disponibles (voir Case, 2002 pour une recension plus complète que celle présentée ci-dessous).

2.1 La RI : séquence ou cycle ?

Une première tentative de modélisation psychologique de l'activité cognitive d'un individu en situation de recherche d'information a été proposée par Armbruster et Armstrong (1993), Dreher (1992), Guthrie et Mosenthal (1987) et Guthrie (1988). À quelques nuances près, ces auteurs sont d'accord sur le cycle suivant :

- (a) formation d'un but cognitif ;
- (b) sélection d'une catégorie d'informations ;
- (c) extraction d'information ;
- (d) intégration à l'information préalablement extraite ;
- (e) recommencement jusqu'à ce que le but soit atteint.

Plus récemment, des auteurs comme Sutcliffe et Ennis (1998) ont exploité cette idée d'activité séquentielle ou cyclique. Pour eux, la mise en œuvre de l'activité de recherche d'information dépendrait à la fois du type de système d'information et des connaissances antérieures de l'individu, que ces connaissances soient de contenu, instrumentales ou informationnelles. Le processus comporte quatre étapes :

- (a) identification du besoin ;
- (b) articulation - conceptualisation du besoin ;
- (c) formulation de la requête ;
- (d) évaluation du résultat.

Toutefois, l'évaluation empirique de ce modèle n'est pas concluante. Un des problèmes essentiels de ces approches vient du fait qu'elles expliquent difficilement pourquoi de nombreuses RI échouent. Elles n'expliquent pas comment la représentation du but informationnel et le traitement des informations non pertinentes peuvent interférer. Elles n'articulent pas l'activité de RI avec la tâche principale.

Il est possible de décrire l'activité de recherche d'information de façon plus rigoureuse, en mettant en relation, d'une part, un cycle évaluation sélection traitement et, d'autre part, les activités de planification, contrôle métacognitif et régulation (Rouet & Tricot, 1995, 1996, 1998). Cette description, appelée EST, articule la compréhension du contenu des documents et le processus de RI 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 RI dans un document électronique 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 le sujet. 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.

Le module évaluation du cycle 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 d'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, ...) 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 à la représentation 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, plus de généralité, des exemples, etc. 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, ...) ;
- 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 changer d'une autre manière.

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 (i.e. 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 (MDT) une représentation du but, mais aussi une représentation de la stratégie de sélection et une représentation du contenu traité. Si, pour une raison ou une autre, la gestion de l'activité de sélection ou de traitement entraîne une surcharge cognitive, alors la représentation du but est momentanément hors de la MDT. Par exemple, la présence simultanée de textes et du menu, qui fait apparaître la structure du document électronique, devrait faciliter le processus de sélection. Mais Rouet (1990) a montré qu'une pagination groupée a tendance à créer une désorientation chez l'utilisateur inexpérimenté : la nécessité de traiter aussi le contenu (compréhension d'unités thématiques) peut provoquer une surcharge cognitive qui rend la gestion du module d'évaluation difficile. Lors de la phase d'évaluation, la représentation du but est réactivée, avec très probablement une déformation ou une perte d'information. Ce phénomène, très largement étayé expérimentalement dans les études sur la mémoire, a été illustré par Gray (1990) dans le domaine de l'utilisation des hypertextes. Il est aussi probable (Ericsson & Kintsch, 1995) que cette déformation de la représentation du but soit considérablement amoindrie, voire absente, quand l'individu est expert du domaine (à la fois du contenu traité et de l'utilisation de ce type de système d'information : une étude en cours avec l'URFIST de Toulouse montre en effet que les documentalistes professionnels sont souvent simultanément experts en documentation, en RI et dans un domaine de contenu).

2.2 La RI : un dialogue ?

Mizzaro (1998) a consacré une revue de littérature recensant près de 160 articles proposant une définition de la notion de pertinence. Pour cet auteur, la pertinence est une relation entre deux entités :

- un document, des descripteurs de ce document et de l'information (ce qui est reçu par l'utilisateur) ;
- un utilisateur, avec un problème, une traduction de ce problème comme besoin d'information, une représentation du problème en termes informatifs, une requête « naturelle » et sa requête « formalisée ».

Chacune de ces entités est décomposée en trois registres : le domaine de contenu, l'exploitation l'information trouvée par l'utilisateur et l'environnement (de travail, d'apprentissage, de recherche, etc.). La pertinence est l'adéquation entre les deux entités (le document d'un côté, l'utilisateur de l'autre), pour les trois registres (domaine, exploitation, environnement). La RI est donc la mise en œuvre de cette activité. Mizzaro décrit remarquablement des caractéristiques du contexte de mise en œuvre de la RI, mais pas l'activité elle-même, pas plus qu'il ne décrit ses liens avec l'apprentissage (tel n'est d'ailleurs pas son propos).

Toutefois, le modèle de la RI de Mizzaro présente l'avantage d'être fondé sur le concept de

pertinence, selon une acception proche de celle de la linguistique pragmatique (Sperber & Wilson, 1989). Ainsi, l'interaction entre un humain et un document pourrait être modélisable comme un dialogue, le but de l'activité comme une recherche de pertinence et l'interprétation de certains aspects saillants du document (la liste des résultats d'une requête, les mots en gras, en couleur ou soulignés, etc.) comme des marqueurs de pertinence (des ostensions pour Sperber et Wilson). De nombreux résultats empiriques, notamment ceux de Diné (e.g. Diné, Passerault & Rouet, 2001) ou les miens (Tricot & Lafontaine, 2002) viennent étayer cette position. Tout se passe comme si, dans une tâche de recherche d'information, de nombreuses erreurs de manipulation de moteurs de recherche ou de documents électroniques pouvaient relever d'erreurs de dialogue, et plus précisément d'erreurs d'interprétation des marqueurs de pertinence.

2.3 La RI : une activité à multiples niveaux de traitement ?

Dillon (1994, 1996) a lui aussi proposé une approche générale de la description de l'activité du lecteur de documents électroniques complexes, le cadre TIMS. À l'opposé des modélisations inspirées de Guthrie (1988), TIMS décrit les composantes de l'activité (que l'on peut considérer comme des niveaux de traitement) et non le déroulement de l'activité. Chaque niveau de traitement correspond à un type d'habileté différent, chacune de ces habiletés peut être plus ou plus bien développée chez l'individu, en fonction notamment de son expertise dans d'autres domaines (e.g. lecture, planification).

Les quatre composantes ou niveaux de traitement de TIMS sont :

- La gestion de la tâche : les lecteurs lisent l'information dans les systèmes d'information dans un but (apprendre par exemple). Pour atteindre ce but, ils doivent décider ce qu'ils veulent extraire du système d'information. Ce but peut évoluer en cours de consultation.
- Le modèle informationnel : les lecteurs possèdent une représentation mentale de la structure du document (structure au sens formel et au sens rhétorique). À partir du moment où ils identifient le type de structure du document, ils instancient une structure particulière du document qu'ils lisent.
- Les habiletés à manipuler le système d'information : le lecteur manipule des objets et des fonctions pour manipuler le système d'information. Ceci est particulièrement crucial et difficile dans les systèmes d'information nouveaux qui n'obéissent pas encore à des conventions (comparativement, un enfant apprend dès l'école primaire à manipuler un livre).

- L'activité de lecture du texte : classique, elle n'appelle aucun commentaire particulier. TIMS présente l'intérêt d'être applicable. Dillon montre que TIMS peut accompagner une activité de conception ergonomique de documents électroniques. Ce cadre est étayé par de nombreux résultats empiriques et une grande pratique du développement de produits. La psychologie cognitive, en relation avec d'autres disciplines, tente donc depuis une vingtaine d'années, de modéliser l'activité de recherche d'information dans les documents électroniques et les systèmes d'information ouverts tel Internet. Ces modèles ne sont pas suffisamment formalisés pour être implémentables et considèrent l'apprentissage, au mieux, comme un processus de compréhension. Le modèle ACT-IF, plus récent, présente la particularité d'être formalisé et implémentable (donc réfutable). Il aborde aussi l'apprentissage comme un processus général, non limité à la compréhension.

3. Le modèle ACT-IF

3.1 Rapide description du modèle

Pirolli et Card (1999) fondent leur théorie de la fouille d'information (information foraging) sur l'hypothèse suivante : quand cela est faisable, le système d'information naturel évolue

vers des états stables qui maximisent les gains de l'information, gains qu'il est possible d'exprimer selon un rapport bénéfice / coût. La recherche d'information chez les humains présenterait des analogies avec la recherche de nourriture chez les animaux que des travaux en éthologie ont mis à jour au milieu des années 1980 selon une perspective évolutionniste. Les systèmes cognitifs engagés dans de la recherche d'information affichent des tendances adaptatives. L'analyse rationnelle, déjà utilisée par Anderson dans sa modélisation de la cognition (Anderson, 1990) ou de la mémoire – apprentissage (Anderson & Milson, 1989), sert aussi de référence à l'élaboration de la théorie de la fouille d'information. Le problème principal de la collecte et de la compréhension de l'information réside dans l'allocation d'attention par l'utilisateur. Le coût de l'information est fonction, d'un côté, du fournisseur, du coût du stockage et de l'accessibilité, d'un autre côté, de l'utilisateur, du temps d'accès, du coût de la fouille et du traitement. Sa valeur peut être définie comme son utilité à la réalisation d'une tâche en cours. L'activité de fouille d'information chez le sujet humain obéirait à un principe de maximisation du rapport bénéfice / coût, non seulement dans le choix des sources mais aussi des interfaces qui donnent accès à l'information. Une façon de maximiser ce rapport est l'apprentissage.

Comme dans la plupart des modèles adaptatifs ou évolutionnistes, la maximisation du rapport bénéfice / coût, et plus généralement la rationalité, sont limitées : toutes les possibilités ne sont pas traitées, notamment parce qu'elles ne sont pas connues. L'adaptation est une dynamique locale.

Le modèle présenté par Pirolli et Card est fondé sur le développement d'équations d'optimisation de la fouille d'information : il s'agit du développement assez complet des termes de l'équation valeur = bénéfice / coût appliqué au domaine de la recherche d'information. L'engagement dans une RI est fonction de la valeur R de l'information recherchée

$$(1) R = \frac{G}{T_B + T_W}$$

avec G = bénéfice (i.e. utilité de l'information à la réalisation de la tâche) ; T_B = temps passé à chercher un document ; T_W = temps passé à traiter le contenu du document sélectionné. Puis les auteurs présentent ACT-IF, leur modèle ACT appliqué à la fouille d'information (adaptive control of thought in information foraging), à partir de l'analyse des protocoles d'utilisation d'un navigateur. Un ensemble de règles de production est proposé. L'évaluation de G à travers son titre, son résumé c sur un écran s, est notée :

$$(2) g(c, s) = \exp\left(\frac{\sum_{i \in Q} A_i}{T}\right)$$

La somme étant celle de l'activation des mots i de la requête Q ; T étant une estimation des documents pertinents dans le système. L'activation A d'un mot i de la requête Q est

$$(3) A_i = B_i + \sum_j W_j S_{ji}$$

avec B_i le niveau de base d'activation de i (le logarithme de sa fréquence) ; S_{ji} la force de l'association entre le mot j à l'écran et du mot i de la requête Q (le logarithme de la probabilité conditionnelle pour le mot j d'apparaître dans le contexte du mot i) ; W_j le niveau de base de l'activation de j.

En résumé, la probabilité pour que l'individu sélectionne un document est fonction (a) de la disponibilité du titre ou du résumé et (b) de la proximité lexicale entre le titre ou le résumé de ce document et la requête (ou le besoin) de l'individu.

Le modèle intègre le temps : si le titre de X est placé avant le titre de Y dans la liste, cela diminue la valeur de Y.

Filant la métaphore de la recherche alimentaire, Pirolli et Card (1999) conçoivent que

« l'homme informationnel » consomme de l'information pour survivre, et réciproquement l'information consomme des comportements humains pour survivre. Par exemple, pour exister sur Google, un site Web a besoin d'être visité par des humains.

D'autres études dans le domaine de la recherche d'information ont adopté une perspective adaptative, comme celle de Morrison et Vancouver (2000) ou celle de Smith, Venkatraman et Dholakia (1999). Ces derniers ont étudié les relations entre RI dans les mémoires artificielles et la mémoire naturelle, en termes de conflits et prises de décision de rechercher une information. Les auteurs se demandent si le coût d'une RI dans une mémoire artificielle est le même quand l'individu a ou n'a pas de connaissances préalables relativement au thème de la RI. La mesure de coût cognitif montre que les individus avec des connaissances ont une meilleure performance en RI mais seulement quand celle-ci est cognitivement coûteuse, c'est-à-dire quand les sources à traiter sont complexes, exigeantes à traiter. Contrairement aux novices, ils n'hésitent pas alors à s'engager dans des démarches complexes. La perspective adaptative apparaît donc comme prometteuse dans le domaine de la modélisation des processus de RI et particulièrement dans l'analyse des décisions de rechercher une information. Le modèle ACT-IF prend en compte les connaissances de l'individu, son besoin d'information, l'information disponible dans le système où l'individu recherche et la disposition de la cette information sur l'interface, et enfin, le temps.

3.2 Limites du modèle ACT-IF

Le modèle ACT-IF ne répond pas à quatre grandes questions : le besoin d'information, l'anxiété et le stress, la planification de la RI et enfin, les différences experts-novices.

3.2.1 Qu'est-ce que le besoin d'information ?

En identifiant le besoin d'information à la représentation (linguistique) du besoin d'information, ACT-IF ne rend pas compte de deux aspects essentiels du besoin d'information : il est parfois implicite et, loin de s'épuiser quand il est satisfait, il s'auto-alimente.

Le besoin d'information est une notion centrale tout autant que discutable dans le domaine de la RI (Case, 2002 ; Tricot, 2004). Il a été défini, au cours de l'histoire des sciences de l'information, comme :

- (a) rechercher une réponse à sa question (Taylor, 1968) ;
- (b) réduire l'incertitude (Atkin, 1973) ou un état de connaissance insatisfaisant (Belkin & Robertson, 1976) ;
- (c) donner du sens (Dervin, 1983).

Il peut être défini comme la prise de conscience par un individu d'un manque de connaissance pour atteindre un but dans une situation. Le besoin d'information est instrumental, car il implique la recherche du but visé. Il est secondaire, c'est-à-dire qu'il dépend d'autres besoins, plus élevés ou basiques (santé, autonomie, apprentissage, production, communication).

Mais le besoin d'information est contestable, dans le sens où ce que l'individu veut peut être différent de ce dont l'individu a besoin. Par exemple, un étudiant croit avoir besoin de lire intégralement 10 ouvrages pour préparer une synthèse, alors qu'il a besoin d'en lire les tables des matières. En outre, le besoin d'information n'est pas forcément conscient ou ne correspond pas forcément à une demande explicite de l'individu. La recherche en psychologie montre que pour prendre conscience d'un manque de connaissance, un individu doit posséder des connaissances, notamment des méta-connaissances (connaissances sur le thème, compétences documentaires, savoir-faire). Dans un domaine particulier, comme celui de l'orientation scolaire et professionnelle, on montre que les individus qui ont très peu de connaissances recherchent moins d'information que les individus qui ont déjà des connaissances. La décision de rechercher une information est influencée positivement par le fait d'avoir préalablement recherché de l'information (Millar & Shevlin, 2003). Toujours dans

ce domaine, Gati, Saka et Krauz (2001) montrent que les difficultés relatives à la prise de décision d'orientation sont d'autant plus importantes que le processus de décision est peu avancé et que l'individu manque d'information. Par ailleurs, dans le domaine de la recherche d'information sur les retraites, une étude menée aux États-Unis auprès de 711 travailleurs, montre que ceux qui recherchent le plus d'information sur le financement de leur retraite sont les plus informés et les plus riches (Joo & Grable, 2001). Enfin, dans le domaine scolaire, Van der Meij (1990) analyse les décisions de recherche d'information concernant des connaissances lexicales, de 84 élèves de fin d'école élémentaire. Il montre que les élèves avec peu de connaissances posent plus de questions, mais que ces questions indifférenciées et moins efficaces. Les élèves avec plus de connaissances posent plus de questions utiles et moins de questions inutiles. Ils sont plus capables de poser des questions spécifiques qui débouchent sur de l'information utile.

Le besoin d'information est donc paradoxal. Il correspond à une prise de conscience, dans une situation particulière, d'un manque de connaissance. Mais cette prise de conscience requiert des connaissances et des métaconnaissances. Plus encore, la satisfaction du besoin d'information ne l'épuise pas, au contraire elle le développe.

Si, comme Atkin (1973) on conçoit le besoin d'information comme un besoin de réduire l'incertitude, alors l'incertitude doit être définie non pas comme un manque de connaissances mais comme la prise de conscience d'un manque de connaissance. Le biais de confirmation, étudié en psychologie du raisonnement, illustre ce phénomène. Après avoir pris une décision ou émis un jugement, les individus ne recherchent pas d'information qui pourrait contredire cette décision. Ils ont même tendance à ne pas traiter ces informations quand elles leur sont présentées. Jonas, Schulz, Frey et Thelen (2001) montrent que ce biais est particulièrement fort quand l'information est présentée de façon séquentielle (alors qu'il est classiquement attesté dans des présentations simultanées). Pour ces auteurs, ce biais provient d'une focalisation de plus en plus importante des individus sur leur décision. Plus on tient à une décision, à une opinion, à une position, moins on a besoin de l'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.

Ainsi les processus cognitifs et métacognitifs qui conduisent l'individu à prendre conscience qu'il a besoin d'une information sont complexes et encore pas complètement élucidés. La prise de conscience du besoin d'information serait un processus métacognitif, d'autant plus précis qu'il est supporté par des connaissances importantes dans le domaine où s'exprime ce besoin. Ce processus métacognitif de diagnostic d'un manque de connaissances dans une situation peut être décrit comme le développement d'une incertitude. Cette incertitude peut être freinée par l'engagement de l'individu dans une opinion, une position, une décision. De nombreux résultats empiriques (Raufaste, 2001) montrent que l'augmentation de connaissances et de métaconnaissances s'accompagne d'une augmentation de la capacité de l'individu à se désengager d'une première hypothèse ou interprétation de la situation. Ainsi, l'engagement dans une RI serait fonction de la valeur de l'information recherchée (R), dans le contexte de l'activité en cours, valeur à laquelle il faudrait rajouter les métaconnaissances de l'utilisateur et retrancher la « certitude » de celui-ci, c'est-à-dire son « engagement » vis-à-vis de ses connaissances dans le domaine d'activité.

3.2.2 Pourquoi l'anxiété et le stress entraînent-ils de la recherche d'information ?

ACT-IF ne prend en compte que des aspects cognitifs de la recherche d'information.

Or des aspects non-cognitifs, comme l'anxiété et le stress, peuvent déclencher un comportement de recherche d'information. C'est notamment le cas lors de l'annonce d'une maladie comme le cancer. L'information trouvée peut contribuer à faire baisser cette anxiété ou ce stress, en « rassurant » l'individu. L'anxiété et le stress peuvent donc provoquer un besoin d'information et diminuer lorsque le besoin est satisfait (Bar-Tal, Sivia Bar & Zisser,

2005). Une enquête de Mesters, van den Borne, De Boer et Pruyn (2001) menée auprès de deux échantillons de 498 et 133 patients atteints d'un cancer, montre que le besoin d'information est d'autant plus important que le patient est anxieux, dépressif et qu'il souffre psychologiquement. Le besoin d'information est orienté soit vers l'action (e.g. recherche d'aides) soit vers la connaissance de la maladie. Au cours du temps, les besoins d'information concernant la maladie et le traitement diminuent, alors que ceux concernant l'accès à des aides restent importants.

3.2.3 Pourquoi les individus font-ils des erreurs de planification ?

On peut émettre une critique classique envers ACT-IF : elle fait l'impasse sur les erreurs d'interprétation de la situation et de planification de l'activité.

Dans l'expérience de Tricot et al. (2000), nous avons montré que la recherche d'information, et le choix même de s'y engager, peuvent être non-rationnels, notamment à cause d'erreurs stratégiques. Certains individus décident d'utiliser un outil qui ne leur permettra pas de trouver l'information recherchée. Ils ne parviennent pas à évaluer la pertinence de l'outil et persistent plusieurs dizaines de minutes dans cette utilisation inefficace. Comparativement, d'autres individus, qui choisissent de ne pas rechercher de l'information mais de résoudre le problème qui leur est posé en raisonnant, font un choix plus rationnel. ACT-IF ne prend pas en compte de façon satisfaisante la composante évaluative de l'activité de recherche d'information. On peut raisonnablement émettre l'hypothèse selon laquelle l'engagement dans la RI dépend de l'évaluation de la valeur de l'information préalablement et en cours d'activité, c'est-à-dire l'évaluation du bénéfice et l'évaluation du coût. Cette hypothèse permettrait d'interpréter le résultat de Tricot et al. (2000) comme une erreur d'évaluation du coût de la RI.

3.2.4 Pourquoi les experts sont-ils plus lents et utilisent-ils moins les outils de guidage ?

Une dernière critique importante de ACT-IF concerne la façon dont il omet de rendre compte des effets de l'expertise sur la manipulation du système d'information.

Dans une étude empirique, Brand-Gruwel, Wopereis et Vermetten (2005) montrent que les experts passent plus de temps que les novices à définir le problème et à traiter le contenu des documents. En revanche, ces auteurs n'obtiennent pas de différence pour les autres aspects de l'activité. Avec des connaissances dans le domaine du contenu, plus de ressources sont disponibles pour engager une activité métacognitive d'évaluation du besoin de RI, pour élaborer le problème de RI et pour planifier l'activité. Mais il doit aussi arriver que l'individu ait plus de connaissances métacognitives pour mettre en œuvre cette activité. ACT-IF ne fournit pas un cadre satisfaisant pour rendre compte des effets de compensation mis en évidence par McDonald et Stevenson (1998). En effet, plus il a de connaissances dans le domaine de contenu moins l'individu est dépendant des aides à la navigation fournies par le document électronique. Au contraire, pour les individus ayant peu de connaissances dans le domaine de contenu, le coût de traitement des contenus est tellement élevé qu'ils doivent se contenter de se faire guider dans leur navigation. Prendre des décisions dans ce registre représenterait un surcoût.

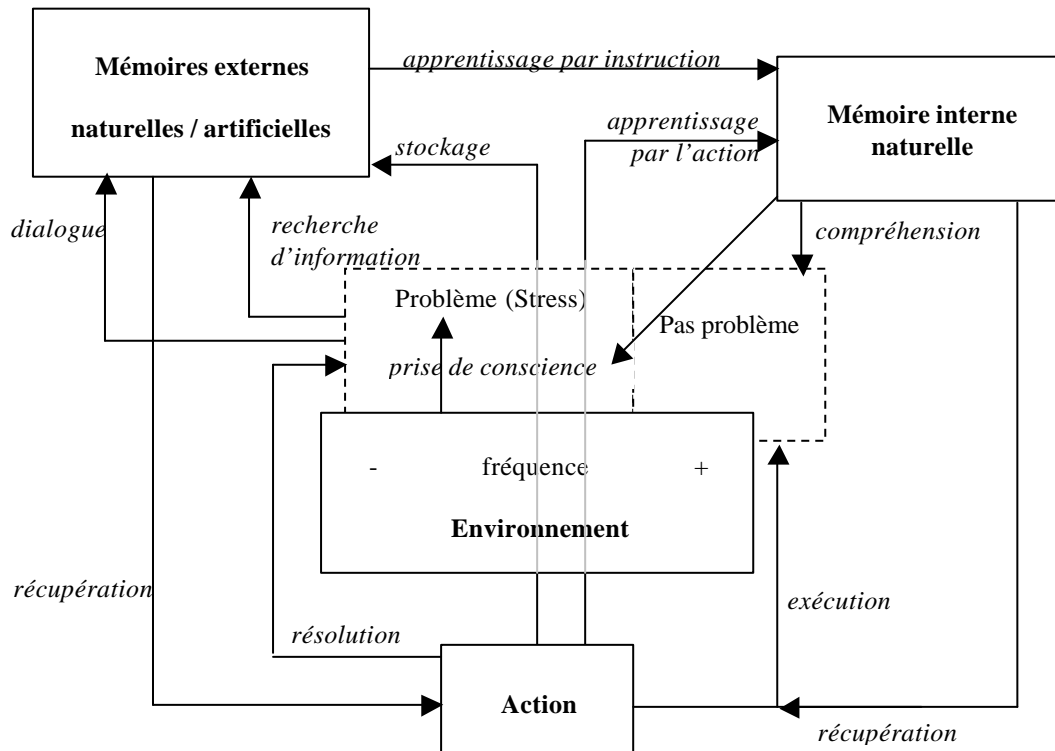
Les quatre limites du modèle ACT-IF, qui concernent le besoin d'information, l'anxiété et le stress, la planification de la RI ainsi que les différences experts-novices, permettent d'ébaucher une première synthèse des différents aspects de l'activité de RI et de leur articulation avec l'apprentissage.

4. Synthèse de l'articulation entre la RI et l'apprentissage

La synthèse présentée ici (Figure 1) tente de rendre compte de façon cohérente de l'ensemble des phénomènes qui ont été mentionnés plus haut. Elle met en relation quatre composantes principales : l'environnement de l'individu, son action, sa mémoire (qualifiée ici d'interne et

de naturelle) et enfin, les mémoires externes avec lesquelles il peut interagir, que celles-ci soient artificielles (les documents) ou naturelles (les individus). Les flèches entre ces quatre composantes correspondent à des processus cognitifs. Cette synthèse est conçue comme un complément au modèle ACT-IF. Il s'agit en quelques sortes d'un cahier des charges pour un futur modèle cognitif de la recherche d'information et de l'apprentissage avec des documents électroniques.

Figure 1. Schéma synthétique des relations entre RI et apprentissage



Si dans un environnement avec un but (une consigne venant d'un tiers), un projet (venant de lui-même) ou une contrainte (imposée par l'environnement), un sujet humain ne comprend pas et/ou ne sait pas comment faire, ce qui peut générer un stress, deux solutions sont possibles :

- la RI, c'est-à-dire la recherche d'une solution dans une mémoire externe,
- la résolution de problème, c'est-à-dire la recherche d'une solution par action sur la situation et interprétation du feedback renvoyé par les modifications de la situation.

Ces deux solutions impliquent que l'individu prenne conscience de son absence de compréhension et de son impossibilité d'organiser une activité adéquate à la situation. Ces trois activités (prise de conscience, RI et résolution de problème), impliquent la mobilisation de connaissances antérieures.

La plupart des tâches de la vie quotidienne peuvent être décrites comme des problèmes mal définis. Aussi, elles requièrent l'acquisition de connaissances, notamment à partir de sources externes. Il existerait différentes sortes de recherches de solutions dans des mémoires externes, dont la RI dans les documents (les mémoires externes artificielles) et le fait de poser une question à un tiers, d'engager un dialogue avec lui (les mémoires externes naturelles). La mise en œuvre d'une RI ou d'un dialogue requiert des savoir-faire. Il s'agit d'activités cognitivement coûteuses, sans résultat garanti. Les situations de résolution coopérative de problème ou d'apprentissage par conflit sociocognitif combinerait le dialogue (la RI dans des mémoires externes naturelles) et la résolution de problème. La RI peut aussi être utilisée

comme étape intermédiaire de la résolution de problème.

Si un environnement rare devient fréquent et si l'individu trouve régulièrement un moyen de satisfaire ses buts (projets, contraintes) en fonction de son environnement, alors il pourra acquérir des schémas puis les automatiser. Si au contraire l'environnement reste rare, il peut y avoir résolution (par l'action ou par recherche d'information) mais pas apprentissage. En résumé, l'apprentissage est une fonction adaptative.

Il est nécessaire d'avoir des connaissances pour prendre conscience du manque de connaissances, pour avoir de l'incertitude, même si, avec ou sans connaissances, l'individu peut avoir l'« illusion de savoir ». On peut émettre l'hypothèse que face à un environnement très rare, il n'y a pas de connaissances, donc pas de prise de conscience, mais que le risque adaptatif est relativement faible. On a des connaissances pour prendre conscience de problèmes susceptibles de devenir fréquents ou importants. Pour qu'il y ait problème et donc représentation d'un but à atteindre, il faut nécessairement disposer des connaissances qui permettent d'élaborer cette représentation. Dans le cas contraire, il n'y a pas de problème. Le reste doit donc être considéré comme accidentel, anecdotique... ou bien dangereux. En résumé, la prise de conscience est aussi une fonction adaptative.

Selon cette perspective, l'apprentissage par instruction (i.e. celui qui correspond à un enseignement) aurait lieu quand des individus provoquent un apprentissage alors que celui-ci n'est pas adaptatif. Les enseignants, l'école, l'éducation, anticiperaient que tel ou tel apprentissage, alors qu'il ne correspond pas à une adaptation à l'environnement, doit quand même avoir lieu, en prévision des environnements auxquels l'apprenant sera confronté dans le futur.

Quand un individu prend conscience qu'il manque de connaissances, il procède probablement à une évaluation du rapport bénéfice / coût des différentes possibilités qui s'offrent à lui, dont la possibilité de ne pas combler ce manque de connaissances. On peut décrire six grandes catégories de possibilités d'élaboration d'une solution au problème posé :

- la recherche d'une information ponctuelle qui sera utilisée uniquement dans cette situation ;
- la recherche d'une information qui sera maintenue dans une mémoire externe ;
- la recherche d'une information qui sera maintenue en mémoire interne ;
- la résolution ponctuelle du problème, dont la solution sera utilisée uniquement dans cette situation ;
- la résolution du problème et le maintien de la solution en mémoire interne si le problème se représente régulièrement (apprentissage implicite, apprentissage par l'action, par répétition) ;
- la résolution du problème et l'apprentissage délibéré c'est-à-dire le maintien en mémoire interne même si le problème ne se représente pas régulièrement (apprentissage explicite).

La prise de conscience d'un besoin de connaissances et l'évaluation du rapport bénéfice / coût d'une recherche d'information ou d'un apprentissage seraient des activités mentales influencées par des facteurs externes à l'individu (contraintes de la situation, présence d'un nouvel artefact dans la situation) et par ses connaissances antérieures (dans le domaine de contenu, compétences techniques et compétences informationnelles). Il est très probable que cette évaluation du rapport bénéfice / coût des différentes solutions possibles est calculée en fonction de représentations de la probabilité subjective d'atteindre le but, de façon absolue (est-ce que cela vaut le coût de rechercher à atteindre ce but ?) et de façon relative à la manière (cette manière est-elle moins coûteuse, plus sûre, que telle autre ?).

5. Conclusion

Il me semble donc que l'on puisse élaborer une théorie écologique de l'utilisation des

documents (ou de l'utilisation des mémoires artificielles). Confronté à une situation qu'il ne peut pas traiter automatiquement, l'individu décide ou non d'utiliser un document en se fondant sur une évaluation relative de la valeur de celui-ci (bénéfice / coût) comparativement à des alternatives (utiliser ses connaissances, raisonner, résoudre le problème par l'action). Seule une approche stricte et contextuelle du besoin d'information est défendable, car sinon le besoin d'information reste un concept ad hoc permettant d'expliquer le pourquoi des comportements de recherche d'information, un peu comme si l'on voulait expliquer un comportement meurtrier par un besoin de meurtre (Bosman & Renckstorf, 1996). La prise de conscience du besoin d'information et l'estimation de la valeur du document seraient d'autant plus justes que l'individu a des connaissances. La mise en œuvre de la recherche d'information serait d'autant plus efficace que l'individu a des connaissances. Une recherche d'information réussie peut générer de nouvelles connaissances pour l'individu, que selon la fréquence de la situation, il maintiendra disponible ou non. Les connaissances de l'individu concernent ici trois registres : domaine de contenu, technique, informationnel. Bref, la relation entre les mémoires artificielles et naturelles serait une relation que, faute de mieux, on peut qualifier de « potentialisation réciproque » : plus de connaissances => plus d'incertitude pertinente => plus de besoins d'information pertinents => plus de recherche d'information => plus d'apprentissage => plus de connaissances, etc. La régulation d'un tel système de mémoires résiderait sans doute plus dans la limite de temps, le nombre de situations différentes à traiter et celle des capacités de travail, que dans une limite de capacités du système de mémoires. En outre, certaines erreurs d'experts rappellent que parfois « plus de connaissances » n'entraîne pas « plus d'incertitude ».

Références

- Amadiou, F. & Tricot, A. (soumis). Utilisation d'un hypermédia et apprentissage : deux activités concurrentes ou complémentaires ? *Psychologie Française*
- Anderson, J.R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum.
- Anderson, J.R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703-719.
- Armbruster, B.B., & Armstrong, J.O. (1993). Locating information in text: A focus on children in the elementary grades. *Contemporary Educational Psychology*, 18, 139-161.
- Atkin, C. (1973). Instrumental utilities and information-seeking. In P. Clarke (Ed.), *New models for mass communication research* (pp. 205-242). Beverly Hills, CA : Sage.
- Bar-Tal, Y., Sivia Barnoy, S., & Zisser, B. (2005). Whose informational needs are considered? A comparison between cancer patients and their spouses' perceptions of their own and their partners' knowledge and informational needs. *Social Science & Medicine*, 60, 1459-1465
- Belkin, N.J., & Robertson, S. (1976). Information science and the phenomenon of information. *Journal of the American Society for Information Science*, 27, 197-204.
- Bosman, J. & Renckstorf, K. (1996). Information needs : problems, interests and consumption. In K. Renckstorf (Ed.), *Media use as social action* (pp. 43-52). London : John Libbey.
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*, 21, 487-508
- Case, D.O. (2002). *Looking for information. A survey of research on information seeking, needs, and behavior*. New York : Academic Press.
- Dervin, B. (1983). Information as a user construct : the relevance of perceived information needs to synthesis and interpretation. In S.A. Ward & L.J. Reed (Eds.), *Knowledge structure*

- and use : implication for synthesis and interpretation* (pp. 153-184). Philadelphia : Temple University Press.
- Dillon, A. (1994). *Designing usable electronic text: Ergonomic aspects of human information usage*. London: Taylor and Francis
- Dillon, A. (1996). Myths, misconceptions and an alternative perspective on information usage and the electronic medium. In J.-F. Rouet, J. Levonen, A. Dillon & R.J. Spiro (Eds.), *Hypertext and cognition* (pp. 25-42). Mahwah NJ: Erlbaum.
- Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: a review of the empirical literature on learner comprehension, control and style. *Review of Educational Research*, 68, 322-349.
- Dinet, J., Passerault, J.-M., & Rouet, J.-F. (2001). La recherche documentaire informatisée à l'école. In E. De Vries, J.-P. Pernin & J.-P. Peyrin (Eds.), *Actes du cinquième colloque Hypermédias et apprentissages* (pp. 135-150). Grenoble : EPI, INRP.
- Dreher M.J. (1992). Searching for information in textbooks. *Journal of Reading*, 35, 364-371.
- Ericsson, K.A., & Kintsch, W. (1995). Long term working memory. *Psychological Review*, 102, 211-245.
- Foss, C. (1988). Effective browsing in hypertext systems. *Proceedings RIAO'88 Conference, User oriented context-based text and image handling*, Cambridge, MA, 82-98.
- Gati, I., Saka, N., & Krausz, M. (2001). 'Should I use a computer-assisted career guidance system?' It depends on where your career decision-making difficulties lie. *British Journal of Guidance and Counseling*, 29, 301-321.
- 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., & Mosenthal, P. (1987). Literacy as multidimensional : locating information and reading comprehension. *Educational Psychologist*, 22, 279-297.
- 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.
- McDonald, S., & Stevenson, R.J. (1998). Effects of text structure and prior knowledge of the learner on navigation in hypertext. *Human Factors*, 40, 18-27.
- 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
- Mizzaro, S. (1998). Relevance: The whole history. In T. Bellardo Hahn & M. Buckland (Eds.), *Historical studies in information science* (pp. 221-244). Silver Spring, ML: ASIS.
- Morrison, E.W., & Vancouver, J.B. (2000). Within-person analysis of information seeking: The effects of perceived costs and benefits. *Journal of Management*, 26, 119-137
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106, 643-675.
- Raufaste, E. (2001). *Les mécanismes cognitifs du diagnostic médical : optimisation et expertise*. Paris: PUF.
- 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., & Tricot, A. (1995). Recherche d'informations dans les systèmes hypertextes : des représentations de la tâche à un modèle de l'activité cognitive. *Sciences et Techniques Éducatives*, 2, 307-331.
- Rouet, J.-F., & Tricot, A. (1996). Task and activity models in hypertext usage. In H. van Oostendorp & S. de Mul (Eds.), *Cognitive aspects of electronic text processing* (Advances in Discourse Processes, vol. 58, pp. 239-264). Norwood, NJ: Ablex Publishing.
- Rouet, J.-F., & Tricot, A. (1998). Chercher de l'information dans un hypertexte : vers un modèle des processus cognitifs. *Hypertextes et Hypermédias*, n° hors série, 57-74.
- Smith, G.E., Venkatraman, M.P., & Dholakia, R.R. (1999). Diagnosing the search cost effect: Waiting time and the moderating impact of prior category knowledge. *Journal of Economic Psychology*, 20, 285-314
- Sperber, D., & Wilson, D. (1989). *La pertinence. Communication et cognition*. Paris : Minuit.
- Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. *College and Research Science*, 13, 391-396.
- Tricot, A., & Lafontaine, J. (2002). Une méthode pour évaluer ensemble l'utilisation un outil multimédia et l'apprentissage réalisé avec celui-ci. *Le Français dans le Monde, Janvier*, 41-52.
- Tricot, A., (2004). La prise de conscience du besoin d'information : une compétence documentaire fantôme ? [en ligne, accédé le 19 septembre 2005 : <http://docsdocs.free.fr/modules.php?name=Sections&op=listarticles&secid=2>].
- 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.
- Van der Meij, H. (1990). Question asking : to know that you do not know is not enough. *Journal of Educational Psychology*, 8, 505-512.