La recherche d'informations dans les documents complexes : Processus cognitifs, apprentissage et développement

Jean-François Rouet, Béatrice Coutelet et Jérôme Dinet
Laboratoire Langage et Cognition (LaCo)
Centre National de la Recherche Scientifique et Université de Poitiers, France
jean-francois.rouet@univ-poitiers.fr

Conférence invitée à la Journée d'Etudes sur le
Traitement Cognitif des Systèmes d'Informations Complexes (JETCSIC)
Genève, 17-18 juin 2004

La recherche d'informations commence à émerger comme un thème de recherche central en psychologie et dans d'autres disciplines. Conséquence de l'explosion des systèmes d'informations, les gens sont de plus en plus souvent confrontés à des documents complexes voire à des bases de données bibliographiques ou autres, qu'ils doivent utiliser pour trouver les informations dont ils ont besoin. La recherche d'informations est aussi entrée dans la panoplie des activités pédagogiques pratiquées dans les classes primaires, secondaires et supérieures. Les enseignants ont en général l'impression que recherche d'information rime avec construction de connaissances, et que cette activité peut produire des savoirs plus "authentiques" que d'autres formes d'interactions avec les documents.

Cette popularisation croissante de la recherche d'informations (RI) fait pour le moment peu de cas des compétences requises par cette activité, des éventuels obstacles à sa mise en œuvre, ou tout simplement des conditions qui peuvent influencer le succès de la recherche. Or la recherche d'informations n'a rien de trivial, comme en témoignent les difficultés fréquentes rencontrées tant par les enfants que

par les adultes. Comment faisons-nous pour trouver une information particulière dans un texte ou un hypertexte ? Que faut-il savoir ou savoir faire pour chercher vite et bien ? Pourquoi les personnes ont-elles autant de difficultés à trouver les informations qu'elles cherchent ? Voilà quelques-unes des questions qui font l'objet de nos recherches depuis quelques années. Dans la suite de ce texte, je propose de faire le point sur les réponses que ces recherches nous ont apportées jusqu'ici.

1. Le processus de recherche d'informations

Lorsqu'elle s'exerce à partir de textes, la recherche d'informations entre d'emblée dans le répertoire des activités langagières, au même titre que la lecture, l'écriture ou le dialogue. L'ignorance un peu surprenante des psycholinguistes à l'égard de la recherche d'information peut sans doute s'expliquer en partie par la conception théorique de l'activité de compréhension qui domine depuis une vingtaine d'années. Comprendre, nous disent les théoriciens, c'est se construire un modèle mental de la situation décrite par le texte (par ex., Blanc & Brouillet, 2003; Kintsch, 1998; van Oostendorp & Goldman, 1999). Cette conception véhicule au moins trois présupposés :

(a) le texte contient une représentation nécessaire et suffisante d'une situation de référence;

(b) la personne qui se saisit d'un texte se donne pour but de comprendre cette situation;

(c) la personne lit intégralement le texte et construit une représentation mentale qui dérive en grande partie de la représentation verbale donnée par le texte.

Ces trois présupposés peuvent se résumer par la notion de "stratégie de cohérence": le lecteur cherche à obtenir une représentation cohérente à la fois avec ce qu'il sait du monde et avec les idées du texte. Cette conception de la compréhension n'a pas favorisé l'émergence des études sur la recherche d'informations, précisément parce que cette activité demande une stratégie totalement différente, que l'on pourrait qualifier de "stratégie de pertinence". Celui qui cherche une information ne souhaite pas nécessairement se construire une représentation de la situation décrite par le texte. Une partie de texte peut suffire, ou bien plusieurs textes peuvent être nécessaires pour réaliser la recherche. Le sujet doit donc se servir de sa représentation de la question ou du problème qui motivent la recherche pour déterminer quelle combinaison d'informations est la plus

appropriée. Cela peut le conduire à traiter les informations du texte à des degrés variables de profondeur, en fonction de l'utilité de l'information, et non pas en fonction de son importance au sein du texte. Lors d'une recherche d'information, l'individu peut être amené à traiter superficiellement des informations textuellement importantes, ou bien au contraire traiter de manière approfondie des détails textuellement mineurs.

Ce mode de traitement, sélectif et orienté par les objectifs préalables du lecteur, met l'accent sur deux aspects de l'interaction texte-lecteur qui, sans être étrangers aux situations de lecture plus courantes, prennent dans la recherche d'informations une importance particulière. Il s'agit d'une part de la gestion d'informations en mémoire de travail, et d'autre part de l'usage des organisateurs textuels. Nos recherches nous ont montré que la capacité à chercher des informations dépendait étroitement du degré de maîtrise de ces deux types de processus.

2. Recherche et gestion de l'information en mémoire de travail

Toute activité de compréhension de texte suppose la gestion de flux importants d'informations en mémoire de travail. Par exemple, l'adulte entraîné peut lire un texte au rythme de 200 mots par minute, c'est-à-dire environ 4 mots par seconde. La construction d'une représentation du sens du texte suppose a minima la création d'une "image" de chacun de ces mots en mémoire de travail. A ces mots "explicites" du texte s'ajoutent les connaissances activités en mémoire à long terme : concepts, images, traces perceptives et émotionnelles, qui, associées aux informations du texte, forment la représentation mentale. Dans la lecture "normale", le flux d'informations est contrôlé par la stratégie de cohérence (Kintsch, 1998, van den Broek et al., 1999). Le sujet cherche à garder le fil de l'histoire ou de l'explication, à garder une trace des différents plans (chronologique, causal, émotionnel, etc.) qui structurent le modèle de situation (Zwaan, 1999). Pour cela, il active ou inhibe sélectivement les significations lexicales (Gernsbacher, Varner & Faust, 1990), et produit les inférences nécessaires à la cohérence locale et, dans certains cas, globale (Albrecht & O'Brien, 1993).

Ces processus sélectifs d'activation, inhibition et inférence sont encore plus importants dans la recherche d'informations. En effet, si le lecteur d'un texte peut toujours parvenir à un "certain degré" de compréhension, le chercheur d'informations

possède en général un critère de réussite beaucoup plus précis : il lui faut en effet trouver une information précise, répondre à une question, ou réunir une quantité suffisante d'éléments pour traiter un problème. L'existence de ce critère (l'objectif de la recherche) représente un facteur de charge en mémoire de travail. Pour bien chercher, je dois me rappeler de ce que je suis en train de chercher. Selon la conception classique de la mémoire de travail, ce maintien doit logiquement occuper une certaine place, place qui n'est dès lors plus disponible pour les autres traitements requis par la recherche. A moins que le chercheur soit en mesure de "désactiver" temporairement sa représentation de l'objectif (de ne plus y penser), pour se concentrer, par exemple, sur le prochain choix à faire dans un menu, ou bien sur la compréhension d 'un passage difficile mais peut-être pertinent. Il lui faudra alors, à l'issue de cet épisode "d'oubli volontaire", être capable de réactiver sa représentation de l'objectif afin de poursuivre son activité. On voit alors que, davantage qu'un problème d'empan, la recherche d'informations pose le problème du contrôle volontaire de l'activation et de la désactivation d'informations en mémoire de travail.

Dans une série d'expériences récentes, nous avons étudié l'hypothèse selon laquelle cette exigence de contrôle cognitif serait à l'origine des difficultés importantes que rencontrent les enfants dans des tâches de recherche. Par exemple, Rouet, Coutelet et Chollet (subm.) ont observé qu'il fallait en moyenne quatre minutes à des enfants de 9 ans pour répondre à une question simple sur un animal à partir d'une encyclopédie. Lorsque la question demandait une inférence, ou la comparaison de deux animaux, le taux d'échec augmentait énormément. Ces résultats ne font que reproduire des observations similaires d'autres chercheurs (Kobasigawa, 1983; Raphael & Pearson, 1985).

L'idée qu'un problème de mémoire de travail soit à l'origine de ces difficultés est indirectement étayée par le constat plus général d'un lien entre mémoire de travail et compréhension chez les enfants de 8-10 ans. Yuill, Oakhill et Parkin (1989) ont par exemple établi que, à capacité de lecture constante, les mauvais compreneurs présentaient des déficits spécifiques des tâches sollicitant le traitement et la rétention simultanés de mots ou de nombres. Bien que d'autres chercheurs aient critiqué les procédures employées dans ces études (Nation et al., 1999), cette idée d'un lien entre mémoire de travail et compréhension en lecture est assez largement partagée (Seigneuric et al. 2000).

Dans des expériences en cours, nous avons établi un lien entre la capacité de mémoire de travail et la performance d'enfants de 10 ans à une tâche de recherche d'informations. La tâche consistait à rechercher, dans une liste d'une quinzaine de titres, celui qui correspondait le mieux à un thème de recherche donné. Afin d'estimer la capacité de contrôle attentionnel des enfants, nous avons utilisé une tâche non verbale, le "double barrage de Zazzo" (Zazzo, 1972). L'emploi d'une tâche non verbale coupe court à toute critique de tautologie (une tâche verbale prédit une autre tâche verbale).

Nous avons montré d'une part que la probabilité de trouver la bonne réponse dans la liste diminuait avec la position sérielle de cette réponse dans la liste. Les réponses placées en bas de la liste étaient moins bien détectées que les réponses placées en haut. Une interprétation possible est que la lecture des expressions non pertinentes entraîne une interférence qui "contamine" peu à peu la représentation de l'objectif. Les enfants ne savent plus exactement ce qu'ils cherchent, et ils ont donc des difficultés à le trouver. Par ailleurs, et c'est le résultat essentiel, nous avons montré que cet effet d'interférence était plus fort chez les enfants qui ont de mauvaises performances au test de Zazzo, ce qui suggère en lien entre le contrôle attentionnel et la recherche d'informations verbales.

Nous pourrions résumer ces résultats ainsi : pour devenir un "bon lecteur", l'enfant doit savoir décoder les mots, comprendre le sens des énoncés, et utiliser les textes pour y trouver les informations utiles. Ce troisième aspect de la maîtrise de l'écrit est dépendant d'une bonne capacité de contrôle attentionnel. Or le contrôle attentionnel se développe tout au long de l'enfance et de l'adolescence. Par conséquent, un certain nombre d'enfants et de préadolescents se trouvent gênés dès lors que la tâche de lecture repose sur une stratégie de pertinence.

3. Recherche et maîtrise des dispositifs métatextuels

Lorsqu'elle s'exerce au moyen de dispositifs textuels complexes (pages, ouvrages entiers, voire ressources électroniques…), la recherche d'informations demande l'utilisation de dispositifs qui signalent le contenu des documents examinés. Par exemple, la recherche sur le Web demande la lecture de menus, qui mènent à des sous-menus puis finalement aux pages d'informations recherchées. Les documents compelxes comportent toute une hiérarchie de signaux et de dispositifs verbaux et visuels permettant au lecteur de s'orienter vers l'information pertinente. Au

niveau local, on trouve les marqueurs typographiques, comme l'utilisation de tailles de caractères variables, l'utilisation du gras, du souligné, des lettres capitales etc. A un niveau intermédiaire, on trouve les titres, les introductions, les marques de paragraphe. Enfin, au niveau le plus global, on trouve les tables de matières, les index, les menus, les en-tête et bien d'autres dispositifs permettant d'avoir une vue global sur le contenu d'un document. Pour utiliser ces systèmes le lecteur doit en connaître le fonctionnement. Ce qui paraît trivial au lecteur adulte l'est nettement moins pour le jeune enfant. Par exemple, avant 10 ans seule une minorité d'élèves peuvent définir la fonction des titres et des tables de matières. Et même chez les lecteurs adultes, certaines notions avancées comme le thésaurus ou l'index n'ont pas toujours une représentation claire.

Nos études sur la recherche d'informations ont montré que l'usage des dispositifs métatextuels était un problème complexe. Certains résultats sont d'ailleurs assez paradoxaux. Par exemple, la présence de marques typographiques peut avoir un effet attracteur néfaste à la recherche d'informations. Dans sa thèse de doctorat, Jérôme Dinet a montré que les enfants tendaient à s'attacher aux indices de surface des informations à choisir (typographie et lexique) au détriment des indices sémantiques "profonds" (Dinet & Rouet, 2002; Dinet, Rouet & Passerault, 1999). Ces résultats ont été récemment répliqués et généralisés (Dinet & Rouet, in prep.). Mais d'un autre côté, on observe chez ces mêmes enfants une sous-utilisation d'autres dispositifs comme les titres ou les index. Rouet, Chollet et Coutelet (subm.) ont montré qu'une grande partie des élèves de 9 et 11 ans lisent de façon linéaire un texte dans lequel ils recherchent une information préciser. Il faut attendre l'âge de 13 ans pour observer des stratégies systématiques de lecture en diagonale appuyées sur l'utilisation des sous-titres. A cet âge, on observe que la capacité de lecture et les stratégies sont deux prédicteurs indépendants de la recherche d'informations. Là encore, ces résultats confirment des observations réalisées par ailleurs (par ex., Cataldo & Oakhill, 2000).

4. Conclusions

Jusqu'à une date récente, les chercheurs ne disposaient que de très peu de données empiriques pour analyser les processus cognitifs en jeu dans la recherche d'informations. Depuis quelques années, ce thème de recherche s'est développé, et on commence à avoir une idée plus précise de ces processus et des connaissances

qui leurs sont associés. La recherche d'information demande une approche du texte totalement différente de la simple lecture-compréhension. Le lecteur doit adopter une stratégie de pertinence, qui le conduit à hiérarchiser les informations non pas en fonction de leur importance textuelle, mais en fonction de leur pertinence par rapport à l'objectif de recherche. Nous avons montré que cette stratégie demande à la fois un bon contrôle attentionnel et l'usage raisonné des dispositifs métatextuels.

L'ensemble de ces résultats suggère que l'apprentissage de la recherche d'informations est sous la dépendance d'un facteur cognitivo-linguistique général et de deux facteurs spécifiques. Le facteur général est la capacité de lecture, mesurée par la rapidité et l'automaticité des processus de décodage. Cette capacité est partiellement liée à l'acquisition du vocabulaire, élément critique en recherche d'informations. Les deux facteurs spécifiques sont d'une part le développement des fonctions exécutives, qui permet à l'enfant de contrôler le flot d'informations en mémoire de travail et d'inhiber les informations non pertinentes, et d'autre part la connaissance des organisateurs textuels, qui permet de construire des stratégies plus économiques. Les données convergent pour indiquer que ces facteurs évoluent considérablement jusqu'à l'adolescence et qu'ils sont vecteurs de différences individuelles considérables y compris chez les jeunes adultes (Rouet, 2003).

Plus généralement, ces données nous invitent à donner une définition renouvelée de la compétence à comprendre les textes. Cette définition donne une place moindre aux stratégies basées sur la cohérence (Kintsch, 1998, van den Broek et al., 1999), et une place plus grande à un facteur jusqu'ici négligé, à savoir la pertinence des informations. Ces résultats ouvrent également un nouvel agenda de recherche qui devra se centrer sur la prise en compte des facteurs pragmatiques dans le traitement de l'écrit.

Références

Albrecht, J.E. & O'Brien, E.J. (1993). Updating a mental model: Maintaining both local and global coherence. Journal of Experimental Psychology: Learning, Memory and Cognition, 19, 1061-1070.

Blanc., N., & Brouillet, D. (2003). Mémoire et Compréhension. Paris : Armand Colin.

Cataldo, M.G. & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial

memory on the ability to locate information in text. Journal of Educational Psychology, 92, 791-799.

Dinet, J., & Rouet, J.-F. (2002). La recherche d'informations : processus cognitifs, facteurs de difficultés et dimensions de l'expertise. In C. Paganelli (dir.), Interaction homme-machine et recherche d'informations (p.133-161). Paris : Hermès Science.

Dinet, J., Rouet, J.-F. & Passerault, J.M. (1999). Recherche documentaire et nouvelles technologies de l'information: Les "nouveaux outils" sont-ils compatibles avec les stratégies cognitives des élèves? in J.F. Rouet & B. de La Passardière (sous la direction de). Actes du Quatrième Colloque Hypermédias et Apprentissages (pp. 149-162). Paris: INRP et EPI.

Dinet, J., & Rouet, J.-F. (in prep.). Rôle du marquage typographique dans la sélection d'informations pertinentes: Etude chez l'enfant et l'adulte. Manuscrit en préparation.

Gernsbacher, M.A., Varmer, K.R., & Faust, M.E. (1990). Investigating differences in general comprehension skill. Journal of Experimental Psychology: Learning, Memory and Cognition. 16(3), 430-445.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge, MA: Cambridge University Press.

Kobasigawa, A. (1983). Children's retrieval skills for school learning. The Alberta Journal for Educational Research, 29(4), 259-271.

Nation, K., Adams, J.W., Bowyer-Crane, C.A. & Snowling, M.J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. Journal of Experimental Child Psychology, 73, 139-158.

Raphael, T.E., & Pearson, P.D. (1985). Increasing students' awareness of sources of information for answering questions. American Educational Research Journal, 22, 217-235.

Rouet, J.-F. (2003). "What was I looking for?" The influence of task specificity and prior knowledge on students' search strategies in hypertext. Interacting with Computers, 15, 409-428.

Rouet, J.-F., Coutelet, B. & Chollet (2004). The acquisition of information search skills by 9 to 13 year-old students. Submitted for publication.

Seigneuric, A., Ehrlich, M.-F., Oakhill, J.V., & Yuill, N.M. (2000). Working memory resources and children's reading comprehension. Reading and Writing: An Interdisciplinary Journal, 13, 81-103.

van den Broek, P., Young, M., Tzeng, Y. & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. in H. van Oostendorp & S.R. Goldman (Eds.), The construction of mental representations during reading (pp. 71-98). Mahwah, NJ: Lawrence Erlbaum Associates.

van Oostendorp, H. & Goldman, S. (1999, Eds.). The construction of mental representations during reading. Mahwah, NJ: Lawrence Erlbaum Associates.

Yuill, N.M., Oakhill, J.V., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. British Journal of Psychology, 80, 351-361.

Zazzo, R. (1972). Manuel pour l'examen psychologique de l'enfant. Paris: Delachaux et Niestlé.

Zwaan, R.A. (1999). Five dimensions of narrative comprehension: The event-indexing model. In S. Goldman, A.C. Graesser & P. van den Broek (Eds.) Narrative comprehension, causality and coherence: Essays in honor of Tom Trabasso (pp. 93-110). Mahwah, NJ: Erlbaum.