

# **Les Activités Documentaires Complexes**

## **Aspects Cognitifs et Développementaux**

Rapport en vue de l'Habilitation à Diriger des Recherches

**Jean-François Rouet**

Chargé de Recherche au CNRS  
**Laboratoire Langage et Cognition**  
CNRS et Université de Poitiers

Avril 2001

## Remerciements

Ce travail est dédié à Stéphane Ehrlich, qui a su éveiller et stimuler mon intérêt pour la recherche. Je lui garde une sincère reconnaissance pour m'avoir montré, entre mille autres choses, comment concilier les exigences d'une démarche rigoureuse de psychologie scientifique et une réflexion plus générale sur les méthodes, les outils et les controverses qui traversent la pédagogie contemporaine. J'espère qu'il trouvera ce texte digne de son enseignement (et qu'il me pardonnera de n'avoir pas encore écrit un livre sur l'enseignement de la compréhension - mais nous en reparlerons).

Je voudrais également exprimer ici l'expression de ma reconnaissance envers Charles Perfetti pour son sens de l'amitié, et pour m'avoir fait partager sa grande expertise de la recherche lors de mon séjour au Learning Research and Development Center (Université de Pittsburgh). Une grande partie des travaux présentés dans cet ouvrage est le fruit de notre collaboration. Mes remerciements aussi à André Bisseret pour l'accueil bienveillant qu'il m'a fait dans son équipe de l'Institut National de Recherche en Informatique et en Automatique à Grenoble. J'ai appris auprès de lui une autre facette essentielle du métier de chercheur: l'art de dialoguer avec des partenaires industriels.

Eric Espéret m'a soutenu moralement durant mes séjours post-doctoraux (et dans mes efforts pour intégrer le Centre National de la Recherche Scientifique!), puis il m'a accueilli au Laboratoire Langage et Cognition lors de mon recrutement au CNRS. Je lui dois de m'avoir convaincu d'entreprendre un "véritable" rapport de synthèse plutôt qu'un simple coupé-collé de quelques articles. Je l'en remercie aujourd'hui.

Daniel Gaonac'h est pour moi depuis le début de cette entreprise (et de toutes les autres!) un directeur de laboratoire attentif et encourageant. Je lui dois de m'avoir encouragé à terminer ce travail. Je le remercie pour l'intérêt, l'attention et l'optimisme dont il fait preuve à l'égard de mes projets et j'espère que ce travail nous donnera de nouvelles occasions de collaborer.

A Pittsburgh, à Grenoble, à Poitiers et ailleurs j'ai toujours trouvé des collègues enthousiastes dont beaucoup sont cités ou se reconnaîtront dans les pages qui suivent. Je les remercie tous ici sincèrement, avec une mention spéciale à Alain, Béatrice, Daniel, Hervé, Jérôme, Ludovic, Marie-Françoise et Sylvie pour leur lecture attentive et leurs commentaires sur une version précédente de ce texte. Idem aux étudiants stagiaires sans qui les expériences relatées dans ce rapport n'auraient jamais été réalisées, aux enseignants qui m'ont ouvert les portes de leurs classes, ainsi qu'aux élèves, petits ou grands, qui ont accepté de travailler avec moi.

J'adresse enfin un remerciement tout particulier à Elsa pour le voyage fait ensemble, son soutien et ses encouragements durant la réalisation de ce travail.

## Table des matières

NB. Les astérisques signalent les parties du rapport qui présentent les travaux personnels de l'auteur.

<b><i>Introduction Générale</i></b>	<b><i>1</i></b>
<b><i>Chapitre 1 La représentation mentale des textes: état de la question et problématique</i></b>	<b><i>4</i></b>
1.1. La théorie propositionnelle de la compréhension	5
1.2. Les modèles de situation	8
1.3. Problèmes théoriques des conceptions actuelles de la compréhension	16
1.4. Synthèse	20
<b><i>Chapitre 2 Le rôle des connaissances métatextuelles dans la maîtrise de l'écrit</i></b>	<b><i>21</i></b>
2.1. Introduction	22
2.2. Activation des connaissances et compréhension de textes	23
2.3. Les connaissances métatextuelles	29 *
2.4. La mise en oeuvre des connaissances métatextuelles au cours de la compréhension	40
<b>2.5. Synthèse et conclusion</b>	<b>43</b>
<b><i>Chapitre 3. La compréhension des documents multiples</i></b>	<b><i>45</i></b>
3.1. Introduction	46
3.2. Compréhension experte et intégration de sources multiples	48
3.3. La théorie des représentations multidocumentaires	50 *
3.4. Propriétés des représentations multidocumentaires	60 *
3.5. Synthèse et conclusion	75
<b><i>Chapitre 4. La recherche d'informations dans les documents complexes</i></b>	<b><i>76</i></b>
4.1. Introduction	77
4.2. Compréhension de textes et traitement des questions	77
4.3. Un modèle cognitif de la recherche d'informations	87 *
4.4. Facteurs influençant les stratégies de recherche chez l'adulte	94 *
4.5. Le développement des capacités de recherche chez l'enfant	99 *
4.6. Synthèse et conclusion	104
<b><i>Chapitre 5. Aspects cognitifs des nouvelles technologies de l'information</i></b>	<b><i>106</i></b>
5.1. Introduction	107
5.2. Lecture et compréhension des hypertextes	109 *
5.3. Recherche d'informations dans les hypertextes	128
5.4. Synthèse et conclusion	136

<b>Chapitre 6. L'intégration multimédia</b>	<b>138</b>
6.1. Description, dépicition et modèles mentaux: aperçus notionnels	139
6.2. Illustrations et compréhension de textes	142
6.3. Les facteurs matériels d'intégration multimédia	146
6.4. Différences individuelles et développement des capacités d'intégration multimédia	150
6.5. Les animations graphiques complexes	155 *
6.6. Synthèse et conclusion	164
<b>Chapitre 7. Conclusions et implications pédagogiques</b>	<b>165</b>
7.1. Synthèse générale	166
7.2. La maîtrise de l'écrit revisitée	168
7.3. Apprentissages scolaires et technologies de l'information	170
<b>Bibliographie</b>	<b>173</b>

## Introduction Générale

La compréhension, c'est ce qui nous permet de nous représenter le monde qui nous entoure: les objets et leurs propriétés, les événements et leurs relations causales, les états psychologiques de nos congénères, les idées, les opinions, les théories bref: tout ce qui fait de nous des "bipèdes pensants". La compréhension intervient dans la plupart des activités cognitives. Elle peut être le fruit de l'expérience directe. Elle est aussi souvent le résultat du traitement d'informations symboliques, en particulier le langage écrit, cet instrument puissant par lequel la plupart des civilisations ont cherché à se constituer des mémoires externes capables de survivre aux éphémères mémoires individuelles, et de dépasser les limites inhérentes à la tradition orale.

Un lecteur expert peut lire un texte au rythme de 200 à 300 mots par minute, soit entre 3 et 5 mots par seconde (Baccino & Colé, 1995). Pendant ce temps, les mots sont identifiés, structurés en phrases, les idées sont extraites et connectées les unes aux autres de façon à former une représentation mentale de la situation. Le lecteur effectue de surcroît les inférences nécessaires pour établir ou compléter les liens logiques, temporels ou causaux entre concepts, objets et événements, et dans une certaine mesure pour anticiper sur la suite du texte. Le tout deux à trois fois plus rapidement qu'avec le langage oral. Ce processus rapide et en apparence continu peut se dérouler pendant des heures sans difficulté majeure. L'être humain doté de la maîtrise de l'écrit peut ainsi participer à l'actualité politique ou sportive, devenir expert dans toutes sortes de domaines de compétences, se plonger dans des enquêtes policières, des aventures romantiques ou des mondes futurs, le tout sans bouger de sa chaise. A moins qu'il ne décide de se lever pour se faire un café ou répondre au téléphone, ce qui ne cause en général qu'une perturbation mineure dans son activité de compréhension.

Inutile de préciser qu'aucun autre primate supérieur ne peut effectuer de tâche comparable. Les ordinateurs non plus. A l'heure où l'informatique pilote des chaînes de montage, rivalise avec les champions d'échecs ou modélise le climat de la planète, aucun logiciel n'est capable ne serait-ce que de s'approcher de la performance humaine pour ce qui concerne la compréhension du langage. Les programmes de traitement automatique du langage naturel (TALN), pour impressionnants qu'aient été leurs progrès en 20 ans, n'en sont encore qu'à traiter approximativement de petites unités très bien contrôlées, ou à simuler la compréhension à partir de toutes sortes d'artifices. La construction de représentations internes des situations à partir de textes, et les opérations qui en découlent (comme le raisonnement déductif) restent l'apanage des humains, du moins de ceux qui ont acquis les savoir-faire nécessaires à ces activités.

Là réside en effet le secret de la magie du langage et en particulier de l'écrit: un apprentissage long et parfois difficile, qui va de pair avec le développement général du système cognitif. Mais qu'est-ce donc que la maîtrise de l'écrit et comment l'acquiert-on? La psychologie contemporaine en donne essentiellement une définition à deux niveaux. Tout d'abord, pour comprendre le langage écrit, l'individu doit apprendre à en maîtriser les codes: c'est l'apprentissage de la lecture (Cf. Rieben & Perfetti, 1992). Cet apprentissage est absolument nécessaire, mais non suffisant pour maîtriser l'écrit. Comprendre un texte nécessite également des opérations de plus "haut niveau", telles que la production

d'inférences, qui parfois posent problème aux individus pourtant normo-lecteurs (Oakhill, 1994).

Jusqu'à présent, les psychologues (qui essaient de comprendre la maîtrise de l'écrit) et les éducateurs (qui essaient de l'enseigner) n'ont pas vraiment dépassé cette conception dichotomique de la maîtrise de l'écrit: savoir bien lire, et bien comprendre ce qu'on lit. Pourtant, des milliers d'élèves, au collège, au lycée, à l'université et dans les centres de formation professionnelle continuent d'échouer dans leurs activités, alors même qu'ils ont appris à lire et qu'ils sont dans une certaine mesure capables de comprendre. Pourquoi? L'hypothèse que je défends dans cet ouvrage est celle de l'existence d'un troisième niveau de compétence, qui permet la maîtrise fonctionnelle de l'écrit dans les situations finalisées. A ce niveau, la compétence du sujet se compose de connaissances procédurales et déclaratives des formes et des fonctions discursives globales (textes), et des règles qui gouvernent leur mise en oeuvre dans des situations spécifiques. Contrairement aux compétences plus élémentaires, cette maîtrise fonctionnelle n'est que très partiellement acquise au terme de la scolarité primaire. Or elle n'est pas directement enseignée par la suite, alors qu'elle est de plus en plus fortement sollicitée dans les situations d'apprentissage. C'est parce qu'ils n'ont pas une bonne maîtrise fonctionnelle de l'écrit que beaucoup d'élèves échouent dans les activités d'apprentissages à partir de documents complexes.

Je ne prétends pas que la maîtrise fonctionnelle de l'écrit constitue un processus inhérent au développement du système cognitif humain. Il faut plutôt l'envisager en termes d'apprentissage et acquisition de compétences. Toutefois, je prétends que qu'une théorie générale de la maîtrise du langage écrit se doit d'intégrer cette dimension fonctionnelle. En effet, dans les sociétés industrielles et post-industrielles ("sociétés de l'information"), l'écrit est devenu un mode de communication fondamental dans la vie professionnelle et sociale. Pour vivre une existence autonome, l'individu doit posséder une solide maîtrise des formes et des fonctions discursives complexes. On peut donc considérer la maîtrise fonctionnelle de l'écrit comme un aspect du développement cognitif fortement socialisé qui passe par des apprentissages explicites.

Ce thème de recherche présente également un intérêt pratique évident. Dans de nombreux pays, la densité et la complexité des systèmes d'information et de communication mettant en oeuvre le langage écrit augmentent bien plus vite que le niveau de compétence moyen de la population, même si ce dernier a lui même fortement progressé depuis un demi-siècle. Les services publics, les librairies, les parkings et les supermarchés sont désormais truffés de machines dont l'usage demande le traitement d'informations symboliques au moyen d'interfaces informatiques. Cette informatisation brutale et irréfléchie entraîne le glissement discret mais réel de toute une partie de la population vers une forme d'exclusion que l'on peut résumer par l'expression "d'illectronisme". L'institution scolaire, quant à elle, vit encore au rythme d'une mythologie informatique née dans les années 60 avec l'enseignement programmé (Skinner, 1964), selon laquelle l'ordinateur serait un outil pédagogique générateur de motivation, d'individualisation et donc d'amélioration des apprentissages. L'Ecole tend à intégrer de façon aveugle les nouvelles technologies, Internet en tête, sans questionner sérieusement la pertinence ou l'efficacité réelles de ces technologies. Pourtant, des outils d'informations puissants mais inadaptés aux besoins pédagogiques des élèves ne bénéficient réellement qu'aux meilleurs. Les élèves moins doués, en retard, ou tout simplement moins exposés à l'écrit dans leur environnement familial ne peuvent qu'échouer un peu plus dans ces écoles fortement mais bêtement "technologisées".

La mythologie de l'ordinateur neutre et bienveillant, qui motive intrinsèquement les élèves et garantit un développement harmonieux (et coopératif) de leur(s) intelligence(s) ne résistera pas longtemps aux études de terrain, si un jour l'institution scolaire entreprend d'en mener dans des conditions d'objectivité raisonnable. A l'approche du "techno-push" qui a prévalu jusqu'à présent, il faudra substituer une approche dans laquelle les choix informatiques seront effectués en fonction de modèles cognitifs et pédagogiques explicites, puis validés selon des méthodes empiriques rigoureuses. Il sera alors temps de se demander ce qu'il faut enseigner aux élèves pour mieux les préparer à cette société où le virtuel le dispute au réel et où tous les actes de la vie quotidienne se voient inexorablement contaminés par les systèmes d'informations.

Ce changement de paradigme supposera que soient identifiées les raisons qui font que l'on peut parfois "comprendre" sans être capable d'utiliser l'information. Il faudra également disposer d'un modèle général des processus cognitifs qui sous-tendent cette maîtrise fonctionnelle de l'écrit, et d'un modèle développemental de leur acquisition. A partir de ces modèles, on peut espérer mettre au point des méthodes pédagogiques efficaces et, finalement, être en mesure de considérer les apports possibles des nouvelles technologies à ces méthodes.

Les travaux que j'ai entrepris depuis près de 10 ans, d'abord au Learning Research and Development Center (Université de Pittsburgh), puis au Laboratoire Langage et Cognition (CNRS et Université de Poitiers), après un détour par l'INRIA (Grenoble), essaient de contribuer à ces objectifs. Après une rapide présentation de l'état actuel des théories de la compréhension (Chapitre 1), c'est à une visite guidée des principaux aspects de ma réflexion que j'aimerais convier le lecteur. Plus précisément, le Chapitre 2 sera consacré à un domaine de connaissances fondamental mais jusqu'à présent relativement négligé: celui des propriétés rhétoriques des textes et des discours. Le Chapitre 3 traitera de l'intégration d'informations à partir de sources multiples, ou synthèse documentaire. Il s'agit d'un type d'activité tout aussi fréquent que la compréhension de phrases ou de textes isolés qui ont jusqu'à présent retenu l'attention des chercheurs. Le Chapitre 4 concerne une autre activité courante mais mal connue: la recherche d'informations dans des documents complexes. Je m'efforce de montrer que ces deux types d'activités reposent sur des compétences spécifiques, caractérisées par l'intégration de paramètres sémantiques, épisodiques et pragmatiques des textes et des situations de lecture.

Les deux chapitres suivants traiteront de considérations plus spécifiquement liées aux nouvelles technologies de l'information et à leurs impacts sur le traitement cognitif de l'écrit au sens large. J'examinerai tour à tour les aspects cognitifs des hypertextes, et ceux de l'intégration multimédia. Finalement, je propose dans un chapitre conclusif quelques implications de ces travaux pour une redéfinition de la maîtrise de l'écrit, pour un usage raisonnable des technologies de l'information en éducation, ainsi que quelques perspectives nouvelles de recherche.