

**Psychologie cognitive et compréhension de texte :  
Une démarche théorique et expérimentale.**

Par

**Guy Denhière\*, Benoît Lemaire\*\*, Cédric Bellissens\*\*\* & Sandra Jhean-Larose\*\*\*\***

\* Guy Denhière, Laboratoire de Psychologie Cognitive, U.M.R. 6146, CNRS et Université de Aix-Marseille I, 3 place Victor Hugo, 13331 Marseille Cedex 3, France. [denhiere@up.univ-mrs.fr](mailto:denhiere@up.univ-mrs.fr)

\*\* Benoît Lemaire, Laboratoire des Sciences de l'Éducation, Université de Grenoble II, BP 47, 38040 Grenoble Cedex 9, France. [Benoit.lemaire@upmf-grenoble.fr](mailto:Benoit.lemaire@upmf-grenoble.fr)

\*\*\* Cédric Bellissens, Laboratoire Cognition et Activités Finalisées, Université de Paris VIII, 2 rue de la Liberté, 93526 Saint-Denis Cedex 02, France. [cbellissens@univ-paris8.fr](mailto:cbellissens@univ-paris8.fr)

\*\*\*\* Sandra Jhean-Larose, IUFM de Paris, Laboratoire Cognition et Activités Finalisées, Université de Paris VIII, 2 rue de la Liberté, 93526 Saint-Denis Cedex 02, France. [jhean@paris.iufm.fr](mailto:jhean@paris.iufm.fr)

In : S. Porhiel & D. Klingler (Eds) (2004). *L'unité texte* (pp. 74-95). Pleyben : Perspectives.

Envisagée du point de vue de la psychologie cognitive, l'étude expérimentale de la compréhension de textes peut faire l'objet d'approches différentes (voir Denhière, 1984 ; Denhière & Rossi, 1991) selon l'intérêt que les chercheurs accordent à des facteurs tels que la reconnaissance du langage (Gineste et Le Ny, 2002 ; Marouby-Terriou et Denhière, 2002), l'activation automatique du lexique mental (Kintsch, 1988 ; Mullet et Denhière, 1997), la production d'inférences (Kintsch, 1993 ; McKoon & Ratcliff, 1992), les activités cognitives de construction du sens (Le Ny, 1979 ; François & Denhière, 1997), l'établissement de la cohérence locale ou globale à partir d'indices de la structure syntaxique de surface (Costermans et Fayol, 1997), les limites de la mémoire de travail à court-terme (Ehrlich, 1994), les propriétés de la mémoire permanente (Blanc et Brouillet, 2003 ; Caillies, Denhière et Kintsch, 2002), les limites de la mémoire de travail à long-terme (Ericsson et Kintsch, 1995) en rapport avec les connaissances initiales du lecteur/auditeur (Bellissens & Denhière, 2004 a,b ; Kintsch, Patel & Ericsson, 1999 ; McNamara & Kintsch, 1996) et le type de structure textuelle (Caillies & Denhière, 2001 ; Caillies, Denhière et Jhean-Larose, 1999).

Ainsi, l'étude de la compréhension de textes implique une conception théorique homogène qui permette la prise en compte simultanée : (i) des caractéristiques textuelles, objet principal du colloque, (ii) des propriétés structurales et fonctionnelles de la mémoire humaine (O'Reilly & Rudy, 2000 ; Tiberghien, 1997), (iii) ainsi que des processus, non seulement intentionnels, mais également automatiques mis en œuvre lors de la lecture (Denhière & Baudet, 1992).

À propos des processus, il est important de rendre compte non seulement des processus délibérés et conscients qui contribuent à l'obtention des résultats observés mais également des processus automatiques, non contrôlés, et de leur dynamique temporelle dont sont dérivés les précédents. Ainsi, si l'on suppose les lexèmes et les formes verbales reconnus et identifiés, l'étude des activités de compréhension doit prendre en compte l'activation initiale des signifiés, mono- ou polysémiques, et du cours temporel de leur activation (Mullet et Denhière, 1997). De ce point de vue, la grande majorité des recherches expérimentales ont mis en évidence l'activation automatique initiale des *n* acceptions d'un terme ou d'une expression, imperméable au contexte pendant une durée brève de 100 à 150 msec., et suivie d'une sélection du sens pertinent par rapport au contexte, qu'il s'agisse d'homographes nominaux (voir Théroüanne et Denhière, 2002, 2004) ou d'expressions idiomatiques (Denhière & Verstiggel, 1997 ; Pariollaud, Denhière et Verstiggel, 2003). Autrement dit, le chaos caractérise les phases initiales du traitement du flux de l'information et l'ordre émerge progressivement en fonction des contraintes imposées par les informations déjà traitées, les connaissances langagières et les connaissances du domaine abordé par le texte : la signification est alors conçue comme une structure émergente des données traitées et non comme une structure élaborée en fonction de règles formelles pré-établies (Denhière & Tapiero, 1996 ; Kintsch, 1998a).

C'est un des mérites principaux du modèle de Construction-Intégration proposé par Kintsch (1988) que de permettre l'étude de la compréhension de textes en respectant la double contrainte posée par (i) les structures de connaissances et de croyances initiales des agents cognitifs et (ii) par les processus automatiques initiaux d'accès au lexique mental et de traitement des différentes formes d'ambiguïté, syntaxique ou sémantique. Ainsi, d'une part, Till, Mross & Kintsch (1988) ont pu montrer que des inférences pertinentes comme non pertinentes étaient initialement activées et que, d'autre part, des ajouts d'information au texte traité généralement considérés comme une production d'inférences n'était en réalité que des

informations récupérées en mémoire à long-terme (Kintsch, 1993 ; Aguilar & Denhière, à paraître).

Les considérations qui précèdent commandent l'organisation du chapitre ci-dessous. Nous commencerons par rappeler les traits essentiels du modèle de Construction-Intégration de Kintsch (1988), après quoi nous présenterons brièvement le système de mémoire permanente couplé au modèle de compréhension : l'Analyse Sémantique Latente (Landauer et Dumais, 1997 ; Landauer, 2002), pour terminer par l'adaptation que nous avons faite de ces deux modèles (Denhière et Lemaire, 2004) et un exemple d'application (Lemaire et Denhière, 2004).

## **1. Le modèle de Construction-Intégration**

Kintsch postule qu'un modèle cognitivement plausible de la compréhension de textes se doit d'être guidé par les connaissances du lecteur. Les connaissances générales sur les mots, la syntaxe, les relations spatiales, le monde, etc. contraignent et rendent possible la construction des différents niveaux de représentation du discours : structure de surface, micro- et macrostructure sémantiques, modèle de situation (voir van Dijk & Kintsch, 1983). Dans la phase de construction de la signification, un système de production contenant des règles assez puissantes et souples pour permettre au candidat-mot correct d'appartenir à l'ensemble élaboré, alors même que d'autres non pertinents ou inappropriés sont produits. La sortie de ce système, partiellement incohérente et contradictoire, se présente sous la forme d'un réseau associatif, qui est transformé en une base de texte cohérente par des procédures de relaxation connexionniste (Rumelhart & McClelland, 1987). Un processus d'intégration est ensuite mis en oeuvre pour renforcer les éléments contextuellement appropriés et inhiber les éléments non reliés ou inappropriés.

### **1.1. La représentation des connaissances en mémoire**

Le processus de construction de la représentation du discours met en oeuvre les connaissances. Les connaissances sont représentées par un réseau associatif dont les noeuds sont des concepts ou des propositions qui sont formellement traités de la même façon. Les noeuds dans ce réseau sont interconnectés et la valeur de la force de connexion peut être positive, nulle, ou négative. La signification des concepts est construite à partir de leur position dans le réseau : ce sont les associés immédiats et les voisins sémantiques d'un noeud qui constituent le noyau de sa signification. La signification complète correspondant à un noeud ne peut être obtenue qu'en explorant les relations de ce noeud avec tous les autres noeuds du réseau. La signification doit être créée. Le premier niveau de la signification correspond à toutes les propositions dans le réseau qui sont directement reliées à un noeud. C'est seulement ensuite que toutes les propositions directement reliées aux propositions du premier niveau pourront être ajoutées pour former un second niveau, et ainsi de suite, jusqu'à ce que la totalité du réseau de connaissances soit impliquée. En d'autres termes, la signification d'un concept est toujours dépendante d'une situation spécifique et du contexte. Elle est nécessairement incomplète et instable : des noeuds supplémentaires peuvent toujours être ajoutés au sous-réseau activé qui constitue la signification temporaire d'un concept, et cela au prix de la perte de certains noeuds déjà activés.

### **1.2. Le processus de construction**

Les étapes dans la construction de la base de texte selon le modèle de construction-intégration implique : (a) la formation des concepts et des propositions correspondant directement à l'entrée linguistique ; (b) l'élaboration de chacun de ces éléments par la sélection

d'un petit nombre de voisins associés du réseau de connaissances ; (c) l'inférence de certaines propositions supplémentaires ; et (d) l'assignation des forces de connexion à toutes les paires d'éléments qui ont été créées. Le résultat est une base de texte initiale riche, incohérente et contradictoire, qui est alors soumise à un processus d'intégration permettant la formation d'une structure cohérente.

**Lors de la formation des concepts et des propositions**, une représentation propositionnelle du texte est construite à partir de l'entrée linguistique syntaxiquement analysée et du réseau de connaissances. Les règles d'élaboration permettent la formation en ligne de propositions incomplètes ou fausses car l'information pertinente pour l'élaboration des propositions n'est pas toujours disponible.

**L'étape suivante consiste dans la sélection de voisins associés à la proposition construite.** Dans cette étape, chaque concept ou proposition élaboré précédemment sert d'indice pour la récupération de noeuds associés dans le réseau de connaissances. Ce processus de récupération est lui-même modélisé d'après des théories relatives à la mémoire (voir Raaijmakers & Shiffrin, 1981). Chaque concept ou proposition de la base de texte sert d'indice de récupération et active les propositions qui lui sont étroitement associées dans le réseau général de connaissances, et ceci indépendamment du contexte. Ces propositions de connaissances fonctionnent comme des inférences potentielles.

**La production d'inférences spécifiques et contrôlées.** Au cours de cette étape, la production d'inférences supplémentaires est nécessaire car toutes les inférences requises pour la compréhension ne sont pas obtenues par le mécanisme aléatoire d'élaboration envisagé précédemment. Deux types d'inférences sont d'une importance particulière dans la compréhension : les inférences de liaison nécessaires quand la base de texte construite est incohérente et les macropropositions élaborées associativement comme les micropropositions (voir Kintsch, 2004).

**L'assignation des forces de connexion.** Les étapes décrites jusque là débouchent sur la construction d'un ensemble de propositions contenant les micropropositions directement dérivées du texte, un ensemble aléatoirement sélectionné d'associés pour chacune d'entre elles et des macropropositions ainsi que leurs associés. L'ultime étape de ce processus de construction implique la spécification des interconnexions entre tous les éléments. Les propositions du texte sont positivement interconnectées avec des pondérations proportionnelles à leur proximité dans la base de texte. Les propositions de la base de texte héritent de leurs interconnexions dans le réseau général de connaissances. Les pondérations sont additives jusqu'à un maximum de 1, dans les cas où une valeur de force héritée se combine avec une connexion déterminée par la base de texte. Quant aux inférences, elles héritent d'interconnexions positives et négatives du réseau général de connaissances. Le résultat du processus de construction est un réseau exprimable sous la forme d'une matrice de connexité, constituée par tous les noeuds lexicaux, par toutes les propositions ainsi que par toutes les inférences et les élaborations qui ont été produites au niveau local et global.

### **1.3. Le processus d'intégration**

Le réseau construit jusqu'alors est incohérent. A tous les niveaux de la représentation, les composants associés aux éléments du texte sont inclus indépendamment du contexte, et nombre d'entre eux sont inappropriés. Un processus d'intégration de type connexionniste est alors utilisé pour exclure ces éléments indésirables de la représentation du texte (Rumelhart & McClelland, 1987).

La compréhension d'un texte est supposée s'effectuer par cycles de traitement correspondant grossièrement à de courtes phrases ou à des syntagmes nominaux. A chaque cycle de traitement, un nouveau réseau incluant les propositions ou concepts fortement activés du cycle précédent est construit. Une fois le réseau correspondant à ce cycle construit, le processus d'intégration s'effectue : l'activation se diffuse dans le réseau jusqu'à sa stabilisation. Plus précisément, un vecteur d'activation représentant les valeurs d'activation initiale de tous les noeuds du réseau est multiplié à maintes reprises avec la matrice de connexité. Après chaque multiplication, les valeurs négatives sont remplacées par zéro et chacune des valeurs d'activation positives est divisée par la somme de toutes les valeurs d'activation, ceci afin que l'activation totale de chaque cycle demeure égale à une valeur de 1. Généralement, le système se stabilise assez rapidement.

Le résultat du processus d'intégration est un nouveau vecteur d'activation, le vecteur final d'activation. Ce dernier indique les valeurs d'activation finale obtenues pour chacun des noeuds dans le réseau, les noeuds les plus fortement activés constituant la représentation du discours formée à chaque cycle de traitement<sup>1</sup>.

## **2. Représentation de l'information en mémoire : l'Analyse Sémantique Latente (Landauer & Dumais, 1997)**

Le modèle de mémoire associative auquel se réfère Kintsch pour coupler son modèle d'architecture cognitive à la représentation de l'information n'est pas un modèle classique en réseau avec des relations hiérarchiques étiquetées du style Collins & Quillian (1968) mais la représentation vectorielle développée par Landauer et Dumais (1997) (voir Bellissens, Thérouanne et Denhière, à paraître).

### **2.1. La constitution d'un espace sémantique**

L'Analyse Sémantique Latente est une technique statistique automatique proposée qui permet de déterminer, à partir de la variabilité contextuelle des passages du discours la connaissance de la signification des mots (voir Manning et Schütze, 1999). Cette technique n'utilise pas de dictionnaires construits par l'homme, de bases de connaissances, de réseaux sémantiques, de grammaires, d'analyseurs syntaxiques, de données morpho-syntaxiques, ou tout autre chose d'approchant<sup>2</sup>. L'Analyse Sémantique Latente prend en entrée uniquement un très vaste corpus de textes bruts supposés représentatifs des textes auxquels a été exposée une population donnée<sup>3</sup> (Landauer, 1998).

Un corpus est représenté sous la forme d'un tableau avec en colonne les paragraphes et en ligne les mots ; les cellules comptabilisant le nombre de cooccurrences mot - paragraphe. Les cooccurrences sont transformées et soumises à une analyse en composantes principales appelée « décomposition en valeurs singulières ». Cette analyse factorielle généralisée permet

---

<sup>1</sup> Le modèle de Construction-Intégration est un modèle d'architecture cognitive suffisamment général pour pouvoir considérer des modèles postérieurs tels que le «*modèle paysager*» de van den Broek (1996) et le «*modèle de résonance*» de Myers et O'Brien (1998) comme des variantes à portée plus limitée.

<sup>2</sup> Voir Ploux et Victorri (1998), Gaume (2003) pour une démarche différente consistant à utiliser des dictionnaires informatisés de synonymes.

<sup>3</sup> L'utilisation de vastes corpus pour représenter la signification des mots n'est pas propre à LSA. Une autre technique, essentiellement basée sur l'analyse des co-occurrences, a été proposée par Burgess (voir Burgess et Lund, 1997 ; Burgess Livesay et Lund, 1998).

de déterminer des facteurs indépendants représentés sous la forme de dimensions orthogonales, lesquelles définissent un espace sémantique. Tous les facteurs dégagés par la décomposition en valeurs singulières ne sont pas également importants, ils ne rendent pas compte des mêmes pourcentages de variance. Aussi, les facteurs correspondant aux valeurs singulières les plus faibles sont assimilés à du bruit inhérent au type de données analysées et sont éliminés. Dans la plupart des applications, le nombre de facteurs pertinents est réduit à 300 ou 400. Ceci signifie que les vecteurs signifiants passent de 30000 coordonnées dans un espace à 30000 dimensions à 300 coordonnées dans un espace à 300 dimensions. A la suite de cette réduction, chaque mot est représenté dans cet espace sous la forme d'un vecteur à  $n$  dimensions. La distance sémantique entre deux mots et/ou expressions<sup>4</sup> peut alors être calculée. Cette distance sémantique, correspondant à la distance entre deux vecteurs dans un espace à  $n$  dimensions, est le plus souvent estimée par le cosinus de l'angle entre les vecteurs correspondants<sup>5</sup>. Plus le cosinus de deux vecteurs est proche de 1, plus la distance sémantique entre ces deux vecteurs est faible, et plus les sens considérés sont semblables.

## 2.2. Évaluation du modèle : Le test du meilleur synonyme

LSA a été évaluée dans de nombreuses expériences et de façons très différentes. L'une d'entre elles consistait à comparer les résultats du modèle LSA à ceux d'individus au test du TOEFL (Test of English as a Foreign Language). Les auteurs ont utilisé un corpus de 4,5 millions de mots, constitué de 30.473 contextes comportant chacun environ 500 caractères, le nombre de dimensions étant réduit à 300. Le test de vocabulaire du TOEFL comportait 80 items dont la partie question était constituée d'un seul mot alors que quatre réponses étaient proposées, la tâche du candidat consistant à choisir le meilleur synonyme.

Pour simuler la performance humaine à partir de LSA, la distance sémantique entre le mot inducteur et chaque mot réponse est calculée. Le meilleur synonyme est le mot dont le vecteur correspondant est le plus proche du vecteur correspondant au mot inducteur : le cosinus de l'angle entre les deux vecteurs est supérieur aux autres. Le pourcentage de réponses correctes de LSA est de 65 %, score qui correspond à la performance moyenne des étudiants passant ce test. Par ailleurs, LSA fait également des erreurs lorsque des étudiants n'en font pas. Ces erreurs montrent que LSA est sensible aux relations paradigmatiques entre concepts plutôt qu'aux relations de stricte synonymie.

LSA possède donc une connaissance sur la signification des mots ou expressions d'un corpus. Cette connaissance induite de la variabilité contextuelle du corpus d'entraînement semble être similaire à celle d'être humains puisque LSA obtient les mêmes résultats. D'autres mises à l'épreuve de LSA ont donné des résultats généralement considérés comme satisfaisants (voir Landauer, Foltz et Laham, 1998).

---

<sup>4</sup> Le vecteur d'une expression est la somme des vecteurs des mots qui la composent. En conséquence, limite souvent énoncée de ce type de représentation, le vecteur représentant "le chat mange la souris" est égal à celui représentant "la souris mange le chat". Nous y reviendrons plus loin.

<sup>5</sup> D'autres mesures que le cosinus sont possibles : la distance euclidienne, la distance de Hellinger, la Distance "city block" et la divergence Kullback-Leibler (voir Patel, Bullinaria et Levy, 1997).

Concrètement, divers types de corpora de langue anglaise et de langue française sont accessibles sur le site : <http://lsa.colorado.edu>. Nous y avons également installé des corpora de langue française<sup>6</sup> : Français-Monde correspond à 6 mois de l'année 1993 du journal Le Monde, Français-Littérature à la somme de tous les ouvrages littéraires antérieurs à 1920 que nous avons pu collecter sous forme électronique (près de 400 ouvrages). "Français-Contes" est formé de tous les contes, fables, histoires et textes pour enfants que nous avons pu collecter sur les sites internet, ainsi que des textes scolaires de CP inclus dans le corpus "Novlex" développé par Chesnet et Lambert (2002).

## **2. Le couplage d'un modèle de compréhension : Construction-Intégration, et d'un modèle de Représentation de l'information en mémoire : l'Analyse Sémantique Latente.**

### **2.1. L'espace sémantique « TextEnfant »**

Nous avons décidé de connecter le modèle de construction-intégration à une approximation réaliste d'une mémoire humaine, à savoir la mémoire des enfants de fin de primaire. Pour cela nous avons constitué le corpus « TextEnfant » qui comporte 3,2 millions de mots et 57.878 paragraphes. « TextEnfant » est composé de contes et de récits familiaux (52%), de productions verbales d'enfants de cours préparatoire et de cours élémentaire<sup>7</sup> (23%), d'extraits de manuels de lecture de cours préparatoire<sup>8</sup> (13%) et d'une encyclopédie électronique pour enfants "Je sais tout" (12%). Cette composition traduit la volonté de tendre vers la constitution d'un espace sémantique représentatif de l'ensemble des textes auxquels ont été confrontés les enfants de primaire. Cependant, en l'absence de travaux scientifiques fiables sur l'étendue et le contenu de la mémoire sémantique enfantine, il convient de s'assurer expérimentalement de la représentativité de l'espace ainsi construit (Denhière et Lemaire, 2004a). Aussi, nous avons testé l'espace « TextEnfant » en déterminant dans quelle mesure il rendait compte des normes associatives enfantines (de La Haye, 2003), des définitions lexicales et des rappels et résumés de récits (Denhière et Lemaire, 2004b).

de La Haye (2003) a utilisé 200 mots inducteurs (144 noms, 28 verbes et 28 adjectifs) et elle a demandé à des participants enfants (9, 10 et 11 ans) et adultes de produire les mots qui leur venaient immédiatement à l'esprit à la lecture du terme inducteur. Pour chaque terme inducteur, elle a recensé les termes évoqués ainsi que la fréquence avec laquelle ils l'avaient été dans chaque groupe. Pour tester l'espace « TextEnfant », nous avons d'abord sélectionné les trois termes associés les plus fréquemment produits à chacun des 200 mots inducteurs ainsi que les trois derniers puis nous avons ensuite calculé la distance sémantique entre le mot inducteur et chacun des six associés sélectionnés en calculant le cosinus entre les vecteurs représentant les éléments de chaque couple mot inducteur-mot associé. Si « TextEnfant » s'approche de la mémoire sémantique enfantine, la hiérarchie des valeurs des cosinus obtenus doit tendre à reproduire la hiérarchie des fréquences d'évocation des associés d'un terme

---

<sup>6</sup> « Français-Livres » se compose de 14.622 paragraphes, de 111.094 mots différents et de 5.748.581 occurrences. « Français-Monde » comporte 20.208 paragraphes, 150.756 mots différents et 8.675.391 occurrences. « Français-Total » correspond à la concaténation des deux corpora précédents.

<sup>7</sup> Ce travail de collecte de textes libres d'enfants des écoles Freinet de Louvain La Neuve a été réalisé avec Marie-Anne Schelstraete, que nous remercions ici.

<sup>8</sup> Nous remercions nos collègues Chesnet et Lambert de l'Université de Poitiers d'avoir mis à notre disposition leur corpus "Novlex" de textes de lecture du CP (voir Chesnet et Lambert, 2002).

donné. Globalement, les deux hiérarchies sont significativement corrélées ( $r = .39$ ,  $p < .01$  pour les 200 mots inducteurs,  $r = .57$  pour les mots inducteurs les plus fréquents), la distribution des cosinus moyens entre le mot inducteur et les mots associés, présentée sur le tableau 1, indique un gradient net, toutes les comparaisons 2 à 2 entre les colonnes adjacentes du tableau étant significative ( $p < .03$ ).

Tableau 1 : Distribution des valeurs moyennes des cosinus entre le mot inducteur et les mots associés.

	1 <sup>er</sup> associé	2 <sup>ème</sup> associé	3 <sup>ième</sup> associé	3 derniers associés
Cosinus	.46	.33	.19	.11

En conclusion, l'espace sémantique « TextEnfant » subit ce premier test d'association sémantique en mémoire avec succès : il reproduit de manière significativement acceptable la hiérarchie des termes associés aux 200 mots inducteurs utilisés par de La Haye (2003).

Un second test, plus contraignant que le précédent, consiste à déterminer dans quelle mesure « TextEnfant » est capable de reproduire les performances enfantines de définition de termes alors qu'aucun dictionnaire n'est entré dans le corpus et que les réponses proposées peuvent être des mots ou des phrases. On a donc repris les résultats d'une épreuve de définition de 120 mots (45 Noms, 55 Verbes, 9 Adjectifs, 6 Adverbes) développée par Denhière, Bourguet et Thomas (2000, à paraître) administrée à quatre groupes d'enfants de 8 à 11 ans. Cette épreuve consistait à présenter un mot à définir et à proposer quatre définitions : correcte, proche, éloignée et sans rapport. Ainsi, par exemple, les réponses proposées pour « cigale » étaient respectivement : (1) insecte à quatre ailes qui vit dans les régions méditerranéennes (réponse correcte) ; (2) insecte proche du grillon (réponse proche) ; (3) petit animal de la même famille que la chauve-souris (réponse éloignée) et (4) chenille noire (réponse sans rapport).

Divers indicateurs de la qualité des performances de « TextEnfant » peuvent être utilisés. Une première mesure consiste à calculer le cosinus entre les vecteurs du mot à définir et les quatre mots réponse proposés. Ainsi, dans l'exemple précédent les valeurs de cosinus sont respectivement égales à : .38 (correct), .24 (proche), .16 (éloigné) et .04 (sans relation). A partir de ces résultats, on peut calculer le pourcentage de réponses correctes de « TextEnfant », c'est-à-dire le nombre de fois où la valeur de cosinus pour la réponse correcte est plus élevée que pour les autres réponses. Ce pourcentage est égal à 53 %, ce qui situe les performances de « TextEnfant » entre celles des enfants de 9 et 10 ans. De plus, le pourcentage de réussite en fonction de la distance des réponses à la réponse correcte diminue régulièrement (voir Denhière et Lemaire, 2004a,b pour plus de détails). Une mesure plus sophistiquée mais plus satisfaisante consiste à calculer une différence normalisée  $d$  entre réponses correcte et proche versus réponses éloignée et sans rapport. Plus la valeur de  $d$  est importante, et plus « TextEnfant » discrimine correctement les définitions proposées. Enfin, de manière à déterminer dans quelle mesure l'espace sémantique « TextEnfant » reproduit l'organisation de la mémoire enfantine, on a comparé les performances de « TextEnfant » à celles de cinq corpus formés de textes pour adultes : quatre années du journal Le Monde (1993, 1995, 1997 et 1999) et le corpus « Littérature ». Les résultats sont présentés dans le tableau 2.

Tableau 2 : Comparaison entre les performances de l'espace sémantique « TextEnfant » et les espaces sémantiques pour adultes.

Espaces sémantiques	Millions de mots	% réponses correctes	$d$	Corrélation

TextEnfant	<b>3,2</b>	<b>53</b>	<b>69</b>	<b>.39</b>
Littérature	14,1	38	52	.34
Le Monde 1993	19,3	44	23	.31
Le Monde 1995	20,6	37	21	.26
Le Monde 1997	24,7	40	28	.26
Le Monde 1999	24,2	34	25	.24

En dépit de la différence très significative de taille des corpora adultes comparés au corpus enfant (de 14 à 20 millions de mots versus 3 millions), les performances de l'espace sémantique « TextEnfant » sont significativement supérieures à celles des espaces adultes quel que soit l'indicateur considéré, la corrélation entre les résultats des enfants de 8 à 11 ans et les performances de « TextEnfant » étant significativement supérieures à toutes les corrélations avec les corpus adultes ( $p < .03$ ).

**Troisième test de l'espace sémantique** « TextEnfant », sa capacité à être utilisé pour fournir une évaluation automatique des rappels et des résumés de textes, de récits en particulier, produits par des enfants de 7 à 11 ans. Classiquement, le texte à rappeler ou à résumer est analysé en propositions de type [prédicat(argument(s))], les rappels ou les résumés produits sont analysés de la même manière de manière à déterminer quels sont les concepts et propositions les plus fréquemment présents dans les restitutions obtenues. Dans le cadre de l'analyse sémantique latente, la mesure de la quantité d'information redonnée est estimée par la valeur du cosinus entre le vecteur représentant le texte source (à rappeler ou à résumer) et le rappel ou résumé produit. La démarche suivie ici a consisté à reprendre les résultats obtenus dans six épreuves de rappel (immédiat et différé) et trois épreuves de résumé et à calculer, pour chaque texte, la corrélation entre le nombre de propositions restituées et la valeur de cosinus obtenu<sup>9</sup>.

Si notre espace sémantique permet de mesurer les performances de rappel et de résumé, on obtiendra une corrélation positive et significative entre le nombre de propositions restituées et la valeur du cosinus entre le texte-source et les restitutions. Le tableau 3 présente les corrélations obtenues entre les valeurs de cosinus et le nombre de propositions redonnées pour chacune des 9 expériences réexaminées.

Tableau 3 : Corrélations entre les performances de rappel et de résumé estimées par le nombre de propositions restituées et les valeurs de cosinus entre les textes-sources et les productions de rappel et de résumé.

Récit	Tâche	Nombre de participants	Corrélation
Dragon	Résumé	56	.71
Clown	Résumé	24	.92
Géant	Résumé	105	.58
Poule	Rappel immédiat	52	.45
Araignée	Rappel immédiat	41	.65
Clown	Rappel immédiat	56	.67

<sup>9</sup> Nous remercions Marianne Bourguet, Cécile Hupet et Hervé Thomas, de nous avoir donné accès aux protocoles individuels des expérimentations qu'ils avaient réalisées dans le cadre de travaux de thèse ou de mémoire dirigés par le premier auteur.

Ourson	Rappel immédiat	44	.62
Dragon	Rappel différé	44	.55
Taureau	Rappel différé	23	.69

Comme on le voit sur le tableau 3, les corrélations vont de .45 à .92 et sont toutes significativement différentes de zéro. Ceci signifie que l'utilisation de notre espace sémantique « TextEnfant » pour calculer le cosinus entre le texte-source à rappeler ou résumer et les productions enfantines, rappel ou résumé, fournit une bonne estimation de la quantité et de la nature de l'information restituée. Ce qu'avaient montré Foltz (1996) et Lemaire et Dessus (2001), Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch & Landauer (1998) avec des adultes est ici étendu à des enfants : un espace sémantique approprié peut être utilisé pour analyser et évaluer la compréhension et la mémorisation de texte d'une manière automatique (et donc rapide) et fortement corrélée aux résultats que donne une analyse prédictive longue et fastidieuse.

## 2.2. Un modèle de compréhension de texte utilisant en temps réel un espace sémantique « TextEnfant »

Le modèle de compréhension que nous proposons s'inspire des modèles de construction-intégration de Kintsch (1988) et de représentation en mémoire de l'Analyse sémantique Latente de Landauer et Dumais (1997), il est opérationnalisé sous la forme d'un programme informatique entièrement automatique (Lemaire et Denhière, à paraître) dont nous allons essayer d'illustrer le fonctionnement par un exemple. Soit le texte formé des deux phrases suivantes et une analyse prédictive sommaire, actuellement effectuée manuellement :

*L'abeille aspire le nectar de la fleur avec sa trompe. Elle apporte ensuite le liquide sucré à la ruche pour le transformer en miel.*

*P1. ASPIRER (abeille, nectar, fleur)*

*P2. TRANSPORTER (abeille, nectar, ruche)*

*P3. TRANSFORMER (abeille, nectar, miel)*

*P4. POUR (P2, P3)*

### La mémoire de travail et construction et intégration en ligne

En plus de la mémoire permanente (la mémoire à long-terme), le processus de compréhension met en jeu une structure dénommée mémoire de travail (Ehrlich, 1994) qui comprend les éléments essentiels de la phrase précédemment traitée et les éléments de la phrase en cours de traitement. Comme indiqué plus haut, l'agent cognitif active automatiquement des éléments de sa mémoire permanente ; en conséquence la mémoire de travail comporte trois types d'éléments : des éléments précédemment traités, des éléments en cours de traitement et des concepts automatiquement activés. La nouveauté importante qu'apporte le couplage d'une structure de mémoire comme l'espace sémantique « TextEnfant » à un modèle de compréhension, c'est que l'activation des éléments de la mémoire permanente par les éléments du texte lu ou entendu est entièrement automatique et paramétrable<sup>10</sup>. Le traitement en mémoire de travail porte donc à la fois sur les éléments du

<sup>10</sup> En conséquence, les éléments activés varieront automatiquement selon l'espace sémantique associé et pourront reproduire le processus de compréhension d'experts ou de novices (Caillies et Denhière, 2001 et Caillies, Denhière et Jhean-Larose, 1999).

texte et les éléments activés en mémoire, ce qui est parfois qualifié de manière impropre “d’inférences” alors qu’il s’agit de connaissances automatiquement récupérées en mémoire à long-terme (voir Kintsch, 1993).

La première proposition active automatiquement les éléments suivants :

*Insecte, larve, vole, ruche, miel, guêpe, bourdonnement, bouquet, violet, pétale, cueillir, bourgeon.*

Dans cette phase de construction, les similitudes sémantiques entre toutes les paires d’éléments sont calculées et un réseau sémantique comportant les propositions et concepts du texte ainsi que les éléments activés qu’ils soient pertinents ou pas. Ensuite, un mécanisme de diffusion de l’activation parcourt le réseau sémantique ainsi formé jusqu’à stabilisation des valeurs d’activation (comprises entre 0 et 1.000). Cette phase d’intégration a pour effet de sélectionner les éléments les plus pertinents, c’est-à-dire les éléments les plus reliés entre eux. Ainsi, après traitement de la première proposition, la mémoire de travail contient les éléments suivants :

*ASPIRER (abeille, nectar, fleur) : 1.000*

*Abeille : .903 ; fleur : .852 ; ruche : .778 ; bouquet : .678 ; bourdonnement : .634 ; miel : .615 ; pétale : .607 ; cueillir : .606 ; violet : .605.*

La première proposition de la phrase 2 active :

*travailleur, ruche, miel, bourdonnement, voler*

alors que la seconde proposition active :

*prendre, minéral, signification, sucre, vivant, ruche, abeille, ours, faisan.*

Comme précédemment, un réseau est formé avec les propositions et les éléments ajoutés et le mécanisme de diffusion de l’activation est lancé jusqu’à **stabilisation**. Au terme du traitement de ces deux phrases, les éléments les plus activés de la mémoire de travail sont les suivants :

*ASPIRER (abeille, nectar, fleur) : 1.000*

*TRANSPORTER (abeille, nectar, ruche) : .997*

*Abeille : .949*

*Ruche : .868*

*TRANSFORMER (abeille, nectar, miel) : .813*

*Miel : .805*

*Bourdonnement : .612*

Comme la mémoire de travail possède une “capacité limitée”, seuls les éléments les plus activés y demeurent, les autres étant “transférés” dans une mémoire épisodique.

### **La mémoire épisodique**

Les éléments “chassés” de la mémoire de travail et qui ne sont pas nécessaires aux traitements en ligne successifs sont transférés dans une mémoire dite “épisodique” qui garde la trace de tous les éléments traités en mémoire de travail. Une fois placés dans cette mémoire épisodique, certains éléments peuvent être replacés en mémoire de travail si nécessaire. Les éléments en mémoire épisodique sont stockés avec une valeur d’activation donnée qui diminue avec le temps à moins qu’elles ne soient réactivées par une réinsertion en mémoire de travail.

Il est à noter que, durant la phase de construction, la mémoire épisodique peut également fournir automatiquement des éléments à la mémoire de travail si ces éléments sont suffisamment proches sémantiquement des éléments (concepts ou propositions) du texte en

cours de traitement (Montoya, Denhière et Baccino, 2002). Ce mécanisme est similaire au processus d'inférence décrit à propos de la mémoire à long-terme. La figure 1 résume la manière dont l'information est susceptible de transiter à travers les divers registres de mémoire.

### INSERER ICI LA FIGURE 1

A la fin du traitement du texte, la mémoire épisodique contient tous les concepts et propositions du texte ainsi que leur degré d'activation. Les valeurs d'activation ainsi obtenues sont comparées à la distribution des fréquences avec laquelle les concepts et propositions sont rappelés par des groupes de participants humains pour déterminer le degré d'adéquation de la simulation par rapport aux données obtenues (Tapiero et Denhière, 1996). Il faut également noter que, chaque état de la mémoire épisodique étant stocké par le programme, il est possible de tracer le cours temporel du degré d'activation des concepts ou propositions donnés et de le comparer à des données comportementales (Jovet, Jhean-Larose, Bellissens et Denhière, 2004).

### Conclusion

Dans ce chapitre, nous avons présenté et illustré une démarche cognitive d'étude expérimentale de la compréhension de textes dont les caractéristiques principales consistent à :

- prendre en entrée du système les produits des phases initiales de traitement du langage, notamment des activités automatiques de traitement lexical impliquant à la fois le lexique mental et les connaissances stockées en mémoire permanente ;
- intégrer intimement traitement du langage et structures de mémoire : mémoire de travail à court-terme, mémoire épisodique et mémoire à long-terme ;
- modéliser la structure de la mémoire permanente en construisant des espaces sémantiques issus de vastes corpora supposés représentatifs des textes auxquels ont été exposés des individus d'un groupe donné ;
- expliciter les processus hypothétiques de traitement de façon suffisamment détaillée et explicite pour les implanter dans un logiciel permettant de paramétrer les composantes essentielles du modèle de compréhension ;
- connecter automatiquement processus de compréhension et structures mnésiques de manière à simuler non seulement le produit terminal de la compréhension mais également le cours temporel des activités élémentaires impliquées dans les processus de traitement et de mise en mémoire des produits (transitoires) de ces traitements.

Comme nous l'avons dit plus haut, ce travail s'inspire du modèle de construction-intégration proposé par Walter Kintsch (1988) et du modèle de mémoire : l'Analyse Sémantique Latente, formulée par Landauer et Dumais (1997). Nous remercions ces collègues d'avoir accepté de collaborer avec nous depuis 1989, date du premier contrat de collaboration CNRS-NSF entre Guy Denhière et Walter Kintsch. Ce travail est également le fruit d'une collaboration étroite entre psychologie cognitive et informatique<sup>11</sup>. Naturellement, ce travail nécessite d'être poursuivi et développé.

---

<sup>11</sup> Pour une large part, le travail décrit ici a été initié au Laboratoire de Psychologie Cognitive de l'Université d'Aix-Marseille 1 où Benoît Lemaire a passé un semestre dit

Les corpora peuvent et doivent être améliorés et rendus plus discriminants pour rendre compte des niveaux de performance à l'intérieur de la classe d'âge 8-12 ans. Nous testons actuellement l'hypothèse selon laquelle l'introduction de dictionnaires devrait améliorer les performances de "TextEnfant". Nous fondions beaucoup d'espairs sur la lemmatisation du corpus, espoirs vains ! Contre toute attente, ramener verbes, noms et adjectifs à une forme canonique n'améliore pas les performances de "TextEnfant" alors que la lemmatisation des seuls verbes s'accompagne d'une amélioration ! Empiriquement, on constate que les contextes dans lesquels sont utilisés un grand nombre de noms sont significativement différents selon leur forme, singulier ou pluriel . Ainsi, à titre d'exemple, alors que le cosinus entre «rose " et "fleur" est égal à .51, le cosinus entre "rose" et "fleurs" est de .14 ; peigne-cheveux : .56 et peigne-cheveu : .02 !

Le modèle de représentation de l'information en mémoire peut être modifié pour s'approcher des modèles en réseau classiques et rendre compte de l'apprentissage (Baudet, Jhean-Larose et Legros, 1994). Les premières tentatives faites dans ce sens sont encourageantes puisque le modèle en réseau ICAN développé par Lemaire et Denhière (2004) montre des performances au moins égales à celles de LSA.

Le modèle de compréhension peut être également amélioré, notamment par l'implémentation du concept de mémoire de travail à long-terme qui s'est révélé nécessaire pour rendre compte des effets d'expertise et d'interruption de la lecture notamment (voir Bellissens et Denhière, 2004a).

Last but not least, évidemment, reste à prendre automatiquement en compte la syntaxe, ce que nous tentons actuellement de faire avec les collègues du Laboratoire d'Informatique Fondamentale de Lumigny : Robert Paséro, Monique Rolbert et Jean Royauté. C'est au prix d'une collaboration réfléchie et équilibrée entre psychologues cognitivistes, linguistes et informaticiens que la compréhension de textes peut faire l'objets de progrès notables, les psychologues devant adapter leurs plans expérimentaux aux exigences de leurs collègues des autres disciplines tandis que linguistes et informaticiens doivent accepter de prendre en compte le produit final des activités des agents cognitifs humains comme les processus qui ont conduit aux résultats observables.

---

« sabbatique » de recherche. Que notre collègue Benoît Lemaire, informaticien dans un département de sciences de l'éducation, soit ici remercié pour nous avoir fait bénéficier, au-delà de ses compétences professionnelles et de ses qualités humaines, de sa grande patience, de la qualité de son écoute et de son souci d'opérationnaliser des hypothèses cognitives sous forme falsifiable.

## Bibliographie

- Aguilar, N., & Denhière G. (à paraître). La production des inférences sur la conséquence probable des événements et des actions. *L'Année Psychologique*.
- Baudet, S., Jhean-Larose, S., & Legros, D. (1994). Coherence and truth : A cognitive model of propositional truth attribution. *International Journal of Psychology*, 29, 319-350.
- Bellissens, C., & Denhière, G. (2002). Word order or environnement sharing : A comparison of two semantic models. *Current Psychology Letters*, 9, 43-60.
- Bellissens, C., & Denhière, G. (2004a). Retrieval from Long-term Working Memory : A Skilled Use of Semantic Memory. *Issues in Psycholinguistics*, 185-210.
- Bellissens, C., & Denhière, G. (2004b). Retrieval structure construction during reading : Experimentation and simulation. In *Proceedings of the 26<sup>th</sup> Annual Meeting of the Cognitive Science Society*, Chicago.
- Bellissens C., Thérouanne, P. & Denhière G. (sous presse). Deux modèles vectoriels de la mémoire sémantique : Description, théorie et perspectives. *Le Langage et L'homme : Logopédie, Psychologie, Audiologie*.
- Blanc, N. & Brouillet, D. & (2003). *Mémoire et compréhension. Lire pour comprendre*. Paris : Éditions in Press.
- Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210.
- Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space : Word, sentence, discourse. *Discourse Processes*, 25, 211-257.
- Caillies, S., & Denhière G. (2001). The interaction between textual structures and prior knowledge: Hypotheses, data and simulation. *European Journal of Psychology of Education*, 16, 17-31.
- Caillies, S., Denhière, G., & Jhean-Larose, S. (1999). The intermediate effect : Interaction between prior knowledge and text structure. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representation during reading* (pp. 151-168). Mahwah: Lawrence Erlbaum Associates.
- Caillies, S., Denhière, G., & Kintsch, W. (2002). The effect of prior knowledge on understanding from text : Evidence from primed recognition. *European Journal of Cognitive Psychology*, 14, 267-286.
- Collins, A.M, & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-248.
- Costermans, J. & Fayol, M. (1997). *Processing interclausal relationship*. Mahwah: Lawrence Erlbaum Associates.

Costermans, J. (1998). *Les activités cognitives. Raisonnement, décision et résolution de problèmes*. Bruxelles : De Boeck & Larcier.

de La Haye, F. (2003). Normes d'associations verbales chez des enfants de 9, 10 et 11 ans et des adultes. *L'Année Psychologique* 103, 109-130.

Denhière, G. (1975). Mémoire sémantique, conceptuelle ou lexicale ? *Langages*, n° 40, 41-73.

Denhière, G. & Baudet, S. (1992). *Lecture, compréhension de texte et science cognitive*. Paris : Presses Universitaires de France.

Denhière, G., & Lemaire, B. (2003). Modélisation des effets contextuels par l'analyse de la sémantique latente. In J.M.C. Bastien (Ed.), *Actes des Deuxièmes Journées d'Étude en Psychologie Ergonomique – EPIQUE 2003*, pp. 289-294. Rocquencourt : Inria.

Denhière, G., & Lemaire, B. (2004a). Representing children's semantic knowledge from a multisource corpus. *Fourteenth Annual Meeting Society for Text and Discourse*, Chicago, August 1-4.

Denhière, G., & Lemaire, B. (2004b). A computational model of children's semantic memory. *In Proceedings of the 26<sup>th</sup> Annual Meeting of the Cognitive Science Society*, Chicago.

Denhière, G., Lemaire, B., Bellissens, C., & Jhean-Larose, S. (in press). A semantic space for modeling children's semantic memory. In Graesser, A.C., & McNamara, D. (Eds.), *Latent Semantic Analysis : A road to meaning*. Mahwah: Lawrence Erlbaum Associates.

Denhière, G. & Rossi, J.P. (Eds.) (1991). *Text and text processing*. Amsterdam : Elsevier Science Publishers, North-Holland.

Denhière, G. & Tapiero, I. (1996). La signification comme structure émergente : de l'accès au lexique à la compréhension de textes. In V. Rialle & D. Payette (Eds.), *Penser l'esprit. Des sciences de la cognition à une philosophie cognitive* (pp. 307-336). Grenoble : Presses Universitaires de Grenoble.

Denhière, G., & Théroüanne, P. (2000). *Homograph processing in single-word context*. Communication to the 41th Annual meeting of the Psychonomic Society. New-Orleans, November 16-19.

Ehrlich, M.F. (1994). *Mémoire et compréhension du langage*. Lille : Presses Universitaires de Lille.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 87, 215-251.

Fellbaum, C. (Ed.) (1998). *WordNet : An electronic lexical database*. Cambridge : Cambridge University Press.

Foltz, P.W. (1996). Latent Semantic Analysis for text-based research. *Behavior Research methods, Instruments and Computers*, 28-2, 197-202.

François, J. & Denhière, G. (1997). *Sémantique linguistique et psychologie cognitive : aspects théoriques et expérimentaux*. Grenoble : Presses Universitaires de Grenoble.

Gaume, B. (2003). Analogie et proxémie dans les réseaux petits mondes. Regards croisés sur l'analogie. In Duvignau, K., Gaume, B., et Gasquet, O (Eds.), *Revue d'Intelligence Artificielle* (84-109), N° 5-6, Hermès Sciences.

Gildea, D., & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28, 3, 245-288.

Gineste, M.D. & Le Ny, J.F. (2002). *Psychologie cognitive du langage. De la reconnaissance à la compréhension*. Paris : Dunod.

Jhean-Larose, S. (1991). L'apprentissage d'un système fonctionnel complexe. *Psychologie Française*, 36, 167-176.

Jovet, C., Jhean-Larose, S., Bellissens, C., & Denhière, G. (2004). The accessibility of goal during reading: The role of repetition of subgoal-related information. *Fourteenth Annual Meeting Society for Text and Discourse*, Chicago, August 1-4.

Kintsch, W. (1988). The role of knowledge in discourse comprehension□ A construction-integration model. *Psychological Review*, 2, 163-182.

Kintsch, W. (1993). Information accretion and reduction in text processing□ Inferences. *Discourse Processes*, 16, 193-202.

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

Kintsch, W. (2000). Metaphor comprehension□ A computational theory. *Psychonomic Bulletin and review*, 7, 257-266.

Kintsch, W. (2001). Predication. *Cognitive Science*, 25, 173-202.

Kintsch, W., Patel, V., & Ericsson, K.A. (1999). The role of long-term working memory in text comprehension. *Psychologia*, 42, 186-198.

Landauer, T.K. (1998). Learning and representing verbal meaning : Latent Semantic Analysis theory. *Current directions in Psychological Science*, 7, 161-164.

Landauer, T.K. (2002). On the computational basis of learning and cognition : Arguments from LSA. In B.H. Ross (Ed.), *The psychology of learning and motivation*. New York : Academic Press.

Landauer, T.K., Foltz, P. , & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem : The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Lemaire, B., & Denhière, G. (2004). Incremental construction of an associative network from a corpus. In *Proceedings of the 26<sup>th</sup> Annual Meeting of the Cognitive Science Society*, Chicago.

Lemaire, B., & Denhière, G. (à paraître). A model and a computer program for simulating text comprehension. *Behavior Research Methods, Instruments and Computers*.

Lemaire B. & Dessus P. (2001) A System to Assess the Semantic Content of Student Essays. *Journal of Educational Computing Research* 24(3), 305-320.

Le Ny, J.F. (1979). *La sémantique psychologique*. Paris : Presses Universitaires de France.

Manning, C.D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

Marouby-Terriou, G., & Denhière G. (2002). Identifier l'écrit : Influence des connaissances infra-lexicales. *Enfance*, 4, Vol. 54, 381-407.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466.

Montoya, S., Denhière G., & Baccino, T. (2002). Resolving pronominal ambiguity: Role of contextual constraints. *Current Psychology Letters*, 8, 7-21.

Mullet, V., & Denhière, G. (1997). Accès au lexique et ambiguïtés lexicales nominales : Effet de la polarité des homographes et de la nature du contexte. In J. François & D. Denhière (Eds), *Sémantique linguistique et psychologie cognitive : Aspects théoriques et expérimentaux*. Grenoble : Presses Universitaires de Grenoble.

Myake, A., & Shah, P. (Eds.) (1999). *Models of working memory : Mecanisms of active maintenance and executive control*. New York, Cambridge University Press.

Myers, J.L., & O'Brien, E.J. (1998). Accessing discourse representation during reading. *Discourse Processes*, 26, 131-157.

O'Reilly, R.C., & Rudy, J.W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, 10, 389-397.

Pariollaud, F., Denhière G. & Verstiggel, J.C. (2003). Le traitement des expressions idiomatiques. Intérêt d'un corpus et de l'analyse sémantique latente (307-337). In Tijus, C. & Cornuejol, *Analogie et métaphore*, Paris : Hermès.

Patel, M., Bullinaria, J.A., & Levy, J.P. (1997). Extracting semantic representations from large text corpora. In J.A. Bullinaria, D.W. Glasspool and G. Houghton (Eds.), *Fourth Neural Computation and Psychology workshop : Connectionist representations* (pp. 199-212). London : Springer Verlag.

Ploux, S., & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes. *TAL*, 39.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review*, 88, 93-134.

Rialle, V. & Payette, D. (Eds.) (1996). *Penser l'esprit. Des sciences de la cognition à une philosophie cognitive*. Grenoble : Presses Universitaires de Grenoble.

Rumelhart, D.E. & Mc Clelland, J. (1987) Learning the past tenses of english verbs : implicit rules or parallel distributed processes ?. In B. Mc Whinney (Eds), *Mechanisms of language acquisition*, Hillsdale, NJ, Lawrence Erlbaum.

Till, R.E., Mross, E.F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition*, 16, 283-298.

Tapiero, I., & Denhière, G. (1995). Simulating recall and recognition by using Kintsch's Construction-Integration model. In C.A. Weaver, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension : Essays in honor of Walter Kintsch* (pp. 211-232). Hillsdale, N.J. : Lawrence Erlbaum Associates.

Thérouanne, P., & Denhière G. (2002). Effet d'un contexte lexical sur l'accès à la signification d'homographes. *L'Année psychologique*, 102, 31-63.

Thérouanne, P., & Denhière G. (2004). Normes d'association libre et fréquences relatives des acceptions pour 162 mots homonymes. *L'Année Psychologique*, 104, 537-595.

Tiberghien, G. (1997). *La mémoire oubliée*. Liège : Mardaga.

Till, R.E., Mross, E.F., & Kintsch, W. (1988). Time course of priming associates and inference words in a discourse context. *Memory and Cognition*, 16, 283-298.

van den Broek, P., Ridsen, K., Fletcher, C.R. & Thurlow, R. (1996). A « landscape view of reading » : Fluctuating patterns of activation and the construction of memory representation. In B.K. Britton & A.C. Graesser (Eds.) *Models of understanding text* (pp. 165-187). Mahwah : Lawrence Erlbaum Associates.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. San Diego, C.A. : Academic Press.

Wolfe, M.B.W., Schreiner, M.E., Rehder, B., Laham, D., Foltz, P., Kintsch, W. & Landauer, T.K. (1998). Learning from text : Matching Readers and Texts by Latent Semantic Analysis. *Discourse Processes* 25, 309-336.