



HAL
open science

MÉMOIRE ET LANGAGE: APPORTS DE L’“ ANALYSE DE LA SÉMANTIQUE LATENTE ” À L’ÉTUDE DU DÉVELOPPEMENT

Sandra Jhean-Larose, Guy Denhière

► **To cite this version:**

Sandra Jhean-Larose, Guy Denhière. MÉMOIRE ET LANGAGE: APPORTS DE L’“ ANALYSE DE LA SÉMANTIQUE LATENTE ” À L’ÉTUDE DU DÉVELOPPEMENT. *Enfance*, 2019. hal-03752681

HAL Id: hal-03752681

<https://hal.science/hal-03752681>

Submitted on 17 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MÉMOIRE ET LANGAGE : APPORTS DE L'« ANALYSE DE LA SÉMANTIQUE LATENTE » À L'ÉTUDE DU DÉVELOPPEMENT

Sandra Jhean-Larose et Guy Denhière

Presses Universitaires de France | « [Enfance](#) »

2019/3 N° 3 | pages 395 à 411

ISSN 0013-7545

ISBN 9782130821274

Article disponible en ligne à l'adresse :

<https://www.cairn.info/vue-enfance-2019-3-page-395.htm>

Distribution électronique Cairn.info pour Presses Universitaires de France.

© Presses Universitaires de France. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Mémoire et langage : Apports de l'« Analyse de la Sémantique Latente » à l'étude du développement

Sandra Jhean-Larose^a et Guy Denhière^b

RÉSUMÉ

Les modèles actuels de la mémoire à long terme et de la mémoire sémantique supposent que des représentations abstraites (concept, schémas, connaissances) sont stockées et élaborées à partir d'expériences singulières selon un principe « abstraitif » de généralisation des informations pouvant produire des espaces sémantiques à partir de larges corpus de textes. Nous montrons comment l'Analyse de la Sémantique Latente (LSA) permet de modéliser et simuler chez les enfants : les associations verbales, les significations lexicales, le rappel et le résumé de récits, les Jugements d'Importance Relative et l'acquisition du vocabulaire.

MOTS-CLÉS : ANALYSE DE LA SÉMANTIQUE LATENTE (LSA), ENFANTS, ESPACES SÉMANTIQUES.

ABSTRACT

Memory and language : Contribution of « Latent semantic analysis » to the study of development

Current models of long-term memory and semantic memory assume that abstract representations (concepts, schemas, knowledge) are stored and developed from singular experiments according to an “abstract” principle of generalization of information that can produce semantic spaces from large corpus texts. We show how Latent Semantic Analysis (LSA) simulates verbal associations, lexical meanings, recall and summary of narratives, Judgment of Relative Importance and vocabulary acquisition in younger children.

KEYWORDS: LATENT SEMANTIC ANALYSIS (LSA), SEMANTIC SPACES, NARRATIVES, VERBAL ASSOCIATIONS, LEXICAL MEANINGS, RECALL, SUMMARY, RELATIVE IMPORTANCE, VOCABULARY.

^a ERCAE, université d'Orléans, 72 rue du faubourg de Bourgogne, 45000 Orléans. *E-mail* : sandra.jhean-larose@univ-orleans.fr

^b CHArt, École pratique des hautes études, 4-14 rue Ferrus, 75014 Paris. *E-mail* : Guy.Denhiere@ephe.psl.eu

INTRODUCTION

Le but principal de cet article consiste à exposer une démarche de recherche qui se préoccupe d'articuler explicitement les activités cognitives impliquées dans les différents aspects du traitement du langage à l'organisation de l'information en mémoire – qu'il s'agisse d'enfants ou d'adultes – en prenant en compte les apports récents de la sémantique cognitive (Le Ny, 1979, 2005 ; François & Denhière, 1997) et les possibilités nouvelles apportées par l'implantation informatisée des modèles cognitifs (Kintsch, 2010). La formulation d'hypothèses qui satisfassent simultanément les contraintes du système linguistique et du système cognitif oblige à rechercher des invariants cognitifs de l'expérience du monde, invariants qui se construisent au cours du développement de l'enfant et de l'adolescent (Denhière & Baudet, 1992). Les recherches réalisées ces trente dernières années ont permis d'identifier de tels invariants cognitifs – au niveau local : schémas d'individus, d'états et de propriétés, d'événements et d'actions, organisés en représentations plus générales : système d'état relationnel, système transformationnel et système téléologique (Baudet & Denhière, 1991 ; Jhean-Larose, 1991). La recherche de ces invariants cognitifs (Baudet, 1990) a conduit la psychologie cognitive à replacer au centre de ses préoccupations le « concept-mémoire », un temps « oublié » comme le rappelle le titre de l'ouvrage publié par Tiberghien (1997), et profondément renouvelé par la formulation de modèles à « appariement global », qui représentent les informations stockées sous la forme de vecteurs (TODAM, Murdock, 1993 ; CHARM, Eich-Metcalf, 1985 ; MINERVA, Hintzman, 1984).

Ces « modèles abstractifs » de la mémoire (Tiberghien, 1997) ont largement inspiré les principaux modèles de compréhension de textes proposés par Kintsch (1988, 1998) et Myers et O'Brien (1998), le premier postulant une étape de « construction » suivie d'une phase « d'intégration » des représentations construites aux connaissances stockées en mémoire, le second proposant une conception du « traitement dirigé par la mémoire ». Ainsi envisagée, la compréhension et la mémorisation d'un même texte varieront en fonction des connaissances stockées en mémoire et donc des caractéristiques des apprenants, de leur niveau de développement notamment. Sur un plan formel, ces « modèles abstractifs » décrivent la mémoire comme un « hyper-espace » à plusieurs centaines de dimensions dans lequel les représentations mnésiques, les signifiés sont exprimés par des vecteurs à n dimensions, ce qui permet de les comparer et d'estimer leur similitude. Une telle conception permet également une implémentation informatique des modèles de représentation des informations en mémoire et d'acquisition de connaissances tels « Hyperspace Analog to Language » (HAL, Lund et Burgess, 1996) et l'Analyse de la Sémantique Latente (LSA, Landauer & Dumais, 1997 ; Landauer, McNamara, Dennis & Kintsch, 2007). Ces modèles informatisés implantent l'hypothèse selon laquelle les associations épisodiques stockées en mémoire par un individu au cours de

son développement peuvent être généralisées et transformées en connaissances, et explicitent les rapports entre confrontation à l'environnement (notamment langagier) et élaboration des représentations en mémoire. La mise à l'épreuve d'une telle conjecture implique, sur le plan expérimental, une démarche qui consiste (1) à élaborer une structure de mémoire la plus représentative possible d'une population donnée, enfants ou adultes, (2) à l'implanter sous la forme d'un modèle informatique, (3) tester empiriquement la pertinence cognitive de cette mémoire par la confrontation aux données expérimentales disponibles et, (4) l'utiliser dans le plus grand nombre possible d'expériences de traitement langagier afin d'en déterminer cas par cas la pertinence afin de juger de l'intérêt scientifique de la conjecture initiale. Les pages qui suivent se veulent un reflet de l'utilisation de cette démarche.

PRÉSENTATION DE L'ANALYSE DE LA SÉMANTIQUE LATENTE (LSA)

Pour nous en tenir à l'« Analyse de la Sémantique Latente » (LSA), la construction d'un espace sémantique comporte trois étapes. Premièrement, il s'agit de rassembler un vaste corpus de textes supposé représentatif de l'environnement langagier auquel ont été confrontées les personnes dont on veut étudier les comportements. De la représentativité du corpus ainsi élaboré, découle la pertinence des analyses ultérieurement réalisées. Dans un second temps, une matrice est constituée à partir du corpus. L'information contenue dans cette matrice correspond à la distribution de chaque terme dans l'environnement linguistique des textes du corpus. Enfin, à partir de la matrice, sont constitués les vecteurs qui vont représenter les termes dans un espace de grande dimension. La distance sémantique entre deux termes : associés d'un terme inducteur, ou expressions : texte lu *versus* texte restitué par exemple, peut alors être calculée. La mesure la plus fréquemment utilisée consiste dans le calcul du cosinus de l'angle que forment les vecteurs représentant les termes et expressions comparées dans l'espace considéré.

LE CORPUS « TEXTENFANT »

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 familiers (52 %), de productions verbales d'enfants de cours préparatoire et de cours élémentaire (23 %), d'extraits de manuels de lecture de cours préparatoire (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. 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, Lemaire, Bellissens et Jhean-Larose, 2007).

TESTS DE L'ESPACE SÉMANTIQUE « TEXTENFANT »

1^{er} test : associations verbales

Un premier test de l'espace sémantique « TextEnfant » vise à déterminer dans quelle mesure il rend compte des normes associatives enfantines publiées par De La Haye (2003). 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 sémantique « 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 » constitue une approximation raisonnable 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 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 ^{er} associé	2 ^e associé	3 ^e 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 correcte la hiérarchie des termes associés aux 200 mots inducteurs utilisés par de La Haye (2003).

2^e test : définitions lexicales

Un deuxième 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 (1990) 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 (correcte),.24 (proche),.16 (éloigné) et.04 (sans relation). À 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 » à un niveau intermédiaire entre les réussites 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, 2003).

Une mesure plus sophistiquée mais plus satisfaisante consiste à calculer une différence normalisée d entre réponses correcte et proche *versus* réponses éloignées 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	3,2	53	69	.39
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 ($p < .03$). Ces résultats corroborent de façon éloquente l'hypothèse formulée en début d'article selon laquelle il est nécessaire de prendre en compte les représentations en mémoire pour étudier le traitement du langage. Dans le cas présent, la « mémoire enfantine » représentée par « TextEnfant », bien que quantitativement plus limitée que les corpus adultes (3 millions versus 14 à 20 millions d'occurrences), mais constituée de textes représentatifs de l'environnement langagier auquel ont été confrontés les enfants au cours de leur développement, permet une meilleure appréhension de leurs performances.

3^e test : rappel et résumé de récits

L'espace sémantique « TextEnfant » peut être utilisé pour évaluer automatiquement les rappels et les 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 (LSA), la mesure de la quantité d'information restituée est estimée par la valeur du cosinus entre le vecteur représentant le texte source (à rappeler ou à résumer) et le rappel ou le 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. Les textes utilisés sont extraits de la batterie DiagnosTM qui comporte deux séries de 16 récits calibrés par le nombre d'épisodes, de phrases, de propositions, de mots (Denhière, Legros & Tapiero, 1993). L'annexe reproduit les récits mentionnés dans cette expérience.

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
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 sont largement significatives et s'étagent de 45 à 92. 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 prédicative longue et fastidieuse.

4^e test : jugements d'importance relative (JIR)

Le but de cette recherche était d'étudier les processus cognitifs implicites dans les jugements de l'importance relative (JIR) des phrases d'un récit sachant que, de manière répétée, on a démontré que ces jugements étaient significativement corrélés avec l'ordination de la fréquence de rappel et de résumé d'un texte. L'hypothèse soumise au contrôle expérimental consistait à supposer que le modèle de l'analyse de la sémantique latente (LSA) permettait de simuler les jugements humains du fait que les phrases importantes devaient entretenir un degré de similarité élevé entre elles et avec l'ensemble du texte.

Seize récits extraits de la batterie DIAGNOSTTM (Baudet et Denhière, 1989 ; Denhière et Baudet, 1990) égalisés pour le nombre de mots, de phrases (n=30), de propositions sémantiques, d'arguments étaient utilisés. 4 groupes de 30 étudiants devaient juger de l'importance relative de chacune des 30 phrases de 4 récits sur une échelle en 5 points. Pour chacune des phrases des 16 récits, on disposait donc d'une valeur moyenne d'importance relative, valeur qui était

significativement corrélée avec le rappel et le résumé de ces textes produits par des enfants et des adolescents.

Les résultats obtenus avec l'espace sémantique « TextEnfant », représentant la mémoire enfantine, implantés à l'Institut des Sciences Cognitives de l'Université du Colorado à Boulder¹ à l'Université de Würzburg² et à l'École Pratique des Hautes Etudes (EPHE) à Paris.

Le tableau 4 présente les corrélations obtenues entre les jugements humains et les mesures LSA pour l'espace sémantique « TextEnfant ». À l'évidence, d'importantes différences sont observées entre les récits avec un bloc de très bons résultats pour 6 d'entre eux, acceptables pour 4 et insuffisant pour les autres. Ce constat invite à étudier plus en détail le vocabulaire utilisé dans les récits et le lexique acquis par chaque individu.

5^e test : acquisition du vocabulaire

L'objectif poursuivi consiste à identifier le vocabulaire acquis par les enfants du cycle 2 (CP et CE1) en utilisant une échelle qui postule une difficulté croissante des mots à définir correspondant à un ordre d'acquisition hiérarchisé et stable (Biemiller, 2009) et en analysant les définitions produites à l'aide de l'Analyse de la Sémantique Latente (LSA) dans le but de procéder à une analyse automatique fine, au-delà des bornes traditionnelles (0, 0,5 et 1) des réponses des enfants.

Biemiller (2009) a repris les résultats de Dale et O'Rourke (1981) qui ont établi la liste des mots possédés à chaque âge (de 8 à 20 ans) par 75 % des individus et qui ne l'étaient pas à l'âge précédent. L'échelle se compose de 60 mots divisés en 10 déciles de 6 items ordonnés par difficulté croissante, chaque mot à définir est présenté au sein d'une phrase et il est demandé à l'enfant de définir le mot dans le contexte de la phrase. Nous avons traduit cette échelle en l'adaptant à la réalité française. Sur les 60 items que comporte l'échelle, seuls les 42 premiers (soit les 7 premiers déciles) sont présentés au CP, et les 48 premiers (soit les 8 premiers déciles) au CE1. L'épreuve est administrée individuellement, l'enfant est invité à s'exprimer le plus possible sans être jamais ni détrompé ni corrigé. La consigne, la plus succincte possible, précise qu'une phrase va s'afficher à l'écran, « qu'on va la lire ensemble », « que dans cette phrase un mot est écrit en italiques (penché) en bleu » et qu'il va être interrogé sur « ce mot en bleu », qu'il va « devoir expliquer ce mot, dire ce qu'il veut dire ». La tâche est explicitée à l'aide de 4 exemples de difficulté croissante. L'ensemble de la réponse formulée par l'enfant est relu à voix haute par l'expérimentateur afin d'obtenir son approbation avant de passer à l'item suivant.

Quatre-vingt-sept enfants de CP (47 filles et 40 garçons ; Age : $m = 6,48$, $\sigma = 0,37$), et quatre-vingt-sept enfants de CE1 (39 filles et 48 garçons ; Age :

¹ <http://lsa.colorado.edu>

² <http://www.summa.psychologie.uniwuerzburg.de>

Tableau 4.

Corrélations entre les jugements humains d'importance relative des phrases des 16 récits et les mesures LSA entre les phrases et le texte entier pour les espaces sémantiques Enfants.

	Paris-EPHE TextEnfant	Boulder- Contes	Würzburg TEnfant
Géant	0,60	0,40	0,60
Dragon	0,57	0,48	0,52
Taureau	0,45	0,30	0,42
Âne	0,43	0,20	0,31
Chamois	0,41	0,41	0,32
Clowns	0,33	0,38	0,3
Poule	0,25	0,02	0,35
Rennes	0,22	0,44	0,21
Araignée	0,18	0,17	0,32
Pièces d'or	0,04	0,13	-0,01
Ourson	0,13	0,23	0,07
Camion	0,24	0,26	0,26
Lebrac	0,14	0,24	-0,06
Dauphin	0,19	-0,09	0,16
Lion	0,23	0,22	0,22
Singes	-0,14	-0,02	-0,06
Moyenne	0,27	0,24	0,25
Écart type	0,19	0,17	0,19
Mediane	0,23	0,24	0,28

m = 7.58, $\sigma = 0.46$) fréquentant deux écoles parisiennes ont participé à l'épreuve. Les réponses des enfants sont évaluées selon deux modalités : par des juges humains à partir d'une définition de référence et à l'aide de LSA. La présence d'un contenu minimal rassemblant les traits définitoires essentiels du concept à définir permettait d'obtenir la note 1 ; l'absence de certain(s) trait(s) définitoires, le sens global du mot étant préservé, conduisait à la note 0,5 ; les autres cas à la note 0. Pour l'évaluation LSA, une définition « standard », comportait à la fois les traits définitoires précédents et des éléments extensifs.

L'espace sémantique « TextEnfant » (Denhière, Lemaire, Bellissens & Jhean-Larose, 2007) était utilisé et, pour chaque item, le cosinus entre la définition de référence et la réponse de l'enfant était calculé.

Tableau 6.

Moyennes des résultats obtenus par les élèves de CP (n=39) et de CE1 (n=48) aux critères de réussite considérés : Score total (notes 1 et notes 1+0,5), score atteint après 6, 8 et 10 échecs (omissions et erreurs) et dernier item correctement défini après 6, 8 et 10 échecs. La dernière ligne indique la différence entre les performances moyennes des 2 classes de CE1 et de CP.

	Score total sur 48 items		Score atteint après			Dernier item défini après		
	Note 1	Note 1+0,5	6 échecs	8 échecs	10 échecs	6 échecs	8 échecs	10 échecs
CP	19,5	21,9	14,0	16,3	17,6	21,3	25,7	29,2
CE1	26,3	28,6	20,7	23,2	25,2	27,9	32,7	36,8
Différence	6,8	6,7	6,7	6,9	7,6	6,7	7,0	7,6

Concernant les évaluations à l'aide de l'analyse de la sémantique latente (LSA), comme le montre la figure 1, on observe sur les 8 déciles administrés une diminution du score moyen aux 6 items d'un décile lorsque la difficulté des mots du décile augmente ($F(8,166) = 322.9, p < .001$), la moyenne des scores aux 10 déciles est significativement plus faible en CP ($m = .138, \sigma = .027$) qu'en CE1 ($m = .155, \sigma = .058$) ($F(1,172) = 5.55, p < .02$).

La comparaison entre les jugements humains et l'évaluation par LSA (voir la figure 2), témoigne d'une corrélation significative entre la moyenne des notes établies par le jugement humain et la moyenne des évaluations calculées par LSA : $r(174) = .58, p < .001$.

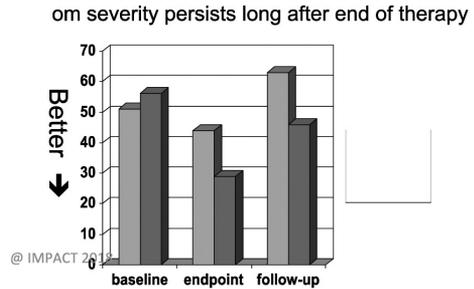
En conclusion, les résultats obtenus mettent en évidence la pertinence de l'approche développée par Biemiller (2009) pour évaluer l'étendue du vocabulaire possédé par les enfants. Par ailleurs, nous avons pu démontrer l'intérêt de l'utilisation de l'Analyse de la Sémantique Latente (LSA) pour évaluer automatiquement les protocoles recueillis.

CONCLUSION

Nous avons montré que LSA, comme exemple des modèles statistiques qui infèrent les structures sémantiques de la distribution des mots dans de vastes corpus (Kintsch, 2010), peut être considéré comme un outil qui permet de modéliser différents aspects du traitement du langage naturel -- non seulement chez les adultes, mais également chez les enfants -- qu'il s'agisse des associations verbales, des significations lexicales, du rappel et du résumé de récits, des Jugements d'Importance Relative et de l'acquisition du vocabulaire. Si LSA peut être décrite comme une approximation acceptable de la mémoire humaine,

Better
↘

14/09/2018



ADOS CSS, Autism Diagnostic Observation Schedule Combined Severity Scale

Figure 2.

Analyse automatique des définitions données par les élèves de CP et de CE1 à l'aide de l'Analyse de la Sémantique Latente (LSA) : diminution significative des cosinus par décile.

PACT therapy effect on child behaviour

Increase in child social communication with parent persisted

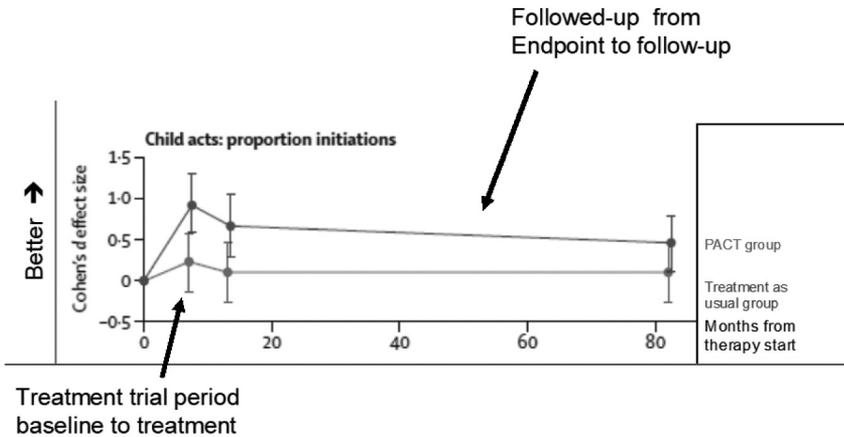


Figure 1.

Analyse automatique des définitions à l'aide de l'Analyse de la Sémantique Latente (LSA) : Corrélations entre les jugements humains et les valeurs de cosinus LSA.

enfantine ou adulte, elle peut également être utilisée en complément d'autres technologies, telles TOPICS (Steyvers & Griffiths, 2007 ; Jhean-Larose, Denhière, Leveau & NGuyen, 2016) ou la prétopologie (Ahat, Ben Amor, Bui, Jhean-Larose & Denhière, 2010). LSA peut également être appliquée à l'acquisition des connaissances évaluée à partir de QCM et de questions ouvertes (Jhean-Larose, Leclercq, Diaz, Denhière & Bouchon-Meunier, 2010 ; Diaz, Rifqi, Bouchon-Meunier, Jhean-Larose & Denhière, 2008 ; Lifchitz, Jhean-Larose & Denhière, 2009) comme à la modélisation et à l'identification automatique de la coloration émotionnelle des textes (Denhière, Leveau & Jhean-Larose, 2010 ; Denhière, Leveau, Jhean-Larose & Nguyen, 2011 ; Leveau, 2011). Il devient ainsi possible d'étudier de manière conjointe les aspects cognitifs et affectifs de la compréhension et de l'acquisition des connaissances, sans dissocier ces deux composantes de la vie intellectuelle et du développement.

REMERCIEMENTS

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.

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 (Chesnet & Lambert, 2002).

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 second auteur.

RÉFÉRENCES

- Ahat, M., Ben Amor, S., Bui, M., Jhean-Larose, S. et Denhière, G. (2010). Document classification with LSA and pretopology, *In*: Jhean-Larose, S. & Denhière, G. (2010), Editors, *Learning, Memory and Semantic Analysis (LSA)*, *Studia Informatica Universalis*, Vol. 8.1, Hermann Informatique, p. 121-138.
- Ahat, M., Lenhard, W., Hoareau, V., Jhean-Larose, S. & Denhière, G. (2007). Le concours DEFT'07 envisagé du point de vue l'Analyse de la Sémantique Latente (LSA). A LSA approach in DEFT'07 contest. *In Actes de l'atelier DEFT'07*, (p. 135-143), Grenoble (France).
- Baier, H., Lenhard, W., Hoffmann, J. & Schneider, W. (2010). SUMMA-A LSA Integrated Development System, *In*: Jhean-Larose, S. & Denhière, G. (2010), *Editors, Learning, Memory and Semantic Analysis (LSA)*, *Studia Informatica Universalis*, Vol. 8.1, Hermann Informatique, 25-49.
- Baudet, S. (1990). Représentation d'état, d'événement et d'action. *In* J. François & G. Denhière, G. (Eds.), *Langages: Cognition et Langage*, 100, p. 45-64.

- Baudet, S., & Denhière, G. (1989). Le fonctionnement cognitif dans la compréhension de texte. In: S. Baudet & G. Denhière (Eds.), *Le diagnostic du fonctionnement cognitif dans la compréhension et la production de textes. Questions de Logopédie*, 21, II, 15-30.
- Baudet, S., & Denhière, G. (1991). Mental models and acquisition of knowledge from text: Representation and acquisition of a functional system. In G. Denhière & J.P. Rossi (Eds), *Text and text processing*. Amsterdam, Elsevier Science Publication, North Holland.
- Bellissens, C, Théroüanne, P. & Denhière, G. (2004). Deux modèles vectoriels de la mémoire sémantique : Description, théorie et perspectives. *Le Langage et l'Homme*, 2004, 1, 94-107.
- Biemiller, A. (2009). *Words Worth Teaching*. Columbus, OH:SRA/McGraw-Hill.
- Dale, E. & O'Rourke, J. (1981). *The Living Word Vocabulary*. Chicago: World Book/Childcraft International.
- de La Haye, F. (2003). Normes d'association verbale chez des enfants de 9, 10 et 11 ans et des adultes, *L'Année Psychologique*, 103, 109-130.
- 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 – ÉPIQUE 2003*, 289-294. Rocquencourt : Inria.
- Denhière, G., & Baudet, S. (1992). *Lecture, compréhension de textes et science cognitive*, Paris, Presses universitaires de France.
- Denhière, G., & Baudet, S. (1990). Le diagnostic du fonctionnement cognitif dans la compréhension de textes. *Glossa*, 20, 10-17.
- Denhière, G., & Lemaire, B. (2004). Representing Children's Semantic Knowledge from a multisource corpus. *Proceedings of the Fourteenth Annual Meeting Society for Text and Discourse*; Chicago, August, 1-4.
- Denhière, G., & Rossi, J.P. (1991) (Eds.). *Text and text processing*. Amsterdam, Elsevier Science Publication, North Holland.
- Denhière, G., Hoareau, V., Jhean-Larose, S., Lehnard, W., Bäier, H., Bellissens, C. (2007). Human Hierarchization of Semantic Information in Narratives and Latent Semantic Analysis, *Proceedings in the 1st European Workshop on Latent Semantic Analysis in Technology-Enhanced Learning*, March 29-30, Heerlen, NL. L 08), Maastricht, The Netherlands, Springer-Verlag, pp. 144-154.
- Denhière, G., Larget, E. (1990). Étude du rappel de récit : influence de l'âge, de la structure des épisodes, de leur ordre de présentation et du délai temporel entre la présentation et le rappel. *Glossa*, 19, 14-34.
- Denhière, G., Legros, D., & Tapiero, I. (1993). Representation in memory and acquisition of knowledge from text and picture: Theoretical, methodological, and practical outcomes. *Educational Psychology Review*, 5(3), 311-324.
- Denhière, G., Lemaire, B., Bellissens, & Jhean-Larose, S. (2004). Psychologie cognitive et compréhension de texte : Une démarche théorique et expérimentale. In S. Porhiel & D. Klingler (Eds.), *L'Unité texte*, (p. 74-95). Pleyben, Perspectives.
- Denhière, G., Lemaire, B., Bellissens, C. & Jhean-Larose, S. (2007). Semantic space for modeling a child semantic memory. In: Landauer, T.K., McNamara, D., Dennis, S. & Kintsch, W. (Eds), *Handbook of Latent Semantic Analysis*, 143-165. Mahwah, NJ: Erlbaum.

- Denhière, G., Lemaire, B., Bellissens, C. & Jhean-Larose, S. (2007). A semantic space for modeling a child semantic memory. Landauer, T.K., McNamara, D., Dennis, S. & Kintsch, W. (Eds), *Handbook of Latent Semantic Analysis* (p. 155-176). Mahwah, NJ: Erlbaum.
- Denhière, G., Leveau, N. & Jhean-Larose, S. (2010). EMOSEM : un modèle d'identification automatique de la coloration émotionnelle des textes. *Actes du Congrès de la Société Française de Psychologie*, Lille, 7-9 septembre, 129-130.
- Denhière, G., Leveau, N. & Jhean-Larose, S. (2010). Mémoire, représentation sémantique et extraction automatique de thèmes. *Actes du congrès de la Société française de psychologie*, Lille, 7-9 septembre, 317-318.
- Denhière, G., Leveau, N. Jhean-Larose, S. & Nguyen, B.L. (2011). How to identify emotions in texts using LSA. *Proceedings of the 21st Annual Meeting of the Society for Text and Discourse*, Poitiers, July 11-13.
- Diaz, J., Rifqi, M., Bouchon-Meunier, B., Jhean-Larose, S., & Denhière, G. (2008). Imperfect answers in multiple choice questionnaires. *Proceedings of the 3rd European Conference on Technology Enhanced Learning: Technologies across learning contexts* (EC-TE Best student paper award).
- Eich-Metcalfe J. (1985) Levels of processing, encoding specificity, elaboration, and CHARM, *Psychological Review*, 91, 1-38.
- François, J. & Denhière, G. (1990). Cognition et Langage. *Langages*, n° 100.
- François, J. & Denhière, G. (1997). *Sémantique linguistique et psychologie cognitive, aspects théoriques et expérimentaux*. Grenoble : Presses universitaires de Grenoble.
- Hintzman, D.L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers*, 16, 96-101.
- Jhean-Larose, S. (1991). Apprentissage d'un système fonctionnel complexe. *Psychologie Française*, 36, 2, 167-177.
- Jhean-Larose, S. & Denhière, G. (2010), Editors, Learning, Memory and Semantic Analysis (LSA), *Studia Informatica Universalis*, vol. 8, 1, Hermann Informatique.
- Jhean-Larose, S., Leclercq, V., Diaz J., Denhière, G. & Bouchon-Meunier, B. (2010). Knowledge evaluation based on LSA: MCQs and free answers. *In: Jhean-Larose, S. & Denhière, G. (2010), Editors, Learning, Memory and Semantic Analysis (LSA), Studia Informatica Universalis*, Vol. 8.1., 53-78.
- Jhean-Larose, S., Leveau, N., Denhière, G. & NGuyen, B.L. (2016). Automatic extraction of topics from documents: five probabilistic Topic model tests. *International Journal of New Technology and Research (IJNTR)*, (2), 6, 1-14.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review*, 95, 163-182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W. (2001). Predication, *Cognitive Science*, 25, 173-202.
- Kintsch, W. (2010). Modeling Semantic Memory. *In: Jhean-Larose, S. & Denhière, G. (2010), Editors, Learning, Memory and Semantic Analysis (LSA), Studia Informatica Universalis*, Hermann Informatique, vol. 8, 1, p. 1-21.
- Landauer, T. K. & Dumais, S.T. (1997). A solution of Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge, *Psychological Review*, 104, 211-240.
- Landauer, T.K., McNamara, D., Dennis, S. & Kintsch, W. (Eds), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

- Le Ny, J.F. (1979). *La Sémantique psychologique*, Paris, Presses universitaires de France.
- Le Ny, J.F. (2005). *Comment l'esprit produit du sens*. Odile Jacob.
- Lemaire, B., Denhière, G., Bellissens, C., Jhean-Larose, S. (2006). A Computational Model for Simulating Text Comprehension. *Behavior Research Methods*, 38, 4, 628-637.
- Lemaire, B., Mandin, S., Dessus, P. & Denhière, G. (2005) Computational cognitive models of summarization assessment skills, in *Proceedings of the 27th Annual Meeting of the Cognitive Science Society (CogSci' 2005)*, p. 1266-1271. Stresa, Italy, July 21-23.
- Leveau, N. (2011). *Émotions, sentiments et opinions véhiculés par des textes : expérimentations, modélisation et simulation*. Thèse de doctorat en psychologie, EPHE, Paris.
- Lifchitz, A., Jhean-Larose, S., & Denhière, G. (2009). Effect of tuned parameters on a LSA multiple choice questions answering model. *Behavior Research Methods*, 41(4), 1201-1209.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic space from lexical cooccurrences. *Behavior Research Methods, Instruments & Computers*, 28, 203-208.
- Murdock, B. B. jr., (1982). A theory for the storage and retrieval of items and associative information, *Psychological Review*, 89, 609-626.
- Myers, J.L., O'Brien E.J. (1998). Accessing discourse representation during reading, *Discourse Processes*, 26, 131-157.
- Steyvers, M. & Griffiths, T.L. (2007). Probabilistic topic models. In: T.K. Landauer, D. McNamara, S. Dennis & W. Kintsch (Eds). *Handbook of Latent Semantic Analysis* (pp. 427-448). Mahwah, NJ: Erlbaum.
- Tiberghien, G. (1997). *La Mémoire oubliée*. Liège : Mardaga.

ANNEXE : RÉCITS DE LA BATTERIE DIAGNOSTTM UTILISÉS DANS L'EXPÉRIMENTATION.

L'araignée. Une petite araignée se promenait sur un sentier plein de cailloux. Tracé dans l'herbe, ce sentier longeait une jolie rivière. La petite araignée, qui s'était échappée de chez elle, alla sur l'herbe, près de la rivière. La petite araignée grimpa sur une grosse pierre et, fatiguée, s'endormit dans un creux de la grosse pierre. Un petit garçon s'amusait à lancer des cailloux dans la rivière alors que son papa pêchait un peu plus loin. Le petit garçon poussa la grosse pierre dans la rivière et la petite araignée tomba avec la grosse pierre. Une truite, qui passait à ce moment-là, avala la petite araignée. Le papa du petit garçon remarqua un ver de terre, dans un creux, près du sentier et l'accrocha à son hameçon. Le papa du petit garçon lança sa ligne, attrapa la truite et la donna à sa femme. La femme vida la truite sur l'herbe et libéra la petite araignée. La petite araignée retourna chez elle où ses parents, contents de la revoir, l'attendaient.

Les clowns. Les trapézistes, qui avaient terminé leur numéro, quittèrent la piste illuminée du cirque et sept clowns firent leur entrée. Le premier était déguisé en clochard alors que le deuxième, habillé en paysan, tirait une drôle de vache.

Un clown portait une tête avec des cornes en carton, tandis qu'un autre, accroché à lui, remuait la queue. Un cinquième clown, enfermé dans une carcasse en bois, ressemblait à un tonneau et roulait à toute vitesse sur la piste. Les deux derniers clowns transportaient une grande échelle. L'un mit l'échelle debout pendant que l'autre montait dessus. Quand son camarade fut arrivé en haut de l'échelle, le clown qui tenait l'échelle s'éloigna. L'échelle balançait dangereusement, mais le clown, perché en haut de l'échelle, réussit à redescendre et, une fois à terre, lâcha l'échelle. En tombant l'échelle assomma le clochard, le paysan, la vache, et le tonneau. Le clown qui avait eu peur en haut de l'échelle, riait à son tour.

Le dragon et l'enfant. Il était une fois un vieux dragon qui effrayait les gens du village car il crachait de longues flammes. Le dragon n'avait qu'un seul ami, un jeune garçon nommé Tino qui l'aimait beaucoup. Ce jeune garçon voulait empêcher le dragon de faire peur aux villageois. Tino emmena le dragon se baigner dans un grand lac. Là le dragon plongea, nagea et but beaucoup d'eau. Quand le dragon sortit de l'eau, il ne crachait plus de flammes. Alors, Tino invita le dragon au restaurant pour fêter son anniversaire. Le dragon mangea de la viande très épicée avec des sauces très piquantes et des piments rouges très forts. Quand le dragon eut fini de manger, il souffla les bougies de son gâteau d'anniversaire. Aussitôt, de longues flammes sortirent de la bouche du dragon. Pris de panique, le dragon quitta le restaurant pour se précipiter dans le lac.

Le géant dans la forêt. Il y a bien longtemps, vivait un géant qui s'appelait Gargantua. Ce géant était bon et il aimait rendre service aux pauvres gens. Un jour, le géant se promenait dans une magnifique forêt qui s'étendait à perte de vue. Le géant vit une vieille dame qui, pliée en deux, ramassait du bois mort. Le géant décida d'aider la vieille dame. En un instant, le géant arracha cinq beaux chênes, les attacha avec une corde et les posa sur son épaule. Marchant le plus lentement possible, le géant accompagna la vieille dame jusque chez elle. Comme la route était longue pour arriver chez la vieille dame, le géant fut bien content de déposer les chênes contre la maison de la vieille dame. Sous le poids des chênes, la maison de la vieille dame s'écroula dans un craquement terrible. Si la vieille dame avait maintenant du bois pour se chauffer, elle n'avait plus de maison pour s'abriter.

L'ourson. Depuis le matin, Winnie l'ourson parcourait la forêt pour trouver de la nourriture. Fatigué, l'ourson s'assit sous un grand chêne et s'endormit. L'ourson se mit à rêver à du miel doré qui lui tombait goutte à goutte dans la bouche. Soudain, un bourdonnement réveilla l'ourson. L'ourson se prit la tête entre les pattes et réfléchit. L'ourson se dit que le bourdonnement provenait des abeilles et que, s'il y avait des abeilles, alors il y avait du miel. L'ourson se mit à grimper à l'arbre, s'arrêta pour souffler, puis recommença à monter. L'ourson était presque arrivé quand la branche sur laquelle il était cassa sous son poids. L'ourson tomba au milieu d'un buisson d'épines, à moitié assommé. L'ourson s'assit à nouveau sous le grand chêne, se rendormit et recommença

La poule. Il était une fois une poule nommée Sidonie qui pondait des œufs magnifiques. Un jour, le fermier apprit à Sidonie qu'en Amérique, les poules poussaient des œufs carrés. Le fermier expliqua à Sidonie les avantages des œufs carrés : ils ne roulent pas et ils tiennent debout tout seuls. Convaincue, Sidonie décida de pondre des œufs carrés et, dans ce but, elle changea sa manière de pondre. Sidonie s'allongea sur le dos et se tourna dans tous les sens mais elle pondait toujours des œufs de forme ovale. Sidonie changea alors de nourriture mais elle pondait toujours des œufs ovales. Sidonie acheta alors un livre qui expliquait la manière de construire des cubes et l'ava. Le résultat fut que Sidonie tomba malade et ne pondit plus un seul œuf. Sidonie pensa alors aux inconvénients des œufs carrés. Les pâtisseries devraient changer leurs moules pour fabriquer des œufs de pâques, et les poussins auraient la forme carrée. Alors, Sidonie se dit que les œufs qu'elle pondait étaient beaux et qu'il était inutile de se rendre malade pour pondre des œufs carrés.

L'enfant et le taureau. Le jeune Robert, en vacances à la campagne, décida d'aller voir l'étang où pêchaient les petits paysans. Soudain, Robert entendit un terrible mugissement derrière lui. Surpris, Robert se retourna brusquement et aperçut un grand taureau noir qui le regardait, immobile. Le taureau fonça aussitôt sur Robert qui courut se cacher derrière un gros peuplier. Robert ne pouvait pas rester derrière le peuplier car ce n'était pas un abri sûr. La tête baissée, le taureau préparait une nouvelle attaque. Robert quitta soudainement son abri pour courir vers l'étang. Poursuivi par le taureau, Robert arriva à l'étang et plongea tout habillé. Bien que ses vêtements le gênaient pour nager, Robert réussit à atteindre l'autre rive de l'étang. Robert se retourna et vit le taureau qui, couché dans l'herbe, ruminait tranquillement.