

RÉSUMÉ:

La compréhension de texte en lecture s'appuie notamment sur une capacité de traitement syntaxique de phrases isolées. Afin d'évaluer cette capacité, un paradigme de jugement d'adéquation sémantique (dire si deux phrases ont un sens proche ou non) a été mis au point en manipulant la fréquence lexicale et la complexité syntaxique. Deux études sont présentées. La première porte sur les données recueillies auprès de 675 élèves de la 6ème à la 3ème où il a été mis en évidence un effet des deux facteurs manipulés sur les performances en compréhension de phrases écrites. Dans la seconde étude, 196 enfants du CE1 au CM2 ont été évalués en compréhension de phrases à l'écrit, à l'oral, en lecture de mots, en vocabulaire, sur leurs capacités mnésiques et leurs connaissances syntaxiques. Deux analyses de régression ont été menées pour étudier le poids respectif des variables cognitives et langagières sur les performances en compréhension de phrases écrites. Chez les plus jeunes (CE1-CE2), on observe le poids important des compétences à l'oral expliquant les performances en compréhension de phrases écrites alors que chez les plus âgés (CM1-CM2) ce sont les performances en lecture de mots qui expliquent le mieux les variations individuelles en compréhension. Ces résultats sont discutés dans le cadre du modèle développemental des habiletés convergentes de la compréhension en lecture de Vellutino et coll..*

MOTS-CLÉS:

Compréhension - Lecture - Développement.

DE LA COMPRÉHENSION EN LECTURE CHEZ L'ENFANT DE 7 À 15 ANS: ÉTUDE D'UN NOUVEAU PARADIGME ET ANALYSE DES DÉTERMINANTS

par Jean Ecalte, Annie Magnan, Houria Bouchafa

SUMMARY: Reading comprehension in children from 7 to 15 y-o: assessment with a new design.

Reading comprehension depends in particular on comprehension of isolated written sentences. To examine this last ability, a new design named judgment of semantic similarity (children were asked to say if two sentences were semantically close) was proposed. Two experiments were carried out. In the first, with children from Grade 6 to Grade 9, it was shown that words frequency and complexity syntactic of sentences are related to performance in written sentences processing. In experiment 2, children from Grade 2 to Grade 5 were administered several tasks, the judgment of semantic similarity in visual modality (reading) and auditory modality (listening comprehension), in word recognition, in vocabulary, in short term memory and working memory, and in syntactic knowledge. Two hierarchical regression analyses were run in order to examine the weight of different cognitive and language variables in comprehension of written sentences. In children of Grades 2-3, listening comprehension was the best predictor to explain the variance of performance in comprehension of written sentences where as in children of Grades 4-5, word recognition was the best predictor. The results were discussed in the frame of Vellutino and al's convergent skills model of reading development.*

KEY-WORDS:

Reading comprehension - Word recognition - Reading development.

Jean ECALLE
Annie MAGNAN
Université Lyon2
Laboratoire Étude des
Mécanismes Cognitifs
EA 3082
5 avenue Mendès France
69676 BRON Cedex
annie.magnan@univ-lyon2.fr

Houria BOUCHAFA
Université d'Angers
Laboratoire Processus de Pensée
et Interventions
EA 2646
Université Catholique de l'Ouest
CERIPSA
Classe A Leroy
BP 10808
490008 ANGERS cedex 01
ouria.bouchafa@uco.fr

DE LA COMPRÉHENSION EN LECTURE CHEZ L'ENFANT DE 7 À 15 ANS : ÉTUDE D'UN NOUVEAU PARADIGME ET ANALYSE DES DÉTERMINANTS

Ce travail porte sur l'utilisation d'un nouveau paradigme d'étude de la compréhension de phrases en lecture qui consiste à juger de l'adéquation sémantique entre deux phrases. Notre objectif est d'étudier le poids respectif des déterminants cognitifs et langagiers sur les performances à cette épreuve.

LA COMPRÉHENSION EN LECTURE : ASPECTS GÉNÉRAUX

La compréhension en lecture met en jeu des processus locaux avec traitement des indices linguistiques (lexicaux, morphologiques et syntaxiques) au cours de la lecture des mots et une activité de plus haut niveau engagée dans la production d'une situation mentale construite au cours de la lecture du texte*. La compréhension est donc issue de processus de nature séquentielle (chaque mot traité successivement apporte de l'information plus ou moins importante) et parallèle avec l'activation de connaissances propres au lecteur en fonction notamment des buts qu'il s'est assignés*. Il faut dès lors distinguer plusieurs étapes dans l'accès à la compréhension d'un texte.

1. *L'accès au lexique mental* : l'identification des mots rapide et précise est essentielle à la compréhension. L'automatisation de ce processus allège la mémoire de travail pour allouer plus de ressources cognitives à la compréhension.
2. *L'analyse syntaxique* s'effectue pas à pas et la prise en compte des informations morphologiques est primordiale.
3. *L'intégration des propositions avec leur signification* repose sur la précision des informations syntaxico-sémantiques relevées et des capacités mnésiques et attentionnelles du lecteur.
4. *La combinaison des différentes propositions* est facilitée par la cohésion du texte (fournie par des éléments linguistiques tels que les connecteurs et anaphores, notamment) et la cohérence entre les propositions que le lecteur construit progressivement. Cette phase nécessite également la prise en compte des informations morphologiques, morpho-syntaxiques, thématiques et pragmatiques. La cohérence, d'une part, est tributaire de la capacité de la mémoire de travail et, d'autre part, s'appuie sur les inférences élaborées par le lecteur. Les inférences sont des ajouts d'informations qui ne sont pas fournies dans le texte. Enfin, cette phase de combinaison de propositions s'accompagne d'une sélection des informations nécessaires pour la phase suivante.
5. *La construction d'un modèle mental de la situation* émerge à partir de la situation décrite par le texte c'est-à-dire des informations hiérarchiquement retenues dans la phase précédente. Le lecteur utilise ses connaissances sur le thème et intègre des schémas ou scripts (ensemble de connaissances communes partagées pour une situation ; exemple « au restaurant ») facilitant la construction du modèle*.

Selon la formule désormais classique $L = R * C$ proposée par Gough et Tunmer*, lire est le produit de deux composantes, la reconnaissance de mots écrits et la compréhension. En d'autres termes, la compréhension écrite est issue de flux d'informations de type *bottom-up* : la mise en œuvre efficace de processus d'identification de mots (processus de bas niveau) est une condition nécessaire (mais non suffisante) à la mise en œuvre des processus de compréhension. Ces derniers, non spécifiques à l'activité lexicale, sont de nature *top-down* dans la mesure où ils engagent des connaissances

*Lecocq et coll., 1996 ; Blanc, Brouillet, 2003

*Fayol, 1992

*Denhière, Baudet, 1992

*1986

diverses de plus haut niveau (thématiques, morpho-syntaxiques, sémantiques) venant contribuer à l'émergence du modèle mental de situation lors de l'intégration successive des informations linguistiques relevées au cours de la lecture de mots*. Une présentation synoptique de ce modèle très général est proposée dans la figure 1.

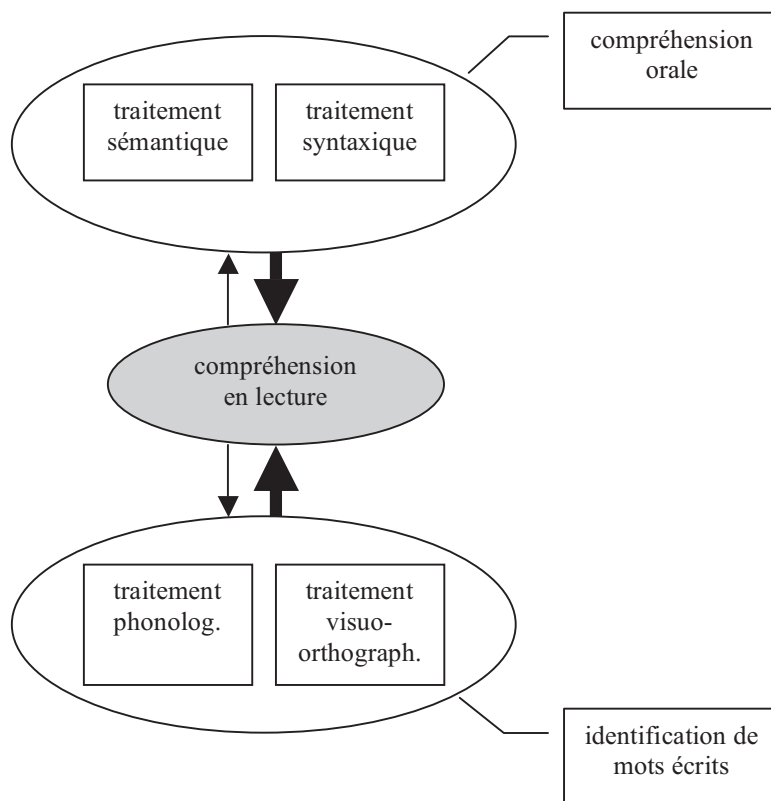


Figure 1 : Cadre général de la mise en œuvre des processus en compréhension de lecture

Comprendre un texte implique donc d'une part, de pouvoir lire tous les mots rapidement et précisément avec la mise en œuvre de processus de type phonologique (décodage) et visuo-orthographique (voie directe). D'autre part, la compréhension est soumise à l'activation de connaissances sémantiques et syntaxiques directement issues du traitement de chaque mot. Toutefois, il convient d'invoquer des flux d'informations « en retour » pour les deux liens compréhension orale-compréhension écrite et compréhension écrite-identification de mots écrits. En effet, la mise en œuvre de processus de compréhension lors de la lecture peut faciliter d'une part, l'activation de connaissances sémantiques et le traitement syntaxique et d'autre part, contribuer à l'identification des mots écrits*.

Lorsque l'identification de mots est déficitaire, la compréhension devient problématique pour au moins deux raisons :

- 1/ si les mots ne sont pas correctement identifiés, la compréhension fait place à une sorte de « devinement » approximatif et
- 2/ le surcoût cognitif engagé lors de la lecture de mots amenuise les ressources cognitives et attentionnelles pouvant être dédiées à la compréhension. Par ailleurs, la lecture de mots peut être efficace mais la compréhension du texte reste soumise à l'efficacité des processus de compréhension à l'oral. C'est le cas (relativement rare) des hyperlexiques qui lisent correctement les mots (« oralisation mécanique ») mais sans comprendre.

*Toutefois, dans cette condition, la part d'information apportée par le contexte est minime, cf. la synthèse sur ce point de Morais, 1999

LE DÉVELOPPEMENT DE LA COMPRÉHENSION EN LECTURE : QUELS DÉTERMINANTS ?

Le modèle componentiel et structural des habiletés convergentes de Vellutino, Tunmer, Jaccard, Chen* met en évidence dans une étude transversale menée auprès d'enfants de CE1-CE2 et de 6^{ème}-5^{ème} à nouveau les deux mêmes composantes expliquant les performances en compréhension de lecture et convoque en outre d'autres facteurs venant contribuer à ces processus.

Sur un plan développemental, lorsque l'enfant commence à apprendre à lire, ses capacités de compréhension (à l'oral) sont déjà efficaces (toutefois des variations interindividuelles existent). Il lui reste donc à développer des processus d'identification de mots écrits rapides et automatisés. Le modèle général (figure 1) et de façon plus large les travaux expérimentaux menés dans le domaine* rappellent les différents déterminants de la compréhension en lecture, à savoir pour l'essentiel, outre les deux composantes précédemment citées (identification de mots et compréhension orale), le vocabulaire, le traitement syntaxique, les capacités inférentielles et de raisonnement, les capacités mnésiques et la capacité à piloter/orienter la compréhension. Quelques travaux sont rappelés pour illustrer ces aspects.

Dans une première étude menée auprès d'une centaine d'enfants de 7 ; 6 ans, Cain, Oakhill et Bryant* montrent que les performances en compréhension de lecture sont expliquées par le QI verbal (15 % de variance expliquée), le vocabulaire (7 %) et une tâche de détection d'intrus (11 %). Pour les auteurs, la compréhension serait liée non pas à des processus d'identification de mots mais à des processus phonologiques impliqués massivement en identification. En outre, en suivant cette population sur 2 ans, les mêmes auteurs* montrent que la compréhension à l'écrit est expliquée par des mesures liées à des processus d'intégration de texte (*text integration*) avec capacité à faire des inférences, liées à une mesure rendant compte d'une compétence métacognitive de type pilotage de stratégie (*metacognitive monitoring*) et enfin par des capacités mnésiques. D'autres travaux* insistent sur le lien entre compréhension en lecture et mémoire de travail. Ce lien serait médiatisé par une habileté à contrôler des informations non pertinentes. En effet, les processus de mise à jour s'appuyant sur la mémoire de travail sont nécessaires à la construction du modèle mental de situation. On voit dès lors en quoi cette capacité de mise à jour joue un rôle via la mémoire de travail pour supprimer des informations non pertinentes ou stocker des informations importantes lors de la lecture d'un texte*.

Enfin, le poids spécifique des processus lexicaux a été examiné. L'identification de mots écrits joue évidemment un rôle dans les performances en compréhension en lecture*. Toutefois, la puissance du lien dépend de ce que l'on mesure en compréhension écrite et des épreuves utilisées en lecture de mots*. Le poids des connaissances lexicales (à l'oral) joue également un rôle en compréhension. Dans une étude avec des enfants de CE2, Tannenbaum, Torgesen et Wagner* investiguent les liens entre trois dimensions évaluant les connaissances lexicales et la compréhension écrite. Les auteurs distinguent l'étendue du vocabulaire (*breadth*) c'est-à-dire sa taille et la profondeur (*depth*) qui renvoient à la richesse c'est-à-dire l'ensemble des significations et des détails conceptuels qu'un mot peut comporter. Enfin, la troisième dimension concerne la fluence, autrement dit la vitesse d'accès au sens. A partir de scores obtenus dans 6 épreuves, ils mettent finalement en évidence deux facteurs, l'étendue du vocabulaire et un facteur regroupant profondeur et fluence. Ils montrent que ces deux composantes lexicales rendent compte de 62 % de la variance en compréhension en lecture. Sur le même registre, pour examiner le poids du vocabulaire, Ouellette* distingue son étendue déterminée par deux mesures, le vocabulaire réceptif (trouver le mot fourni parmi 4 images) et le vocabulaire expressif (nommer une image) et sa profondeur, mesures issues de deux tâches, donner la définition d'un mot et fournir un synonyme (parmi 4 mots). Après contrôle de l'âge et de l'intelligence non verbale (QIP), l'auteur observe que les performances en compréhens-

*2007 ; pour une synthèse en français, voir Ecalte, Magnan, Ramus, 2007

*cf. la synthèse remarquable de Oakhill et Cain, 2003

*2003

*Oakhill, Cain, Bryant, 2003

*Seigneuric, Ehrlich, 2005 ; Carretti et coll., 2006 ; Gathercole et coll., 2006

*Carretti et coll., 2006

*en français, voir Lecocq et coll., 1996
*Cutting, Scarborough, 2006

*2006

*2006

sion sont expliquées par la profondeur du vocabulaire. En conclusion, la lecture serait plus liée à la profondeur des connaissances sémantiques.

Quant au traitement syntaxique, on distinguera le poids du facteur « complexité syntaxique », les performances en compréhension de phrases étant liées au degré de complexité de la structure des propositions* et le poids des connaissances syntaxiques. Sur ce deuxième aspect, il est désormais classique de distinguer ce qui relève d'un savoir implicite (connaissances nécessaires pour extraire le sens de différentes constructions syntaxiques - ex : active/passive - ou pour exercer un jugement de grammaticalité) d'un savoir explicite encore appelé conscience syntaxique qui relève d'une activité réflexive et contrôlée sur la syntaxe (par ex. réplication d'erreurs)*.

D'autres facteurs contribuent à la compréhension en lecture :

- la capacité à repérer la structure d'un texte (hiérarchie des informations et logique interne) ;
- la capacité à faire des inférences : Oakhill et Cain* distinguent les inférences de cohérence (mises en œuvre à partir de connaissances générales sur le monde) et les inférences « en surplus » (*elaborative inference*), non nécessaires à la compréhension, qui ajoutent de l'information, « l'embellissent » ; il faut y rajouter un type particulier d'inférences, la compréhension des métaphores* ;
- la capacité à engager des stratégies pendant la lecture d'un texte : Cain* observe que les bons compreneurs comparativement à des lecteurs de même niveau lexique (même âge de lecture de mots) sont capables de stratégie adaptée : ils savent qu'il faut se souvenir des points importants.

Dans ce travail, pour approcher une partie importante des processus de compréhension en lecture, on examinera une capacité de base, la capacité de traitement syntaxique via un nouveau paradigme de jugement d'adéquation sémantique qui consiste à dire si deux phrases écrites veulent ou non dire la même chose*.

Dans une première étude menée auprès d'élèves de collège, on examinera l'effet de deux facteurs, la fréquence lexicale et la complexité syntaxique. Au cours d'une autre étude transversale menée auprès d'enfants scolarisés en primaire, en utilisant le même paradigme, l'objectif est de rechercher le poids respectif d'un certain nombre de déterminants (compréhension orale, identification de mots écrits, vocabulaire, connaissances syntaxiques implicites/explicites) sur les performances en compréhension de phrases écrites. Selon le modèle développemental de Vellutino et coll.*², la modification du poids de diverses habiletés est liée à l'âge/niveau de lecture des enfants. En particulier, chez les jeunes lecteurs, les liens entre compréhension et identification de mots écrits sont plus puissants que chez les lecteurs plus âgés pour lesquels la compréhension en lecture est plus liée à des processus de compréhension à l'oral, elle-même associée aux connaissances sémantiques et syntaxiques.

MÉTHODOLOGIE

ÉTUDE 1

Population

Cette étude transversale concerne 675 élèves de collège répartis comme suit : 185 scolarisés en 6^{ème} (âge moyen = 11 ; 7 ans, $\sigma = 4,6$ mois), 174 en 5^{ème} (am = 12 ; 7 ans, $\sigma = 3,9$ m), 139 en 4^{ème} (am = 13 ; 7 ans, $\sigma = 7$) et 177 en 3^{ème} (am = 14 ; 7, $\sigma = 7$). Tout élève ayant un an ou plus de retard scolaire a été écarté de l'étude afin d'obtenir un échantillon plus homogène reflétant un niveau scolaire « normal ». En effet, une étude antérieure* avait montré que le retard scolaire était lié à des retards en lecture. Les collèges sont répartis dans 9 départements dont un DOM (Département d'Outre-Mer). Ils ne relèvent pas de ZEP (Zone d'Education Prioritaire).

*Lecocq et coll., 1996

*Demont, Gombert, 1996 ; Gaux, Gombert, 1999

*2003

*cf. le remarquable travail de Lecocq et coll., 1996
*1999

*Ce paradigme se distingue de celui utilisé dans le TCS de Maeder (Maeder, Gilles, 2006) dans la mesure où il est proposé une tâche d'appariement sémantique entre 2 phrases à choix forcé (trouver parmi 4, la phrase qui a le même sens que la phrase cible)

*2007

*Ecalte, Magnan, 2006

Matériel et Procédure

Une tâche de jugement d'adéquation sémantique a été proposée : elle consiste à dire si deux phrases linguistiquement différentes, présentées successivement signifient presque la même chose versus pas du tout la même chose. Les enfants devaient simplement entourer deux points noirs s'ils pensaient que les 2 phrases se ressemblaient fortement sur le plan sémantique ou entourer un point noir et un point blanc pour le cas contraire (figure 2). L'épreuve comporte 48 items (composés de 2 phrases), 24 avec des paires de phrases sémantiquement proches, 24 avec des phrases sémantiquement éloignées.

Le garçon bondit.	L'enfant saute.	<input type="radio"/> <input checked="" type="radio"/>	<input checked="" type="radio"/> <input checked="" type="radio"/>
Le losange qui est dans l'ovale est beige.	Le losange est dans l'ovale beige.	<input type="radio"/> <input checked="" type="radio"/>	<input checked="" type="radio"/> <input checked="" type="radio"/>

Figure 2 : Exemples d'items dans le bloc A (à fréquence élevée ; réponse oui) et dans le bloc C (fréquence faible ; réponse non)

Deux facteurs ont été manipulés, la fréquence lexicale et la complexité syntaxique des phrases. Pour le premier, l'indicateur UG1G5 de Manulex* a été utilisé, norme qui correspond à la fréquence des mots du CP au CM2. Les paires de phrases comportent soit des mots* de fréquence lexicale élevée (F+) dont l'indicateur U est supérieur à 10, soit de fréquence faible (F-, U < 10). La complexité syntaxique a été contrôlée en s'inspirant des travaux de Lecocq et coll.* dans la construction de l'ECOSSE (épreuve de compréhension de phrases avec support imagé). Douze structures différentes de complexité syntaxique croissante ont été retenues et regroupées en 4 blocs.

En résumé, le matériel, présenté sur un format A3 recto verso, est composé de 48 items répartis en 4 blocs (feuilles A, B, C, D) dont la complexité syntaxique croît. Dans chaque bloc, 12 paires de phrases sont proposées réparties sur 3 structures syntaxiques et 2 niveaux de fréquence lexicale. Dans chaque condition, une paire de phrases a pour réponse « oui » (les 2 phrases de la paire veulent dire presque la même chose : paire « semblable ») et une autre paire a pour réponse « non » (les 2 phrases ne veulent pas du tout dire la même chose : paire « différente »). Enfin, sur chaque feuille, les items sont présentés de façon aléatoire.

La passation s'est déroulée par groupes de 10-12 élèves dans un endroit calme où les enfants étaient disposés pour éviter tout copiage. Deux essais avec feedback correctif étaient proposés en début d'épreuve. Puis, les enfants répondaient à l'épreuve à leur rythme. La durée de l'épreuve était limitée à 30 min.

La VD (Variable Dépendante) est un score pondéré tenant compte des réponses correctes oui/oui (réponse attendue oui/réponse donnée oui) moins les erreurs type « fausses alarmes » non/oui (réponse attendue non/réponse donnée oui). Le score maximum est donc de 24.

Résultats

Une analyse de variance a été réalisée sur les scores pondérés avec comme facteurs, la classe C (6^{ème}, 5^{ème}, 4^{ème}, 3^{ème}), la fréquence F (F+ vs F-) et la complexité syntaxique B (blocs A, B, C, D). On observe un effet significatif de la classe, $F(3, 671) = 30.66$, $p < .0001$, les scores progressant de la 6^{ème} à la 3^{ème} (13.5, 14.5, 16.5, 17.1). Les tests post hoc (Tukey) indiquent une différence tendancielle significative (.08) entre les performances en 6^{ème} et celles en 5^{ème}, une différence significative (.0001) entre celles de 5^{ème} et celles de 4^{ème} et pas de différence significative entre celles de 4^{ème} et celles de 3^{ème}. L'effet attendu du facteur Bloc est également significatif, $F(3, 2013) = 305.756$, $p < .0001$. Les performances diminuent progressivement du bloc A au bloc D (4.6, 4.2, 4, 2.6), les tests post hoc montrant des différences significatives (.0001) entre chaque

*Lété, Sprenger-Charolles, Colé, 2004

*Il n'a pas été tenu compte de la fréquence lexicale des petits mots fonctionnels (déterminant, anaphore, connecteur)
*1996

bloc. Enfin, on observe un effet significatif de la fréquence, $F(1, 671) = 297.65$, $p < .0001$, les items avec des mots de fréquence élevée ayant des scores supérieurs aux items avec des mots de fréquence faible (8.4 vs 6.9).

L'interaction Classe*Bloc, significative, $F(9, 2013) = 5.99$, $p < .0001$, est exprimée sur la figure 3a. On observe que les différences de performances entre les blocs diminuent lorsque le niveau scolaire augmente. L'interaction Classe*Fréquence (figure 3b) est significative, $F(3, 671) = 4.37$, $p < .005$, montrant que l'effet de la fréquence est moins important lorsque le niveau scolaire s'élève. L'interaction Bloc*Fréquence (figure 3c) est significative, $F(3, 2013) = 70.79$, $p < .0001$; elle est due aux différences de performances entre les items à fréquence lexicale élevée et celles d'items à fréquence faible, notamment avec une différence quasiment nulle sur le bloc B et une différence très faible sur le bloc D. Enfin, la triple interaction C*B*F significative, $F(9, 2013) = 2.04$, $p = .03$, est exprimée sur la figure 4. Deux phénomènes sont observés, l'un révèle pour les 4 classes une diminution remarquable des performances sur les blocs B, C, D à fréquence lexicale faible, l'autre montre l'effet amoindri de la complexité syntaxique sur les 4 blocs à fréquence lexicale faible lorsque le niveau scolaire augmente.

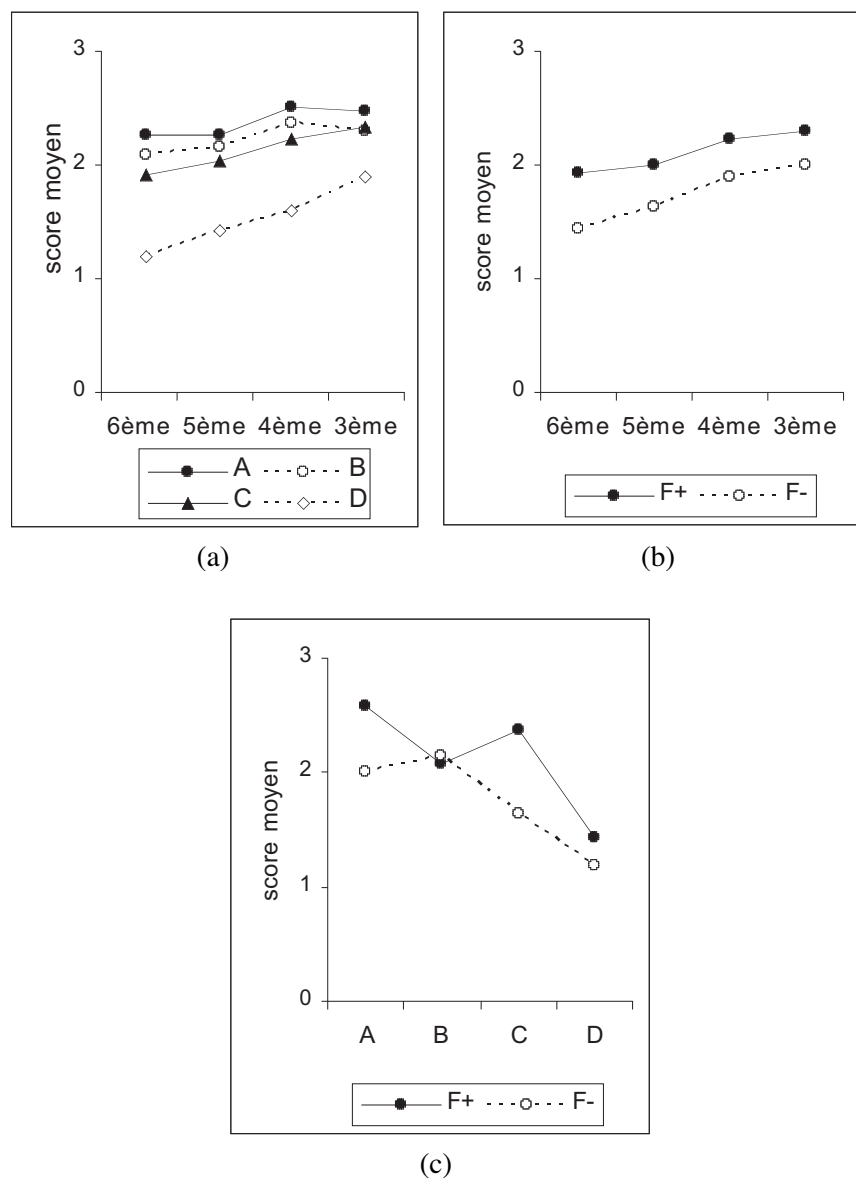


Figure 3: Graphiques d'interaction Classe*Bloc (a), Classe*Fréquence (b), Bloc*Fréquence (c)

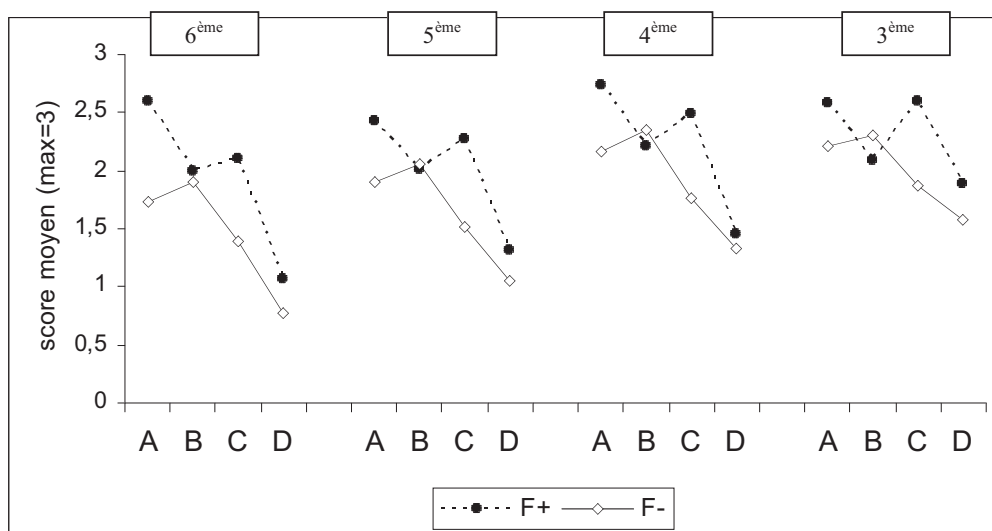


Figure 4: Graphique de l'interaction Classe*Bloc*Fréquence

ÉTUDE 2

Population

Cette nouvelle étude* transversale a porté sur 196 enfants répartis sur 4 niveaux scolaires, CE1 ($N = 54$; âge moyen : 7 ; 4 ans), CE2 ($N = 45$; am : 8 ; 5 ans), CM1 ($N = 53$; am : 9 ; 5 ans), CM2 ($N = 44$; am : 10 ; 5 ans). A nouveau, seules les données issues des enfants « à l'heure » (non redoublants) ont été prises en compte. La population a été testée dans 3 départements dans des écoles hors ZEP.

Dispositif

Plusieurs tâches ont été proposées, en compréhension orale, en compréhension écrite, portant sur les connaissances syntaxiques, le vocabulaire, la lecture de mots et les capacités mnésiques (mémoire à court terme et mémoire de travail).

Compréhension. La même épreuve que dans l'étude 1 a été administrée, d'abord en modalité auditive, évaluant la compréhension de phrases à l'oral puis en modalité visuelle pour évaluer la compréhension de phrases écrites. La passation de la première s'est déroulée en passation individuelle alors que la seconde a été administrée en petits groupes.

Connaissances syntaxiques. Une première épreuve de type épi-linguistique renvoie à des connaissances implicites avec une tâche de jugement de grammaticalité de phrases. Elle consistait à relever les éventuelles erreurs morpho-syntaxiques dans une phrase énoncée oralement. L'enfant devait dire si oui ou non la phrase était correcte (*les poussins dorment dans l'écurie ; le monsieur lit le grand journaux*). Après trois exemples avec feedback correctif, 20 items étaient proposés, 10 avec réponse « oui » (phrases correctes) et 10 réponses « non » (phrases incorrectes). Score maximum : 20.

La seconde épreuve de type méta-linguistique implique la mobilisation de connaissances explicites dans une tâche de répliation d'erreurs. L'enfant devait transformer une phrase correcte en phrase grammaticalement non correcte à partir d'un modèle. Par exemple, la phrase modèle était « *le jardinier un trou creuse* » et la phrase cible est « *maman lit une histoire* », la réponse attendue étant « *maman une histoire lit* ». Les erreurs sont de type syntaxique sur la structure basique SVO (sujet-verbe-objet) ou morpho-syntaxique sur les accords en genre (*un grosse ventre*). Score maximum : 10.

Vocabulaire. L'EVIP* a été proposée. C'est une épreuve de vocabulaire réceptif où l'enfant doit désigner l'image parmi 4 qui correspond au mot oral prononcé par l'expérimentateur. La VD est le score brut.

*Le recueil des données a été réalisé par Cécile Arlabosse et Christelle Charmond, étudiantes à l'Institut des Sciences et Techniques de la Réadaptation - Formation d'Orthophonie - Université Lyon (1)

*Echelle de Vocabulaire en Images Peabody : Dunn, Thériault-Wahlen, Dunn, 1993

Lecture de mots. Le subtest LUM* a permis d'obtenir un indice alliant vitesse et précision, deux aspects des processus d'identification de mots définissant leur degré d'expertise. Il a été demandé à l'enfant de lire à voix haute une série de mots présentés sur une feuille en une minute. La VD tient compte du nombre de mots lus moins les erreurs d'oralisation.

Mémoire. Le subtest de l'ODéDys* a été proposé : il comporte une partie « rappel de chiffres à l'endroit » et une partie « rappel de chiffres à l'envers ». Les capacités mnésiques (mémoire à court terme et mémoire de travail) sont globalement évaluées avec le score obtenu dans les deux domaines.

Résultats

Tableau 1 : Données descriptives

	Comp. écrit /24	Comp. oral /24	Jug. gramm. /20	Répl. erreurs /10	EVIP	LUM	Mémoire
CE1	3.2 (4.6)	4.4 (4.2)	16.9 (2.2)	1.1 (1.3)	82.5 (11.1)	37.9 (10.1)	7.6 (1.4)
CE2	6.2 (5.3)	9.2 (4.9)	18.2 (1.9)	3.5 (2.6)	96.4 (14.6)	52.8 (11)	7.8 (1.2)
CM1	8 (4.4)	10.9 (3.9)	18.9 (1.5)	5 (2.7)	117.5 (17.4)	66.8 (18.3)	9 (1.1)
CM2	10.6 (4.7)	13.2 (4.8)	19 (1.3)	6.4 (2.5)	122.3 (15.6)	76.5 (15.1)	9.5 (1.2)

Les données descriptives sont présentées dans le Tableau 1. Globalement, on observe pour chaque variable un effet significatif ($p < .001$) du niveau scolaire sur les performances, celles-ci augmentant progressivement du CE1 au CM2. On observera que les scores dans la tâche de jugement de grammaticalité sont saturés, le taux de réussite allant de 84 % en CE1 (16,9/20) à 95 % en CM1 (18,9/20).

Pour examiner le poids respectif des déterminants en compréhension écrite de phrases, une analyse de régression hiérarchique pas à pas a été effectuée avec comme VD les scores en compréhension écrite (variable à expliquer) et les scores dans les autres tâches comme VI (variables candidates à l'explication de la VD), à savoir les scores en compréhension de phrases à l'oral (Comp. oral), en lecture de mots (LUM), à la tâche de jugement de grammaticalité (Jug. gramm.), à la tâche de répl. d'erreurs (Répl. erreurs), en vocabulaire (EVIP) et en Mémoire. Pour contrôler l'effet éventuel de l'âge chronologique, cette variable est introduite systématiquement en premier. Enfin, pour renforcer les analyses, on a réalisé les régressions sur les effectifs cumulés en CE (CE1 + CE2) et en CM (CM1 + CM2).

Tableau 2 : Résultats des analyses de régression

	β	r^2	Δr^2	p
<i>CE1-CE2: N = 99</i>				
1- âge	.30	.09	-	.002
2- Comp. oral	.54	.32	.23	.0001
3- Répl. erreurs	.36	.40	.08	.001
4- LUM	.21	.43	.03	.036
<i>CM1-CM2: N = 96</i>				
1- âge	.14	.02	-	.18
2- LUM	.45	.22	.20	.0001
3- Comp. oral	.39	.34	.12	.0001
4- EVIP	.20	.37	.03	.04

Au cours élémentaire (Tableau 2), l'âge explique 9 % de la variance en compréhension écrite, les scores en compréhension à l'oral 23 %, ceux obtenus dans la tâche de répl. d'erreurs 8 %, enfin les scores au LUM 3 %. Au total, 43 % de la variance

en compréhension de phrases à l'écrit sont expliqués par ces 4 variables. Au cours moyen, après la variable âge qui apporte seulement 2 % de variance (non significatif), ce sont d'abord les scores au LUM qui expliquent 20 % de la variance en compréhension écrite, puis ceux en compréhension à l'oral (12 %), enfin le vocabulaire, 3 %. Globalement, ces 4 variables expliquent 37 % de la variance en compréhension de phrases à l'écrit.

CONCLUSION ET DISCUSSION

L'objectif de ce travail consistait à mettre en œuvre un nouveau paradigme d'étude de la compréhension en lecture qui utilise une tâche de jugement d'adéquation sémantique portant sur des phrases dont on a manipulé la fréquence lexicale et la complexité syntaxique. La première étude s'est adressée à une population relativement importante d'élèves de collège, au cours de laquelle les effets de fréquence lexicale et de complexité syntaxique ont été analysés. La seconde étude avait pour but d'examiner auprès d'une population du primaire le poids différentiel de certaines variables cognitives et langagières pouvant affecter le développement des processus de traitement syntaxique. Globalement du CE1 à la 3^{ème}, on a pu observer que les performances augmentent progressivement montrant ainsi que l'épreuve est sensible au niveau scolaire.

Au collège, le taux moyen de performance s'élève à 56.3 % en 6^{ème}, 60.4 % en 5^{ème}, 68.8 % en 4^{ème} et 71.3 % en 3^{ème}. Ce dernier taux montre qu'on est loin d'un effet plafond et suggère la difficulté de l'épreuve. En outre, on note l'effet attendu des facteurs fréquence lexicale et complexité syntaxique. Lorsque les phrases comportent des mots fréquents, les performances sont supérieures à celles des phrases avec des mots de fréquence faible. Par ailleurs, les performances sont dépendantes de la complexité des phrases : globalement, les performances baissent progressivement au fur et à mesure que la complexité augmente. On observe aussi que le bloc D, le plus difficile à traiter, produit des scores bien inférieurs aux blocs A, B, C dans les 4 niveaux scolaires. Une première interaction Classe*Bloc met en évidence que l'effet de complexité est moindre lorsque le niveau augmente, les différences entre les blocs s'amenuisant. L'effet de la fréquence lexicale est moins marqué quand le niveau scolaire progresse, les différences de scores entre les items à fréquence faible et ceux de fréquence élevée diminuent (interaction Classe*Fréquence). Tous niveaux scolaires confondus, la troisième interaction Bloc*Fréquence met en évidence deux faits :

- 1/ les performances sur les items du bloc B ne se distinguent guère en fonction de la fréquence lexicale et
- 2/ la différence de performances entre les items F + et F- sur le bloc D est plus faible que celle sur les blocs A et C.

Enfin, la triple interaction Classe*Bloc*Fréquence montre d'une part, que l'effet de complexité syntaxique est clairement marqué sur les blocs B, C, et D pour les items à fréquence lexicale faible et d'autre part, que sur les items à fréquence lexicale élevée, la différence de performances entre les blocs A, B, et C tend à s'estomper lorsque le niveau scolaire s'élève.

En primaire, pour l'épreuve de compréhension de phrases dans les deux modalités auditive et visuelle (lecture), les données descriptives rappellent ce que Lecocq avait déjà montré, à savoir qu'il y a un décalage de performances entre celles obtenues dans la modalité auditive, supérieures à celles obtenues dans la modalité visuelle. En outre, il a été observé un effet plafond dans l'épreuve de jugement de grammaticalité et de faibles performances dans la tâche de réplcation d'erreurs pour les enfants des classes de CE1 (10 %) et CE2 (30 %). Les résultats de l'analyse de régression montrent que chez les plus jeunes lecteurs (CE1 et CE2), les performances en compréhension de phrases écrites sont expliquées massivement par celles en compréhension orale suivies par les performances obtenues dans la tâche de réplcation d'erreurs, les scores en lecture de mots apportant un pourcentage d'explication de variance très faible. Au contrai-

re, avec les lecteurs plus âgés (CM1 et CM2), c'est l'inverse qui apparaît : les scores en compréhension de phrases écrites sont expliqués massivement par la lecture de mots, suivie de la compréhension orale puis du vocabulaire. Autrement dit l'hypothèse développementale avancée, issue des travaux de Vellutino et coll.* ne se trouve pas confirmée par notre étude.

*2007

En effet, selon les auteurs, la compréhension chez les jeunes enfants serait plus liée aux processus d'identification de mots écrits alors que chez les plus âgés la compréhension en lecture serait plus dépendante des processus de compréhension à l'oral. Si la position de Vellutino et coll.* est parfaitement plausible, par ailleurs renforcée par leurs résultats, l'hypothèse développementale des liens entre compréhension en lecture, lecture de mots et compréhension orale pourrait évoluer non seulement avec l'âge mais surtout, de notre point de vue, pourrait dépendre du matériel utilisé. En effet, nous avons testé seulement la capacité de traitement syntaxique qui n'est qu'un aspect de la compréhension en lecture alors que Vellutino et coll.* évaluent la compréhension via la lecture silencieuse d'un texte narratif suivi de questions. Par ailleurs, il est étonnant que dans notre étude le niveau de vocabulaire n'apporte qu'une contribution réduite à l'explication de la variance en compréhension écrite et les capacités mnésiques aucune explication. Sur ce dernier point, l'épreuve combinée utilisant l'empan endroit/envers ne semble pas la mieux adaptée pour saisir le poids des processus mnésiques impliqués en compréhension de lecture. Une tâche de mise à jour évaluant les capacités de la mémoire de travail serait plus pertinente*.

*2007

*2007

*Carretti et coll., 2005

Les liens entre compréhension et différentes variables cognitives et langagières rapportés ici dans l'étude 2 et dans divers travaux n'établissent en rien une relation de causalité. Des études avec entraînement pourraient étayer le sens d'une telle relation. Sur cet aspect, Oakhill et Cain* présentent une recherche** au cours de laquelle des faibles lecteurs ont été entraînés à lire plus vite des mots. Alors qu'après entraînement, leurs performances en lecture de mots se sont améliorées, celles en compréhension n'ont pas progressé. De même, à partir du lien fort entre vocabulaire et compréhension, on pourrait s'attendre à ce que des enfants entraînés à développer leur niveau de vocabulaire voient leur niveau de compréhension en lecture s'améliorer. Même si une synthèse menée à partir de 8 études* montre des effets différentiels de l'entraînement sur la compréhension, les auteurs concluent en rapportant d'autres études où l'absence d'un effet sur la compréhension ne met pas en évidence le lien de causalité attendu. Pour eux, compréhension et vocabulaire pourraient être reliés indirectement par un troisième facteur, le QI verbal. A l'évidence, la compréhension en lecture est sous-tendue par diverses capacités cognitives et langagières : elle nécessite un certain niveau de vitesse/précision en lecture de mots, un certain niveau de vocabulaire, des connaissances syntaxiques suffisamment développées, des capacités de raisonnement et des capacités métacognitives pour piloter ses propres processus de recherche d'informations. C'est la confluence de ces capacités qui rend les processus de compréhension efficaces. A ce titre, le modèle présenté par Vellutino et coll.* est particulièrement évocateur et stimulant pour mener des recherches complémentaires.

*2003

**Fleisher et coll., 1979, cité

*Mezynski, 1983, citée par Oakhill, Cain, 2003

*2007

L'épreuve de jugement d'adéquation sémantique utilisée dans nos deux études est une bonne alternative aux épreuves de compréhension disponibles et largement utilisées chez les professionnels, l'ECOSSE* et l'épreuve Compréhension dans le LMC (Lecture de Mots et Compréhension*) (pour une analyse critique des tests de compréhension en lecture*). Ces deux épreuves sont fondées sur le même principe, trouver l'image (parmi 4) correspondant à la phrase écrite proposée. Leur composition constitue leur limite, à savoir un contexte imagé qui peut servir d'appui pour compenser une lecture de mots approximative. Si l'ECOSSE est particulièrement bien construite avec une complexité syntaxique croissante au cours des blocs, aucun des deux tests n'a contrôlé le niveau de vocabulaire dont on a vu l'importance en compréhension*.

*Lecocq, 1996

*Khomsi, 1999

*Voir Ecalte et coll., 2007

*Tannenbaum et coll., 2006 ; Ouellette, 2006

Dans les deux études présentées, on a observé une augmentation progressive des performances au cours des 8 niveaux scolaires (du CE1 à la 3^{ème}) ce qui renforce l'idée de développer après expérimentation sur une population plus nombreuse un dispositif

d'évaluation de la compréhension de phrases dont le matériel linguistique a été dûment contrôlé (fréquence lexicale et structure syntaxique). C'est dans ces conditions (un étalonnage important et une sélection rigoureuse) d'items qu'un nouveau test de compréhension de phrases pourrait être mis à disposition des professionnels.

Il reste toutefois que ce dispositif n'évaluera qu'un aspect de la compréhension, les capacités de traitement syntaxique qui constituent une des bases essentielles de la compréhension de texte. D'autres travaux expérimentaux sont nécessaires pour étudier cette hypothèse.

BIBLIOGRAPHIE

- BLANC, N., BROUILLET, P. (2003). *Mémoire et compréhension. Lire pour comprendre*. Paris : Editions InPress, collection Psychologie.
- CAIN, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology*, 17, 295-312.
- CAIN, K., OAKHILL, J.V., BRYANT, P.E. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal*, 13, 31-56.
- CARRETTI, B., CORNOLDI, C., DE BENI, R., ROMANO, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, 91, 45-66.
- CUTTING, L.E., SCARBOROUGH, H.S. (2006). Prediction of reading comprehension: relative language proficiency and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10 (3), 277-299.
- DEMONT, E., GOMBERT, J.E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology*, 66 (3), 315-332.
- DENHIÈRE, G., BAUDET, S. (1992). *Lecture, compréhension de texte et science cognitive*. Paris : PUF.
- DUNN, L.M., THÉRIAULT-WHALEN, C.M., DUNN, L.M. (1993). *Echelle de Vocabulaire en Images Peabody (EVIP), adaptation française du Peabody Picture Vocabulary Test-Revised*. Toronto : PsyCan, 147 p.
- ECALLE, J. (1997). Les représentations chez le lecteur novice ou expert : perspective componentielle. *Enfance*, 2, 285-303.
- ECALLE, J., MAGNAN, A. (2002). *L'apprentissage de la lecture. Fonctionnement et développement cognitifs*. Paris : A. Colin.
- ECALLE, J., MAGNAN, A. (2006). Des difficultés en lecture à la dyslexie : problèmes d'évaluation et de diagnostic. *Glossa*, 97, 4-19.
- ECALLE, J., MAGNAN, A., RAMUS, F. (2007). L'apprentissage de la lecture et ses troubles. In IONESCU, S., BLANCHET, A. (dir.), *Nouveau cours de psychologie. Psychologie du développement et de l'éducation* (volume coordonné par LAUTREY, J.), 229-275. Paris : PUF.
- FAYOL, M. (1992). Comprendre ce qu'on lit : de l'automatisme au contrôle. In FAYOL, M., GOMBERT, J.E., LECOCQ, P., SPRENGER-CHAROLLES, L., ZAGAR, D. (Eds.), *Psychologie cognitive de la lecture*, 73-106. Paris : PUF.
- FLEISCHER, L.S., JENKINS, J.R., PANY, D. (1979). Effects on poor readers' comprehension of training in rapid decoding. *Reading Research Quarterly*, 14, 30-48.
- GATHERCOLE, S.E., ALLOWAY, T.P., WILLIS, C., ADAMS, A.M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, 93, 265-281.
- GAUX, C., GOMBERT, J.E. (1999). La conscience syntaxique chez les pré-adolescents : question de méthodes. *L'Année Psychologique*, 99, 45-74.
- GOUGH, P.B., TUNMER, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7 (1), 6-10.
- JACQUIER-ROUX, M., VALDOIS, S., ZORMAN, M. (2005). *ODEDYS : Outil de DEpistage des DYSlexies*. Grenoble : IUFM, Laboratoire Cogni-Sciences.
- KHOMSI, A. (1999). *Epreuve d'évaluation de la compétence en lecture. LMC-R : lecture de mots et compréhension*. Paris : ECPA
- KHOMSI, A. (2003). *BLI : Batterie de lecture informatisée*. Paris : ECPA.
- LECOCQ, P. (1996). *L'E. CO.S.SE : une épreuve de compréhension syntactico-sémantique*. Lille : Presses Universitaires du Septentrion.
- LECOCQ, P., CASALIS, S., LEUWERS, C., WATTEAU, N. (1996). *Apprentissage de la lecture et compréhension d'énoncés*. Lille : Presses Universitaires du Septentrion.
- LÉTÉ, B., SPRENGER-CHAROLLES, L., COLÉ, P. (2004). Manulex : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, and Computers*, 36 (1), 156-166.
- MAEDER, C., GILLES, P.Y. (2006). Validation d'un test de compréhension syntaxique : TCS. *Glossa*, 98, 22-37.

- MEZYNSKI, K. (1983). *Review of Educational Research*, 53(2), 253-279.
- MORAIS, J. (1994). *L'art de lire*. Paris: Editions Odile Jacob.
- MORAIS, J. (1999). *L'art de lire*. Paris: Edition de Poche, collection Opus.
- OAKHILL, J.V., CAIN, K. (2004). The development of comprehension skills. In NUNES, T., BRYANT, P.E., *Handbook of Children's Literacy*, 155-180. Dordrecht: Kluwer Academic Publishers.
- OAKHILL, J.V., CAIN, K., BRYANT, P.E. (2003). The dissociation of word reading and text comprehension: evidence from component skills. *Language and Cognitive Processes*, 18 (4), 443-468.
- OUELLETTE, G.P. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98 (3), 554-566.
- SEIGNEURIC, A., ERHLICH, M.F. (2005). Contribution of working memory capacity to children's reading comprehension: a longitudinal investigation. *Reading and Writing*, 18, 617-656.
- TANNENBAUM, K.R., TORGESEN, J.K., WAGNER, R.K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10 (4), 381-398.
- VELLUTINO, F.R., TUNMER, W.E., JACCARD, J.J., CHEN, R. (2007). Components of reading ability: multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11 (1), 3-32.