UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RECOURS À LA SÉMANTIQUE LEXICALE DANS LA COMPENSATION EN LECTURE CHEZ DES ÉLÈVES DU PRIMAIRE QUI PRÉSENTENT UN DÉFICIT D'IDENTIFICATION DES MOTS ÉCRITS

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DOCTORAT EN ÉDUCATION

PAR

MÉLANIE BÉDARD

SEPTEMBRE 2023

REMERCIEMENTS

Une thèse sans remerciement, c'est comme une pédagogie sans flexibilité : une chimère. Voici donc un aperçu de toute la gratitude que je ressens. Un proverbe africain dit que « ça prend tout un village pour élever un enfant », moi, je dis que ça prend un nombre incalculable de personnes pour « élever » un doctorant!

Mes premiers remerciements sont sans contredit dédiés à Line Laplante. Si je prenais le temps de déployer complétement les raisons pour lesquelles j'ai envie de la remercier, je doublerais sans doute la longueur de cette thèse! C'est cette professeure qui a accueilli mon désarroi de recommencer un 2e baccalauréat en 2006 en éducation, celle qui m'a fait transcrire des enregistrements d'élèves interminables en alphabet phonétique, qui m'a proposé de faire du découpage et du collage de matériel des soirs et des fins de semaine complètes, qui m'a transmis sa passion ardente pour la profession d'orthopédagogue, qui a été suffisamment généreuse pour me trimballer de conférences en conférences pour me faire rencontrer des gens géniaux et dévoués et me montrer les coulisses du métier. Je ne peux plus calculer le nombre d'heures passées avec toi à travailler dur sur des projets, à rire, à voyager (… parfois seule), à pleurer, ou encore à rêver de vendre un certain type de vêtements dans un magasin à grande surface. Du fond du cœur, je te remercie d'être ma « maman professionnelle », et j'espère vraiment être à la hauteur.

Je tiens également à remercier chaleureusement Séverine Casalis, qui a accepté de me soutenir dans ce projet, de longue haleine, outre-Atlantique. Merci de ta grande disponibilité malgré le décalage horaire, et de ton immense générosité dans le partage d'idées et d'expertise. Tu as su être toujours présente pour moi, malgré ton emploi du temps chargé et la distance, même une certaine veille du jour de l'an à regarder des données. Tes visites à Montréal ont toujours été un plaisir, remplies de travail intense et de moments tout aussi plaisants. J'ai appris énormément de notre collaboration, tant d'un point de vue professionnel que personnel, et je souhaite de tout cœur que cela puisse se poursuivre encore.

La vie de doctorante est remplie de sacrifices, qui touchent notamment nos familles. Merci à mes enfants pour leur patience et leur compréhension envers moi dans certains moments plus difficiles de conciliation travail-études-famille. Vous êtes des enfants extraordinaires que tous les parents voudraient avoir. Vous ne le savez pas, mais c'est souvent vous qui m'avez donné la force de continuer. Merci à mes parents, qui ont toujours été présents pour leurs enfants. Le soutien que vous m'avez offert est indescriptible. Vous

m'avez montré à être persévérante et travaillante dans la vie. Vous m'avez élevée en m'encourageant à me dépasser. J'ai eu besoin de tout ça et encore plus pour réaliser cette thèse. Sans vous, je n'aurais jamais réussi. Merci aussi au père de mes enfants, Marc, pour sa patience, sa compréhension et son soutien. Merci à mon amoureux. Ton soutien inconditionnel et ton enthousiasme à l'égard de ma capacité à réaliser cet exploit (même quand moi-même j'en doutais) ont été vraiment précieux. La fierté que je vois dans tes yeux à mon égard est un message fort pour moi.

Je souhaite remercier des étudiants que j'ai eu le bonheur de croiser dans mon parcours universitaire. Merci, gang : Isabel, Kathleen, Ariane, Ivan, Matthew, Guillaume et Dominique (mes collègues étudiants du NeuroLab), Caroline Girard, Marie-Pier Godin, Marie-France Côté, Geneviève Monette, Dominique Laguë, Émilie Cloutier (et toutes celles que j'oublie). Ce fut toujours agréable de partager nos bonheurs et malheurs d'étudiants, d'avoir votre oreille attentive ou des conseils de votre part. Merci à Marie-Josée Barbeau qui en plus d'être une collègue étudiante est devenue une super amie. Merci de m'écouter parler de mes affaires et de partager mon enthousiasme pour la recherche et l'éducation; et aussi de réinventer le monde dans nos salles à manger. Merci aussi au technicien du NeuroLab, Anthony. Merci de ta patience avec moi dans l'apprentissage des techniques liées à l'EEG.

Je remercie également des professeurs.es que j'ai côtoyés de près ou de loin, mais qui ont fait une différence dans mon parcours : Monique Brodeur, Nathalie Chapleau, Catherine Gosselin, Anila Fejzo, Isabelle Gauvin, Simon Colin, Marie-Christine Beaudry, Danielle Bertrand-Poirier, France Dubé, Chantal Ouellet, Alain Desrochers, Tassos Mikropoulos, Véronique Fortier, etc. Chacun à votre façon, vous avez contribué à m'encourager à poursuivre, à alimenter la passion en moi, merci.

Il ne faut pas se cacher que le soutien financier est également déterminant dans la réalisation d'une thèse. Merci à toutes ces sources officielles de financement de mes études doctorales : Fondation de l'UQAM (Bourse Robert Sheitoyan), SCCUQ, FRQSC, CSLP-UQAM, ISC et Fondation J.A.-de-Sève.

Un merci spécial à l'Institut des troubles d'apprentissage (ITA), aux personnes que j'y ai rencontrées, et aux gens qui travaillaient avec moi lors de mon passage dans cet organisme : Annie, Jérémie, Benjamin, Danna, Christine, Lucille, Hélène, Lise, Marie-Neige, Marie-Julie, Madeleine et Fauve. J'ai toujours adoré la mission de cet organisme et j'ai adoré l'énergie effervescente que nous avions pour la cause. Merci d'avoir été sur ma route et d'avoir contribué à ma formation.

Un immense merci à l'Association des orthopédagogues du Québec (ADOQ), aux gens que j'y ai côtoyés ainsi qu'aux personnes qui soutiennent cette association depuis ses débuts. Sylvie Guertin et Jacynthe Turgeon, votre soutien indéfectible à mon endroit me touche toujours très profondément. Je vous apprécie beaucoup. Christine Pruneau, Nathalie Bisaillon, Marthe; que de souvenirs, dans mes premières armes à m'impliquer bénévolement dans une association professionnelle, merci d'avoir été avec moi. Mathieu, Isabelle, Magalie, Lise, etc., continuez, on a besoin d'une association forte.

Un merci particulier à Pierre, Pascale, David, et même FL, d'avoir accepté de partager un peu de votre vie avec moi. Sans vous et vos sacrifices, ma carrière de chargée de cours et de doctorante n'existerait pas. Je sais à quel point vous êtes importants dans cette équation. ☺

Merci à ma gang du Cégep Saint-Jean-sur-Richelieu qui sont bien patients avec mon double, voire triple, emploi du temps : Michèle Roberge, Marie-Pier Rivard, Dominique Pinard, Josée Fontaine, Julie Marcil, Camille Miclette, et même Pascal, Bianka, Katherine, Line, merci d'être présents dans ces derniers moments de travail sur ma thèse. Les filles du SAIDE, notre douce folie se révèle un moyen indispensable de demeurer saines d'esprit, j'en ai souvent besoin, merci!

Je remercie chaleureusement Audray Grandmaison (année 1 du projet), Marie-France Pagé et Marie-Josée Mongeau pour leur confiance inébranlable envers le présent projet de recherche… et d'avoir convaincu tant de gens que cela valait la peine de participer aux collectes de données, merci! Je remercie également les directions, directions adjointes, enseignants.es, orthpédagogues, suppléants, etc. qui ont collaboré de près ou de loin à la collecte, y compris la fée des enveloppes. ;)

Je tiens finalement surtout à remercier du fond du cœur tous les élèves qui ont participé à cette recherche, notamment les élèves qui se sont déplacés à NeuroLab avec leur accompagnatrice pour venir vivre cette expérience unique avec nous. Sans vous, rien de cela n'aurait été possible.

Merci, merci, merci…

DÉDICACE

À mon papa, qui m'a encouragé dans mes études pendant toute sa vie, malgré qu'il n'ait jamais terminé son secondaire.
Bonne lecture de là-haut… (en diagonale, je sais!)

À Louis-Jacob et Charlie, mes amours. Du haut de vos 11 et 6 ans, merci d'avoir accepté de me partager avec cette thèse. En espérant que cette expérience vous servira de modèle de ténacité. Quand on veut, on peut!

TABLE DES MATIÈRES

REMERCIEMENTS ii

DÉDICACE v

LISTE DES FIGURES ix

LISTE DES TABLEAUX xii

RÉSUMÉ xiv

ABSTRACT xvi

CHAPITRE 1 PROBLÉMATIQUE 18

1.1 Premiers apprentissages en lecture et élèves en difficulté 18

1.2 Évaluation et intervention orthopédagogiques 19

1.3 Compensation des difficultés d'identification des mots écrits 22

1.4 Étude de l'identification des mots écrits 26

1.5 Objectif général de recherche et questions de recherche 27

1.6 Pertinence de la recherche et retombées anticipées 28

1.6.1 Pertinence sociale 28

1.6.2 Pertinence scientifique 29

CHAPITRE 2 CADRE THÉORIQUE 32

2.1 Caractère transdisciplinaire de la thèse au sein des sciences de l'éducation 32

2.2 Processus cognitifs de lecture : de l'identification des mots écrits à la compréhension 35

2.3 Modèles théoriques interactifs des processus cognitifs de lecture 39

2.3.1 Modèle de construction-intégration de van Dijk et Kintsch (1983) 40

2.3.2 Cadre des systèmes de la lecture de Perfetti (1999) et de Perfetti et Stafura (2014) 41

2.4 Modèles théoriques d'identification des mots écrits 43

2.4.1 Considérations préalables concernant les caractéristiques linguistiques de l'orthographe française 44

2.4.2 Modèles théoriques de la lecture experte 46

2.4.2.1 Modèle à deux voies indépendantes 46

2.4.2.2 Modèle à deux voies en cascade (DRC) 48

2.4.2.3 Modèle à traitement parallèle distribué (PDP) 50

2.4.3 Modèles théoriques développementaux de l'identification des mots écrits et le modèle à double fondation 53

2.4.4 Théorie de la granularité psycholinguistique 57

2.4.5 Brève critique des modèles théoriques 61

2.5 Sémantique lexicale 63

2.5.1 Caractéristiques sémantiques des mots et leur représentation interne chez le lecteur 64
2.5.1.1 Distinction forme et sens : homonymie et synonymie 64
2.5.1.2 Liens de sens entre les mots : hyponymie, prototype et antonymie 65
2.5.2 Modèle théorique cognitif de la sémantique lexicale dans l'identification des mots écrits : le modèle computationnel distribué 68
2.6 Électrophysiologie de l'identification des mots écrits 70
2.7 Les difficultés d'identification des mots écrits 75
2.7.1 Profils de lecteurs en difficulté de lecture 76
2.7.2 Manifestations des difficultés d'identification des mots écrits 78
2.7.2.1 Importance des difficultés de décodage 79
2.7.2.2 Précision et latence de l'identification des mots écrits 80
2.7.3 Profil atypique de lecteurs en difficulté 82
2.8 Compensation d'une difficulté d'identification des mots écrits 84
2.8.1 Effet de contexte 84
2.8.2 Le modèle interactif-compensatoire (Stanovich, 1980) 88
2.8.3 Le modèle d'encodage compensatoire de Walczyk (2000) 90
2.9 Étude de la sémantique dans l'identification des mots écrits : l'amorçage sémantique 92
2.9.1 Amorçage sémantique et données comportementales 93
2.9.2 Amorçage sémantique et données électrophysiologiques 96
2.10 Questions spécifiques et hypothèses de recherche 101
CHAPITRE 3 ÉTUDE 1 – MÉTHODOLOGIE ET RÉSULTATS 102
3.1 Méthodologie 102
3.1.1 Description des variables de l'étude 1 102
3.1.2 Participants 104
3.1.2.1 Recrutement 106
3.1.2.2 Caractéristiques de l'échantillon 107
3.1.3 Sélection des instruments de mesure 109
3.1.3.1 Instrument de mesure de la compréhension écrite 109
3.1.3.2 Instrument de mesure de la précision et de la fluence de l'identification des mots écrits 112
3.1.4 Déroulement 116
3.2 Résultats de l'étude 1 117
3.2.1 Données descriptives des trois instruments 117
3.2.2 Corrélations entre les variables 127
3.2.2.1 Corrélations des instruments qui mesurent l'identification des mots écrits 127
3.2.2.2 Corrélations entre la compréhension écrite et l'identification des mots écrits 130
3.3 Interprétation des résultats : vers l'identification de lecteurs compensateurs au sein de l'échantillon 131
3.3.1 Compréhension écrite 132
3.3.2 Identification des mots écrits 134
3.3.2.1 Précision ou fluence de l'identification des mots écrits 136
3.3.2.2 Fluence de l'identification des mots écrits : quel type d'items? 139
3.3.3 Sélection finale des lecteurs compensateurs 142

CHAPITRE 4 ÉTUDE 2 – MÉTHODOLOGIE ET RÉSULTATS 144
4.1 Commentaire général sur l'approche méthodologique 144
4.2 Méthodologie 146
4.2.1 Devis de recherche et description des variables 146
4.2.1.1 Description des variables indépendantes 148
4.2.1.2 Description des variables dépendantes 150
4.2.1.3 Description des variables contrôles 151
4.2.2 Participants 152
4.2.2.1 Recrutement des participants 153
4.2.2.2 Caractéristiques des participants 155
4.2.3 Instruments de mesure 158
4.2.3.1 Élaboration des tâches expérimentales 159
4.2.3.2 Sélection des instruments de mesure pour les composantes associées à la lecture-écriture 168
4.2.4 Déroulement 170
4.2.5 Préparation des données d'électroencéphalographie 172
4.3 Résultats de l'étude 2 174
4.3.1 Résultats des composantes associées à la lecture-écriture 174
4.3.2 Tâche expérimentale 177
4.3.2.1 Données comportementales (précision et latence) 178
4.3.2.2 Données obtenues par les potentiels évoqués (amplitude et latence) 182
4.3.2.2.1 Données d'amplitude du N400 (SOA court) 183
4.3.2.2.2 Données de latence du N400 (SOA court) 184
4.3.2.2.3 Données d'amplitude du P200 (SOA court) 185
4.3.2.2.4 Données de latence du P200 (SOA court) 186
4.3.2.2.5 Données d'amplitude du N400 (SOA long) 187
4.3.2.2.6 Données de latence du N400 (SOA long) 188
4.3.2.2.7 Données d'amplitude et de latence du P200 (SOA long) 189
4.4 Discussion des résultats 189
4.4.1 Composantes associées à la lecture-écriture 190
4.4.2 Mobilisation de la sémantique lexicale en situation de lecture de mots isolés: amorçage sémantique 193
4.4.2.1 Effet facilitateur de la sémantique dans la précision et les latences 194
4.4.2.2 Activation précoce de la sémantique lexicale chez les lecteurs compensateurs 199
CONCLUSION 206
LISTE DE RÉFÉRENCES 211
ANNEXE A Phonèmes du français et symboles de l'alphabet phonétique international 230

LISTE DES FIGURES

Figure 2.1 Cadre des systèmes de la lecture (Perfetti et Stafura, 2014). Traduction libre.....42

Figure 2.2 Représentation schématique du modèle à deux voies indépendantes *(Coltheart, 1978; Coltheart, Patterson et Marshall, 1980)* 47

Figure 2.3 Modèle à deux voies en cascade (DRC), de Coltheart *et al.* (2001) 49

Figure 2.4 Distinction entre représentations locales et distribuées 51

Figure 2.5 Modèle à traitement parallèle distribué (Parallel Distributed Processing, PDP), traduit et adapté de McClelland et Seidenberg (1989) et de Plaut *et al.* (1996).....52

Figure 2.6 Modèle à double fondation, traduit et adapté de Seymour (2008).....55

Figure 2.7 Représentation schématique des trois principaux défis de l'apprentissage de la lecture : disponibilité, consistance et granularité (traduction de Ziegler et Goswami, 2005).....60

Figure 2.8 Illustration de l'organisation hiérarchique d'un réseau sémantique (Collins et Quillian, 1969, traduit par Ferrand, 2007).....66

Figure 2.9 Représentation du fonctionnement du traitement sémantique dans l'amorçage sémantique selon le modèle computationnel distribué (traduction de Plaut et Booth, 2000).....70

Figure 2.10 Représentation du décours temporel de l'identification des mots écrits, selon l'état actuel de la littérature scientifique dans le domaine.....74

Figure 2.11 Profils de difficulté de lecture selon le modèle Simple View of Reading (traduit et adapté de Stuart *et al.*, 2008; Tunmer et Greany, 2010).....77

Figure 2.12 Quelques éléments composant les tâches d'amorçage sémantique « classiques ».....93

Figure 3.1 Variables de l'étude 1 et indicateurs de compensation associés 104

Figure 3.2 Distribution des scores obtenus à l'épreuve de compréhension écrite « Le Lion » (Bianco *et al.*, 2018) 117

Figure 3.3 Distribution des scores de l'identification de mots et de pseudomots du sous-test MIM (BELEC) 118

Figure 3.4 Distribution des scores de précision obtenus sur les mots du sous-test MIM (BELEC) 120

Figure 3.5 Distribution des scores de précision sur les pseudomots au sous-test MIM (BELEC).....120

Figure 3.6 Distribution des scores de la fluence de l'identification de mots et pseudomots du sous-test MIM (BELEC) 122

Figure 3.7 Distribution des scores de fluence de reconnaissance de mots, au sous-test MIM (BELEC) ...123

Figure 3.8 Distribution des scores de fluence du décodage de pseudomots au sous-test MIM (BELEC)..123

Figure 3.9 Distribution des scores de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013).....125

Figure 3.10 Distribution des scores de fluence de la reconnaissance de mots au Test 1 minute (Sprenger-Charolles et Leloup, 2013) 126

Figure 3.11 Distribution des scores de fluence du décodage de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013) 126

Figure 3.12 Terminologie utilisée pour qualifier la performance des élèves en comparaison à la moyenne du groupe, sur une courbe normale hypothétique.....133

Figure 3.13 Les 9 variables de l'identification de mots et de pseudomots du sous-test MIM (BELEC) et du Test 1 minute (Sprenger-Charolles et Leloup, 2013) 135

Figure 3.14 Répartition des élèves selon leur score Z de compréhension écrite (axe des y) et de fluence de décodage des pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013) 143

Figure 4.1 Résumé du devis de recherche de l'étude 2 mettant en évidence les variables indépendantes 148

Figure 4.2 Répartition des élèves de l'échantillon de l'étude 1, selon leur score de compréhension écrite et en fluence de décodage de pseudomots, et identification des participants potentiels à l'étude 2 154

Figure 4.3 Constitution d'une liste d'items pour la tâche expérimentale 162

Figure 4.4 Structure générale de la séquence des stimuli visuels dans la tâche d'amorçage sémantique 166

Figure 4.5 Structure de la tâche d'amorçage sémantique. SOA court à gauche; SOA long à droite.....167

Figure 4.6 Regroupements d'électrodes pour les données d'ERP d'amorçage sémantique. Délai (SOA) court à gauche, délai (SOA) long à droite 174

Figure 4.7 Scores obtenus en mémoire de travail phonologique (WISC) selon le type de lecteurs 176

Figure 4.8 Scores obtenus en étendue de vocabulaire réceptif (EVIP) selon le type de lecteurs.....176

Figure 4.9 Scores obtenus en conscience phonologique (CELF) par type de lecteurs 177

Figure 4.10 Temps obtenus en dénomination rapide de chiffres (BALE) selon le type de lecteurs.....177

Figure 4.11 Moyennes des taux de réussite et des latences pour le délai (SOA) court par type d'items et par type de lecteurs. 180

Figure 4.12 Moyennes des taux de réussite et des latences pour le délai (SOA) long par type d'items et par type de lecteurs. 181

Figure 4.13 Résultats de l'amplitude des potentiels évoqués sur la N400 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite. 184

Figure 4.14 Résultats des latences (en ms) des potentiels évoqués sur la N400 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite. 185

Figure 4.15 Résultats de l'amplitude des potentiels évoqués sur la P200 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite. 186

Figure 4.16 Résultats de l'amplitude des potentiels évoqués sur la N400 selon la latéralité au délai (SOA) long.....188

Figure 4.17 Résultats des latences (en ms) des potentiels évoqués sur la N400 au délai (SOA) long. 189

LISTE DES TABLEAUX

Tableau 2.1 Classification théorique de certaines langues selon la complexité syllabique et l'opacité de leur orthographe (Seymour *et al.*, 2003).....46

Tableau 2.2 Résumé des principales caractéristiques des modèles théoriques de l'identification des mots écrits présentés dans le cadre théorique 62

Tableau 3.1 Répartition des élèves par centre de services scolaires, par écoles et par classes.....107

Tableau 3.2 Répartition des élèves de l'échantillon selon leur langue maternelle 109

Tableau 3.3 Résumé des caractéristiques primordiales de l'instrument de mesure de la compréhension écrite 110

Tableau 3.4 Résumé des caractéristiques primordial à l'instrument de mesure de l'identification des mots écrits.....113

Tableau 3.5 Procédures de calculs du nombre de mots bien lus en une minute selon les deux possibilités de considération du temps.....116

Tableau 3.6 Données desriptives de l'épreuve de compréhension écrite « Le lion ».....118

Tableau 3.7 Données descriptives de précision de l'identification des mots écrits selon MIM (BELEC)...119

Tableau 3.8 Données descriptives de précision du sous-test MIM (BELEC) pour les mots et les pseudomots 121

Tableau 3.9 Données descriptives de fluence de l'identification des mots et des pseudomots au sous-test MIM (BELEC) 122

Tableau 3.10 Données descriptives du fluence d'identification de mots et de pseudomots.....124

Tableau 3.11 Données descriptives de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013).....125

Tableau 3.12 Données descriptives de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013).....127

Tableau 3.13 Corrélations des différentes mesures de fluence en identification de mots et de pseudomots 128

Tableau 3.14 Corrélations de la précision et de la fluence au sous-test MIM (BELEC) 129

Tableau 3.15 Corrélations entre les mesures de précision du sous-test MIM et de fluence du Test 1 minute (Sprenger-Charolles et Leloup, 2013).....129

Tableau 3.16 Corrélations entre la compréhension écrite, la précision et la fluence de l'identification des mots et des pseudomots au sous-test MIM (BELEC) et au Test 1 minute (Sprenger-Charolles et Leloup, 2013) 130

Tableau 3.17 Score Z des élèves « faibles » sur une des mesures du sous-test MIM (BELEC) 137

Tableau 3.18 Score Z des élèves « faibles » sur le plan de la précision de l'identification des mots écrits (à gauche) ou sur le plan de la fluence (à droite), au sous-test MIM (BELEC).....139

Tableau 3.19 Score Z des élèves « faibles » sur une mesure d'identification de mots et de pseudomots au sous-test MIM (BELEC) et au Test 1 minute (Sprenger-Charolles et Leloup, 2013).....140

Tableau 4.1 Répartition des élèves par centre de services scolaires et par type de lecteurs 156

Tableau 4.2 Caractéristiques supplémentaires des participants 158

Tableau 4.3 Description des différentes conditions et des items de la tâche expérimentale d'amorçage sémantique.....159

Tableau 4.4 Caractéristiques linguistiques des items par condition 165

Tableau 4.5 Moyennes et écarts-types des mesures portant sur les composantes associées à la lecture pour chaque type de lecteurs.....175

Tableau 4.6 Moyennes et écarts-types à la tâche d'amorçage sémantique 182

Tableau 4.7 Présence d'un effet facilitateur de la sémantique 194

RÉSUMÉ

Il semble exister une catégorie d'élèves en difficulté de lecture, qui, malgré un déficit important des processus spécifiques d'identification des mots écrits, arrivent tout de même à bien comprendre un texte lu de manière autonome. Ces élèves sont appelés ici des lecteurs compensateurs et font l'objet de la présente thèse.

Une importance accrue est accordée à la prévention des difficultés de lecture dans les milieux scolaires. Ainsi, plusieurs initiatives qui visent à détecter rapidement les élèves à risque de présenter des difficultés de lecture sont mises en place. Malgré ces efforts, il semble que certains élèves qui ont effectivement des difficultés de lecture passent inaperçus (et ne sont donc pas repérés comme étant en difficulté), puisque leurs difficultés sont non apparentes en situation de lecture de texte; ils sont en quelque sorte fonctionnels en classe malgré leurs difficultés. Pourtant, il n'est pas rare de constater que ces difficultés deviennent manifestes plusieurs années plus tard, et empêchent à ce moment la personne d'être fonctionnelle en lecture. De ce fait, il semble pertinent de se pencher sur la question de la compensation en lecture, particulièrement de son fonctionnement.

Une première étude a pour but de repérer les lecteurs compensateurs dans un échantillon de 165 élèves de la 5^{e} année du primaire. Parmi les instruments de mesure utilisés pour ce faire, ce sont les mesures de compréhension écrite avec des questions de compréhension posées à l'oral, ainsi qu'un test « 1 minute » de lecture orale de pseudomots qui ont été sélectionnés pour repérer les lecteurs compensateurs. Une douzaine d'élèves de l'échantillon ont ainsi été identifiés comme étant des lecteurs compensateurs, ce qui représente environ 7% des élèves de l'échantillon. Une deuxième étude a eu lieu avec 14 lecteurs compensateurs (dont la majorité provient de l'étude 1) et 15 lecteurs contrôles, de même niveau scolaire et de même niveau de compréhension écrite. Cette étude, malgré le faible nombre de participants, a permis de mettre en lumière des différences significatives entre les deux groupes d'élèves sur une mesure de conscience phonologique et de dénomination rapide de chiffres, mais pas sur des mesures de vocabulaire réceptif, de fluence verbale ou de mémoire de travail phonologique. Une tâche d'amorçage sémantique a également été réalisée, et a montré que la compensation en situation de lecture de texte se manifeste également en situation de lecture de mots isolés ; les lecteurs compensateurs bénéficient d'un lien sémantique entre deux mots présentés à un délai (SOA) court (150 ms), notamment au niveau des scores de précision. Les lecteurs compensateurs semblent présenter une sensibilité particulière à la

sémantique, qui se manifeste aussi tôt que 200 ms après la présentation du mot ; il y a un effet de l'amorçage sur la composante P200 chez les lecteurs compensateurs uniquement.

Mots clés : identification de mots écrits, compensation en lecture, sémantique, élèves en difficulté, orthopédagogie, ERP, neurosciences éducationnelles

ABSTRACT

Abstract

There seems to be a category of students with reading difficulties who, despite a significant deficit in the specific processes of identifying written words, still manage to understand a text read independently. These students are called here compensating readers and are the subject of this thesis.

Increased importance is given to the prevention of reading difficulties in schools. Thus, several initiatives aimed at quickly detecting students at risk of having reading difficulties have been put in place. Despite these efforts, it seems that some students who do have reading difficulties go unnoticed (and are therefore not identified as being in difficulty), since their difficulties are not apparent when reading text; they are somehow functional in class despite their difficulties. However, it is not uncommon to find that these difficulties become manifest several years later, and at that time prevent the person from being functional in reading. Therefore, it seems relevant to address the issue of read compensation, particularly how it works.

A first study aims to identify compensating readers in a sample of 165 students in the 5th year of primary school. Among the measuring instruments used to do this, reading comprehension measures with comprehension questions asked orally, as well as a "1 minute" test of oral reading of pseudowords were selected to identify readers compensators. A dozen students in the sample were thus identified as compensating readers, which represents approximately 7% of the students in the sample. A second study took place with 14 compensatory readers (the majority of whom come from study 1) and 15 control readers, of the same school level and the same level of written comprehension. This study, despite the small number of participants, made it possible to highlight significant differences between the two groups of students on a measure of phonological awareness and rapid naming of numbers, but not on measures of receptive vocabulary, fluency verbal or phonological working memory. A semantic priming task was also carried out, and showed that compensation when reading text also occurs when reading isolated words; Compensating readers benefit from a semantic link between two words presented at a short SOA (150 ms), especially in terms of accuracy scores. Compensating readers appear to exhibit a particular sensitivity to semantics, which manifests as early as 200 ms after word presentation; there is an effect of priming on the P200 component in compensating readers only.

Keywords : word recognition, word reading, reading compensation, semantics, at risk students, orthopedagogy, ERP, educational neurosciences

CHAPITRE 1

PROBLÉMATIQUE

Au Québec, malgré les améliorations des dernières années, la situation du décrochage scolaire est préoccupante : en 2013-2014, près de 22,3% des élèves sortent du système scolaire sans l'obtention d'un diplôme ou d'une qualification (ministère de l'Éducation et de l'Enseignement supérieur (MEES), 2016). Les coûts sociaux, économiques, ainsi que les conséquences individuelles associés au décrochage sont très importants (ministère de l'Éducation, du Loisir et du Sport (MELS), 2013). En 2009, on estimait que la situation du décrochage scolaire coûte annuellement au Québec environ 1,9 milliard de dollars, soit en revenus non perçus, soit en dépenses sociales additionnelles généralement associées au décrochage et à ses effets dans le temps (MEES, 2016).

Il existe un lien entre les difficultés de lecture vécues au début du primaire et le décrochage scolaire à l'adolescence, tel qu'en témoignent certaines études qui lient les difficultés de lecture à 7 ans à un risque plus élevé de décrochage scolaire à l'adolescence (Janosz *et al.*, 2013), ou des statistiques du Gouvernement du Québec qui rapportent que presque le tiers des décrocheurs de la formation générale des jeunes, lesquels présentent des difficultés graves d'apprentissage, avaient déjà accumulé du retard dès l'âge de 7 ans (ministère de l'Éducation du Québec (MEQ), 1997). Également, environ 90 % des enfants qui ont de faibles habiletés en lecture au cours de la première année de scolarisation demeurent « à risque » tout au long de l'élémentaire (de Jong et van der Leij, 2003; Landerl et Wimmer, 2008; McGill-Franzen et Allington, 1991). Le retard constaté dès le début de l'apprentissage de la lecture semble donc persister, et il peut avoir de lourdes conséquences pour ces individus (voir le rapport de l'Institut de recherche sur la santé au Canada (IRSC), 2004, ainsi que le feuillet de complément du MEES (2016) pour un résumé des principales conséquences du décrochage sur la qualité de vie). Dans ce contexte, s'intéresser à ce qui constitue le défi des premiers apprentissages en lecture s'avère judicieux, afin d'éviter plusieurs problématiques, tant individuelles, que sociales ou économiques.

1.1 Premiers apprentissages en lecture et élèves en difficulté

Bien que la finalité de la lecture soit la compréhension du message véhiculé par le texte (Gough et Tunmer, 1986; Stuart *et al.*, 2008), ce sont les processus d'identification des mots écrits – c'est-à-dire la lecture de

mots[1] – qui constituent le principal défi pour l'apprenti-lecteur. Ceci est appuyé par la grande part de variance de la performance en lecture de texte (aussi appelée compréhension écrite) qui leur est attribuable dans les premières années d'apprentissage. Ainsi, de façon générale, plus un apprenti-lecteur excelle en identification de mots, mieux il performe dans des tâches de lecture de texte (p. ex. Catts et Kamhi, 2005). Inversement, plus un apprenti-lecteur présente des difficultés à identifier les mots, moins il tend à performer en lecture de texte (Landi *et al.*, 2006). Qui plus est, plusieurs études rapportent que ce sont ces processus d'identification de mots qui sont en cause pour la majorité des adultes et des enfants qui manifestent des difficultés de lecture (Aaron, 1989; Desrochers, 2009; Perfetti, 1985; Siegel et Faux, 1989). Il convient donc de se pencher sur ces processus, notamment en s'intéressant à ce qui est mis en place en milieu scolaire pour soutenir les élèves en début d'apprentissage de la lecture, et plus particulièrement pour ceux qui présentent des difficultés à identifier les mots écrits.

1.2 Évaluation orthopédagogique

Dans le contexte scolaire québécois actuel, on accorde de plus en plus d'importance à la prévention des difficultés d'apprentissage de la lecture, comme en témoignent la politique de l'adaptation scolaire (MELS, 1999), la politique de la réussite éducative (MEES, 2017), ou les différents projets d'implantation du modèle de la réponse à l'intervention dans certains centres de services scolaires, anciennement les commissions scolaires (ministère de l'Éducation et de l'Enseignement supérieur, 2020; St-Pierre, 2012). On désire ainsi identifier rapidement les élèves à risque de présenter des difficultés et leur apporter immédiatement un soutien adéquat, ceci afin d'éviter que ces difficultés ne se manifestent ou ne s'aggravent au point d'empêcher les individus d'apprendre à lire efficacement. Il s'agit là d'un des principes directeurs du modèle de réponse à l'intervention (Haager *et al.*, 2007).

En contexte scolaire québécois, l'orthopédagogie joue un rôle important pour les élèves à risque, en difficulté ou en trouble d'apprentissage (Association des orthopédagogues du Québec, 2018; Brodeur *et al.*, 2015; Laplante, 2012; Prud'homme, 2018). L'orthopédagogie est définie par l'ADOQ (2018) comme « la science de l'éducation dont l'objet est l'évaluation et l'intervention relatives aux personnes qui présentent des difficultés d'apprentissage scolaire, incluant les troubles d'apprentissage ». Puisque la population des élèves qui présentent des difficultés de lecture est hétérogène, en ce sens que le déficit peut toucher l'un ou l'autre des nombreux processus de lecture (Braibant, 1994; voir le cadre théorique),

[1] Il est à noter que l'expression « identification de mots » est parfois utilisée dans la thèse pour désigner l'« identification des mots écrits », ceci dans le seul but d'alléger le texte.

c'est l'orthopédagogue qui procède à une évaluation précise des difficultés de lecture, puis qui détermine et met en œuvre des interventions adaptées aux besoins de ces élèves (Laplante, 2011). Tel que mentionné précédemment, les processus d'identification des mots écrits sont fréquemment mis en cause dans les difficultés de lecture. S'inspirant actuellement fortement des sciences cognitives (Prud'homme 2018), l'évaluation orthopédagogique des processus d'identification des mots écrits est réalisée en demandant à l'élève chez qui on soupçonne des difficultés au regard de ces processus de lire des mots présentés de façon isolée. Ces mots sont soigneusement choisis selon leurs caractéristiques linguistiques et sont regroupés en catégories; on distingue par exemple les mots fréquents des mots rares, les mots comportant des graphèmes complexes des mots qui ne comportent que des graphèmes simples, les mots irréguliers des mots réguliers, etc. (p. ex. Alegria et Mousty, 1996). La performance de l'élève à l'égard de ces différentes catégories d'items à lire est interprétée à l'aide de modèles théoriques qui visent à décrire les différents mécanismes cognitifs impliqués dans l'identification de mots. Cette interprétation permet de brosser un portrait des mécanismes qui sont fonctionnels ou déficitaires chez l'élève évalué (voir la section 2.4 pour la description de ces modèles). C'est donc à la suite de cette évaluation précise des mécanismes d'identification de mots que l'orthopédagogue élabore et met en œuvre une intervention adaptée aux besoins de l'élève.

L'orthopédagogue joue également un rôle important dans la prévention des difficultés de lecture-écriture, notamment en lien avec le dépistage des élèves qui sont à risque de présenter des difficultés (Association des orthopédagogues du Québec, 2018; Brodeur *et al.*, 2015). En collaboration avec les enseignants, plusieurs efforts sont déployés dans les milieux scolaires pour cibler rapidement les élèves à risque ou qui présentent des difficultés de lecture. Malgré cela, il demeure une certaine proportion d'élèves dont les difficultés d'identification de mots ne sont repérées que tardivement (Catts *et al.*, 2012). En effet, il n'est pas rare que les enseignants de la fin du primaire ou du secondaire réfèrent en orthopédagogie un élève qui n'avait pas présenté de difficultés de lecture évidentes jusque-là. En ce qui concerne l'identification de mots, des difficultés importantes ne sont parfois remarquées qu'au cours de la 3e ou de la 4e année de scolarité du primaire (van Viersen, *et al.*, 2016), pendant le secondaire (Leach *et al.*, 2003; Catts *et al.*, 2012), voire au postsecondaire (Mimouni et King, 2007). Ce phénomène de difficultés d'identification de mots repérées tardivement est documenté notamment chez des élèves qui manifestent plus de difficultés en écriture qu'en lecture (Andrews, 2008), chez ceux qui présentent un « trouble de l'orthographe sans trouble apparent de la lecture » (Rodrigue, 2006), ou chez les élèves doués qui présentent une dyslexie (van Viersen *et al.*, 2016; Berninger et Abbott, 2013), par exemple.

De manière générale, une difficulté au niveau de l'identification de mots pourrait être repérée rapidement, à condition que l'on adopte de bonnes pratiques de dépistage. D'une part, le principal défi des apprentis lecteurs est l'apprentissage de l'identification des mots écrits (Laplante et Turgeon, 2021). Ceci fait en sorte que les enseignants des niveaux concernés par les premiers apprentissages passent beaucoup de temps à travailler ces processus en classe. D'autre part, comme il a été mentionné précédemment, la capacité en compréhension écrite est fortement reliée aux capacités en identification de mots en début d'apprentissage. Par conséquent, les élèves qui ont des difficultés d'identification des mots écrits devraient être généralement aisément identifiables en classe, peu importe que l'évaluation porte sur la compréhension écrite ou l'identification de mots isolés. Pourtant, il n'en ait pas toujours ainsi, puisque certaines difficultés ne sont remarquées que tardivement. Ceci indique que, malgré leur difficulté d'identification des mots écrits, ces élèves ont été en mesure de « suivre » le rythme de leurs pairs pendant un certain temps en camouflant leurs difficultés de lecture. Vers la fin du primaire, au secondaire ou au postsecondaire, ils peuvent ne plus arriver à camoufler autant leurs difficultés; l'écart avec leurs pairs devient alors manifeste. Ainsi, les difficultés d'identification de mots, bien qu'elles aient été présentes pendant toutes ces années, peuvent passer inaperçues pendant un certain temps, notamment si ces difficultés n'ont eu que très peu d'incidence sur la capacité en compréhension écrite. Ces difficultés n'empêchent donc pas les élèves, pendant un temps en tout cas, de mettre en œuvre leur compétence à lire. Toutefois, il semble que ces difficultés peuvent devenir apparentes à un moment ou à un autre du parcours scolaire; par exemple quand les attentes augmentent en lecture, quand les textes deviennent plus complexes, ou quand les thèmes abordés sont moins connus des lecteurs (Lipka *et al.*, 2006). Par ailleurs, il est difficile de savoir si tous les élèves qui compensent vont voir leurs difficultés devenir manifestes à un moment donné. En effet, certains adultes qui ont un trouble d'apprentissage, qui altère par définition l'identification des mots écrits, sont tout à fait fonctionnels dans leur quotidien (Binder et Lee, 2012). Il apparaît donc qu'il est actuellement difficile de prédire si toutes les difficultés deviennent manifestes ou pas.

Les problèmes soulevés par ce type particulier d'élèves en lien avec la pratique orthopédagogique sont nombreux. D'une part, puisque leurs difficultés ont peu d'incidence sur la compréhension écrite pendant un certain temps, il est difficile de repérer ces élèves au sein des groupes-classes. Ceci est d'autant plus vrai si seule la compétence à lire (selon le programme ministériel) est évaluée, sans pratiques de dépistage qui dépassent cette prescription. Ces élèves risquent ainsi d'évoluer sans que personne ne remarque leurs difficultés sur le plan de l'identification de mots. A priori, il pourrait être tentant de suggérer que puisque

les difficultés de ces élèves ne sont pas manifestes au point d'entraver la compréhension écrite, c'est qu'ils n'ont pas besoin de soutien puisqu'ils performent bien dans la « compétence à lire ». On suggèrerait ainsi qu'arriver à camoufler des difficultés est une excellente manière d'être un lecteur fonctionnel. Dans ce cas, ne pas repérer les élèves qui ont effectivement des difficultés d'identification de mots au sein des groupes-classes ne serait pas à proprement parler problématique. Par contre, sachant que ces lecteurs, que l'on peut qualifier de fonctionnels à un moment de leur cheminement scolaire, peuvent voir leurs difficultés devenir manifestes plus tard, on ne peut tenir ce raisonnement; être fonctionnel en lecture malgré une difficulté d'identification de mots à un moment donné ne garantit pas du tout que ce lecteur sera fonctionnel à un moment ultérieur. Les milieux scolaires ont donc tout intérêt à repérer ces élèves le plus rapidement possible et à leur offrir le soutien dont ils ont besoin pour prévenir une apparition tardive de leur difficulté. De plus, par leur profil de performance un peu inattendu (cf. leurs difficultés d'identification de mots n'entravent pas autant que prévu la compréhension écrite), les évaluations réalisées en orthopédagogie peuvent ne pas tenir compte de la possibilité que des difficultés d'identification de mots puissent être « camouflées ». Certains outils sont systématiquement utilisés en évaluation orthopédagogique au Québec à cause de la disponibilité d'outils (p. ex. texte et questions de compréhension), mais ne constituent sans doute pas le meilleur outil pour mettre en évidence les difficultés d'identification des mots écrits chez ces élèves qui arrivent à les camoufler. La performance de ces élèves, même en évaluation orthopédagogique, peut donc amener les orthopédagogues à sous-estimer la gravité des difficultés en identification des mots écrits, puisque les outils sont construits et interprétés en tenant compte du fait que les difficultés d'identification des mots écrits entraînent généralement une difficulté en compréhension écrite.

1.3 Compensation des difficultés d'identification des mots écrits

Une des hypothèses avancées pour expliquer que des difficultés d'identification de mots puissent passer inaperçues pour un certain nombre d'élèves a trait à la compensation qui peut être mise en œuvre par ces derniers (p. ex. Share, 1995; Stanovich, 2000). Pour la lecture, Stanovich (1984) propose l'hypothèse selon laquelle les lecteurs qui présentent un déficit de l'un des processus de lecture peuvent recourir aux informations fournies par d'autres processus pour compenser cette difficulté (voir chapitre 2 pour davantage de détails). Cette hypothèse est fondée sur un phénomène répliqué de nombreuses fois auprès de participants ayant différentes caractéristiques (voir Stanovich, 2000 pour une revue de littérature), phénomène appelé l'*effet de contexte*. Ce sont les travaux de Perfetti *et al.* (1979) et de Stanovich (1980) qui en sont les études princeps. Ces études ont comparé la performance en identification de mots, lorsque

ces mots sont présentés en contexte, comme au sein d'une phrase, ou lorsque ces mots sont présentés de manière isolée. Elles montrent ainsi que ce sont les lecteurs en difficulté qui identifient mieux et plus rapidement les mots écrits lorsqu'ils sont présentés dans un contexte, alors que ce n'est pas le cas pour les lecteurs sans difficulté. La présence d'un contexte facilite ainsi le processus d'identification de mots chez les lecteurs en difficulté alors qu'elle change peu de choses pour le lecteur sans difficulté; recourir au contexte dans lequel le mot est présenté pourrait ainsi être une manière de compenser une difficulté d'identification des mots écrits.

Les contextes dont il est question dans ces études sont variés. Ils peuvent être constitués de mots, de phrases, de paragraphes ou de texte complet (voir le chapitre 2 pour les détails). Il semble donc que des informations provenant de ces contextes puissent être utilisées dans la compensation. Lorsque le contexte est un texte, un paragraphe ou une phrase, les informations pouvant être utilisées sont nombreuses puisque plusieurs unités entrent en jeu (p. ex. la syntaxe, le thème, la structure du texte, les idées, etc.) Par contre, quand le contexte est constitué d'un seul mot, la nature des informations pouvant être utilisées pour compenser la difficulté à identifier les mots est plus réduite. Dans les études sur l'effet de contexte, lorsque le contexte est un mot, c'est l'amorçage sémantique qui est utilisé (Ferrand, 2007). L'amorçage sémantique consiste en la présentation des mots en paire, le premier mot étant relié sémantiquement au second mot à identifier (McNamara, 2005). Ainsi, maintes fois, il a été documenté que les lecteurs en difficulté d'identification de mots écrits sont ceux pour lesquels la performance est grandement améliorée par la présentation préalable de ce mot sémantique lié (p. ex. Schvaneveldt *et al.*, 1977; voir Stanovich, 2000 pour une revue). Ainsi, il est possible d'envisager que la sémantique, c'est-à-dire le sens qui est véhiculé (Brousseau et Roberge, 2000), soit impliquée dans la compensation d'un déficit d'identification des mots écrits. Au quotidien, au niveau du mot, un élève pourrait identifier un mot de façon erronée, et chercher dans les mots qu'il connaît à l'oral un mot qui conviendrait et qui est similaire à ce qu'il a produit (p. ex. Laplante, 1998; Sprenger-Charolles et Casalis, 1996); au niveau de la phrase, il pourrait utiliser le sens de la phrase (groupes de mots avant et après) pour deviner le mot qu'il n'arrive pas à identifier, etc. Nation et Snowling (1998) ont par contre mis en évidence que tous les lecteurs en difficulté ne sont pas en mesure d'utiliser les informations fournies par le contexte pour compenser leurs difficultés : les élèves qui présentent des difficultés de compréhension orale (que l'on appelle aussi les faibles compreneurs) ne présentent pas cet effet de contexte déjà documenté chez les lecteurs en difficulté sur le plan de de l'identification de mots. Quoi qu'il en soit, l'ensemble de ces données empiriques concernant l'effet de contexte semble suggérer que l'information sémantique d'un mot (sémantique lexicale), d'une phrase

(sémantique phrastique) ou d'un texte (sémantique textuelle) puisse être utilisée pour compenser des difficultés à identifier les mots écrits chez certains lecteurs (tel que formulé par Share, 1995).

Depuis l'hypothèse de compensation formulée par Stanovich (1984) dans son modèle, et suivant les nombreuses études sur l'effet de contexte, le fait que les élèves qui ont une difficulté d'identification des mots écrits puissent compenser cette difficulté dans différentes circonstances qui impliquent la présence d'un contexte est grandement accepté. Ainsi, un nombre incommensurable de chercheurs font référence à ce phénomène : on y retrouve des références dans plusieurs écrits plutôt récents qui recensent les connaissances robustes actuelles à propos de l'apprentissage de la lecture et des difficultés d'identification de mots; dans la synthèse de Snowling (2000) à propos de la dyslexie, dans le Bilan des données scientifiques sur la dyslexie produit par l'INSERM (Institut national de la santé et de la recherche médicale de la France) en 2007, ou encore dans une recension sur l'apprentissage de la lecture en français de Sprenger-Charolles (2016). De ce fait, il n'est pas surprenant de constater que plusieurs études du champ de la psychologie cognitive ou de la didactique cognitive de la lecture réfèrent à cette hypothèse de compensation aussitôt que cela est pertinent pour l'interprétation des résultats obtenus, et surtout lorsque les résultats mettent en évidence une compréhension peu altérée malgré des processus d'identification de mots déficitaires (p. ex. Braibant, 1994; Bruck, 1990; Elbro *et al.*, 1994; Sprenger-Charolles et Colé, 2013; Vellutino *et al.*, 2007; Yuill et Oakhill, 1991, pour ne nommer que ceux-ci). Avec la plus grande disponibilité des techniques d'imagerie cérébrale, certaines études réfèrent également à la compensation, mais à un niveau neuronal (plutôt que comportemental). Un nombre considérable d'études mettent en lumière certaines distinctions d'activation cérébrale chez des élèves qui ont des difficultés d'identification des mots écrits; puisque ces distinctions font appel à des régions homologues à celles utilisées en lecture par des lecteurs sans difficulté mais dans l'hémisphère droit plutôt que dans l'hémisphère gauche, les chercheurs évoquent souvent la possibilité que les régions de l'hémisphère droit correspondent à un circuit neuronal de la compensation (p. ex. Norton *et al.*, 2015; Richlan *et al.*, 2011; Shaywitz *et al.*, 2006). Toutefois, cela ne signifie pas qu'il s'agisse d'une compensation de nature sémantique. L'idée de compensation d'une difficulté d'identification de mots est effectivement omniprésente dans les recherches portant sur la lecture.

Ainsi, il apparaît que l'idée que le contexte est utilisé pour compenser une difficulté d'identification des mots écrits est amplement acceptée, et nommée explicitement par la communauté scientifique, aussi souvent que nécessaire. Par contre, très peu d'études se concentrent à documenter le fonctionnement

précis de ce mécanisme de compensation : en présence d'une difficulté d'identification des mots écrits, quelles informations du contexte sont utilisées exactement? Comment l'information provenant d'un autre niveau ou d'un autre processus arrive-t-elle à rendre l'identification de mots meilleure? À quel moment cela a lieu? De plus, les quelques études qui tentent de décrire ce mécanisme s'intéressent le plus souvent à des populations adultes. Ainsi, quelques recherches se centrent sur les adultes qui ont été identifiés « dyslexiques », mais qui arrivent pourtant à mener des études universitaires puisque leur compréhension écrite est peu affectée, population appelée « dyslexiques compensés » ou « lecteurs résilients » (Cavalli *et al.*, 2016; Horowitz-Kraus et Breznitz, 2012; Shaywitz *et al.*, 2003; Welcome *et al.*, 2009). La compensation chez l'enfant qui est en plein apprentissage de la lecture fait l'objet de très peu de recherches. Les quelques études qui portent précisément sur la compensation en lecture seront rapportées dans le cadre théorique, mais il convient d'indiquer ici que seules des données robustes indirectes à propos de l'influence de la sémantique dans la compensation sont présentes dans la littérature scientifique actuellement, c'est-à-dire des études qui relient la compensation à des éléments qui incluent la sémantique parmi d'autres concepts. Certaines données empiriques concernent notamment le vocabulaire (Cavalli et coll., 2016; van Viersen et coll., 2016), la morphologie (Elbro et Arnbak, 1996; Quémart et Casalis, 2015) mais surtout de bonnes capacités langagières générales (Bishop et Snowling, 2004; Bruck, 1992; Hulme et Snowling, 1992; Nation et Snowling, 1998; Snowling, 2000; van Viersen *et al.*, 2016).

Quoi qu'il en soit, la mise en œuvre efficace d'une telle compensation de nature sémantique en lecture peut effectivement empêcher les enseignants et les orthopédagogues de remarquer des difficultés d'identification de mots, puisque, lorsqu'elles sont bien compensées, elles ne gênent pas l'élève lorsqu'il doit comprendre le texte qu'il lit. Qui plus est, si cette compensation est mise en œuvre efficacement en situation de lecture de mots présentés isolément (sémantique lexicale), cela empêche également l'orthopédagogue de mesurer l'ampleur des difficultés, puisqu'elles peuvent être « non apparentes » même lors de la lecture de certaines catégories de mots en contexte d'évaluation orthopédagogique. Le dépistage de ces élèves qui compensent leurs difficultés à identifier les mots écrits représente ainsi un défi de taille pour les milieux scolaires. Sachant que, à plus ou moins longue échéance, ces élèves pourraient voir ces difficultés altérer la compréhension écrite et, par voie de conséquence, entraver leur performance académique, il convient de s'intéresser au phénomène et de le documenter davantage.

1.4 Étude de l'identification des mots écrits et la contribution possible de l'imagerie cérébrale

Puisque peu de résultats robustes concernant les élèves du niveau primaire sont disponibles actuellement, il convient d'aborder le phénomène de compensation d'abord au niveau du mot (plutôt qu'au niveau de la phrase ou du texte). Par ailleurs, la contribution de l'identification des mots écrits à la compréhension écrite est encore importante au primaire (Sprenger-Charolles et Colé, 2013). En dépit des nombreux chercheurs qui ont référé au recours à la sémantique pour compenser des difficultés d'identification de mots, peu d'études ont tenté, à notre connaissance, de documenter le fonctionnement du mécanisme de compensation par la sémantique en situation de lecture de mots présentés isolément chez des élèves en difficulté du primaire. Cela pourrait être attribuable, entre autres, à l'important défi méthodologique posé par l'étude de l'identification des mots écrits, qu'il convient de souligner ici. Lire un mot est une action qui se déroule extrêmement rapidement, souvent en moins d'un quart de seconde (Rayner, 1998), et qui est relativement automatique et non consciente (Laberge et Samuels, 1974). Un nombre important d'informations sont traitées dans ce court laps de temps : l'information visuelle qui correspond aux lettres du mot écrit, l'information phonologique qui correspond aux phonèmes du mot oral correspondant, et l'information sémantique qui correspond à la signification de ce mot, au(x) concept(s) qu'il désigne. Le traitement extrêmement rapide de ces informations complexes rend les différentes opérations cognitives de l'identification de mots extrêmement difficiles à distinguer les unes des autres. Suivant les principes de la chronométrie mentale (Ferrand, 2001 pour la lecture; Posner, 1978), la plupart des études s'inscrivant en sciences cognitives tiennent à la fois compte des erreurs commises pour une tâche donnée, ainsi que du temps de réaction ou temps de réponse. Le temps de réponse correspond à l'intervalle de temps entre le moment où un stimulus est présenté au participant et le moment où celui-ci fournit une réponse. Les principes de la chronométrie mentale indiquent donc que ce temps pris par le participant pour réaliser la tâche est un « indicateur précieux des processus mentaux […] impliqués dans la lecture (Ferrand, 2007, p. 29) ». Parmi les tâches les plus utilisées dans le domaine, recensées par Haberlandt (1994), se trouvent les tâches de décision lexicale et de décision sémantique, la première étant de loin la plus répandue. Dans une tâche de décision lexicale, on demande au participant d'indiquer si l'item présenté est un mot existant dans sa langue (item que l'on appelle *mot*) ou non (item que l'on appelle pseudomot), en appuyant sur des boutons. Tel que soulevé entre autres par Hauk *et al.* (2012), ce temps de réponse est difficile à interpréter, car il inclut plusieurs processus et mécanismes cognitifs. Certains d'entre eux sont directement reliés au construit à mesurer – ici l'identification de mots –, alors que d'autres sont plutôt reliés à la prise de décision ou encore à la programmation motrice de la réponse pour la pression du bouton, éléments qui ne font pas partie à proprement parler du phénomène d'intérêt. Les temps de réponse les plus rapides

mesurés par la pression d'un bouton sont mesurés autour des 500 millisecondes (ms; Ferrand, 2007), bien au-delà du temps réel pris par le participant pour « identifier » le mot.

Or, les récentes avancées en neurosciences relatives aux techniques d'imagerie cérébrale apportent une piste de solution pour distinguer plus clairement ce qui se passe pendant ces quelques millisecondes prises pour identifier un mot écrit. En effet, l'électroencéphalographie (EEG) est l'une des techniques d'imagerie ayant la meilleure résolution temporelle. Elle permet de mesurer l'activité électrique associée au fonctionnement psychophysiologique du cerveau à chaque milliseconde (Luck, 2005), permettant ainsi d'inférer ce qui se passe au niveau cognitif pendant l'identification d'un mot. La mise en relation de données comportementales telles que la qualité des réponses ainsi que les temps de réponse, et de leur(s) corrélat(s) psychophysiologique(s) au niveau cérébral, offre donc la possibilité d'observer finement le décours temporel des différentes opérations cognitives réalisées pendant l'identification de mots, à chaque milliseconde. Il est ainsi possible de mieux différencier chaque opération de l'identification de mots, des opérations liées à la prise de décision ou à la programmation motrice, en considérant l'activité électrique des premières millisecondes seulement. Un nombre important d'informations est ainsi disponible pour l'étude avant même que le sujet n'ait appuyé sur le bouton pour fournir sa réponse.

1.5 Objectif général de recherche et questions de recherche

Considérant l'état actuel des connaissances scientifiques au regard des lecteurs compensateurs ainsi que la situation qui prévaut actuellement dans le milieu scolaire au regard de ces élèves, il convient d'étudier ce groupe particulier d'élèves en difficulté de lecture. Ainsi, l'objectif général de la présente recherche est d'étudier la compensation en lecture opérée par une certaine proportion d'élèves qui ont une difficulté sévère d'identification des mots écrits sans que leur compréhension écrite ne soit drastiquement affectée.

Considérant le caractère un peu particulier de ce groupe d'élèves présentant des difficultés de lecture non manifestes, et considérant que peu d'études s'intéressent à ce phénomène chez des élèves du primaire, une première question de recherche est la suivante : Comment identifier les lecteurs compensateurs, dont les difficultés de lecture ne sont pas apparentes, au sein d'une population d'élèves de même niveau scolaire?

De plus, l'idée que certains lecteurs puissent recourir à la sémantique pour pallier leur difficulté à identifier les mots écrits, de même que les possibilités offertes par l'EEG pour désambiguïser les différentes

opérations cognitives de l'identification de mots, permettent de formuler la question suivante : le traitement sémantique des lecteurs compensateurs mesuré par le comportement et par l'EEG est-il mobilisé de manière particulière en identification de mots écrits présentés de manière isolée?

1.6 Pertinence de la recherche et retombées anticipées

La pertinence et les retombées anticipées de la présente étude concernent tant l'avancement des connaissances scientifiques que les pratiques en milieu scolaire.

1.6.1 Pertinence sociale

La pertinence sociale de cette recherche a été annoncée à de multiples occasions au cours de cette problématique, puisque, effectivement, la présente recherche prend tout son sens du fait qu'elle est pertinente pour les milieux scolaires et pour l'orthopédagogie. Les principaux éléments sont rapportés ici.

D'une part, les coûts économiques et sociaux associés aux difficultés d'apprentissage de la lecture sont tels qu'il convient de s'intéresser à ce phénomène et de documenter différentes manières de repérer rapidement les élèves qui ont des difficultés. Cette recherche permettra de repérer plus rapidement cette catégorie d'élèves qui compensent, qui, autrement, ne seraient repérés que tardivement et verraient en quelque sorte leurs difficultés d'aggraver. Aussi, en lien avec les milieux de pratique, différentes retombées peuvent être envisagées à propos de la meilleure compréhension du fonctionnement du mécanisme de compensation des difficultés d'identification des mots écrits. D'une part, tel que rapporté en détails à la section 1.2, sachant que ce type d'élèves arrive parfois à faire en sorte que les difficultés ne soient pas apparentes, une meilleure compréhension de ce mécanisme permettra de raffiner les procédures d'évaluation, de dépistage en lecture. Des propositions de changement de pratique peuvent ainsi découler des résultats obtenus et contribuer à un meilleur accompagnement de ces lecteurs qui compensent. En effet, ces propositions permettraient tant aux enseignantes qu'aux orthopédagogues de trouver des moyens de repérer des difficultés non apparentes dans l'actualisation de la compétence à lire des élèves. Une fois ces élèves repérés, un suivi pourra être complété afin d'assurer que l'écart avec les pairs concernant l'identification de mots ne se creuse pas, à condition qu'il y ait des mesures appropriées, au bon moment pour les soutenir. Une meilleure compréhension de la compensation en lecture pourra également fournir des pistes pour orienter mieux les actions à poser pour ces élèves; élaboration d'un plan d'intervention, monitorage fréquent de la performance en compréhension écrite pour assurer qu'elle est

toujours intacte, etc. seront autant de moyens pouvant être suggérés en priorité, selon les résultats obtenus dans la présente recherche.

Dans le cadre actuel, il ne serait pas prudent d'envisager des retombées sur les interventions particulières à réaliser auprès des lecteurs compensateurs. Étant donné que le phénomène est peu documenté, il semble plus rigoureux de mieux comprendre le mécanisme de compensation d'abord, ce qui a des incidences sur l'évaluation orthopédagogique de la lecture chez ces lecteurs. Lorsque le phénomène sera mieux compris, c'est-à-dire lorsque que l'on comprendra comment la compensation est mise en œuvre et dans quelle dynamique, et que l'on aura une meilleure idée de comment la faire ressortir en situation d'évaluation, des hypothèses quant aux interventions à mettre de l'avant pourront être formulées. Ces hypothèses devront ensuite faire l'objet elles aussi d'études empiriques rigoureuses avant de pouvoir être proposées aux milieux scolaires. Il semble ainsi plus prudent d'envisager actuellement des retombées principalement sur l'évaluation de la lecture.

1.6.2 Pertinence scientifique

D'abord, dans les domaines concernés par l'identification des mots écrits, comme dans la majorité des autres domaines, la plupart des études publiées sont réalisées par des locuteurs de l'anglais. Alors que certains phénomènes ne sont pas dépendants de la langue (p. ex. l'environnement ou la physique), la lecture dépend beaucoup des caractéristiques de la langue écrite (voir le cadre théorique pour davantage de détails). La plupart des études publiées à propos de la lecture concernent donc la langue anglaise. Comme le système écrit de l'anglais n'est pas identique à celui du français, il est primordial de documenter les phénomènes d'intérêt en français également. Les données de cette thèse contribueront ainsi à l'avancement des connaissances scientifiques à propos de cette langue spécifiquement. Plus encore, cette recherche vise à informer la compensation en lecture, soit un phénomène souvent évoqué dans la littérature scientifique à propos des difficultés de lecture, mais qui a peu fait l'objet de recherches en soi. Ainsi, le fonctionnement de ce mécanisme de compensation est peu documenté actuellement bien que plusieurs données aient mis en lumière l'existence de cette compensation et que plusieurs chercheurs recourent à cette idée pour expliquer les données obtenues chez certains élèves en difficulté de lecture. En se centrant principalement sur le recours à la sémantique lexicale pour compenser un déficit d'identification des mots écrits, la présente étude documente en partie ce mécanisme de compensation, et apporte ainsi des informations nouvelles et importantes pour le domaine.

De plus, alors que beaucoup d'études sur l'identification des mots écrits sont produites dans le champ de la psychologie, de la psycholinguistique, ou encore de la neuropsychologie, la présente étude relève davantage d'une perspective éducative. Cette perspective éducative se manifeste par différents éléments, qui font en sorte que la présente étude apporte un éclairage distinct des études menées dans d'autres domaines. Les études en neuropsychologie choisissent souvent les participants de leur étude parce qu'ils représentent des cas extrêmes ou une pathologie (Baars et Gage, 2013; Ward, 2015). Dans le cas de la lecture, ces études se centrent donc très souvent sur le trouble spécifique d'apprentissage de la lecture uniquement (Goswami, 2008). En psychologie, les populations d'élèves en difficulté sont souvent sélectionnées selon des critères de performance sur une seule tâche donnée. La population ciblée par la présente recherche se veut plus représentative des élèves en difficulté rencontrés dans les milieux scolaires; les orthopédagogues évaluent et interviennent auprès de tous les élèves qui présentent des difficultés de lecture, et la nécessité de mettre en place des interventions adaptées aux besoins des apprenants fait en sorte que la nature de ces besoins est toujours prise en considération.

La perspective éducative de cette thèse se manifeste également dans la manière prudente et réfléchie avec laquelle les mesures issues des neurosciences sont intégrées à une recherche en éducation. L'introduction de l'EEG est justifiée notamment par la possibilité qu'offre cette technique d'imagerie cérébrale d'aller au-delà des comportements observables par la lecture orale ou les tâches traditionnelles de décision lexicale. Il y a recours à des données similaires à celles utilisées en neurosciences, c'est-à-dire l'activité électrique émanant du cerveau, mais ce n'est pas parce que ce type de données est issu d'un domaine qui se rapproche davantage de la santé, du domaine médical, que ces domaines s'introduisent en éducation ici. En effet, l'ensemble de la démarche de la présente recherche est fondé sur l'état de la recherche en éducation et vise à documenter ce champ à l'aide de techniques empruntées aux sciences cognitives et aux neurosciences (alors que l'objectif des neurosciences est de documenter le fonctionnement du cerveau; Baars et Gage, 2010).

Ce recours à des techniques d'imagerie cérébrale en éducation est relativement nouveau et s'inscrit dans une discipline, appelée les neurosciences éducationnelles. Sa relative récence explique que plusieurs visions de ce que devrait être ce champ de recherche coexistent actuellement. Plusieurs questions sont posées sur la définition du champ ou sur des exemples de travaux qui respecteraient les principes souhaités pour le champ. Ceci est manifeste dans les nombreux congrès dédiés, les séminaires ou les tables de discussion internationales organisés sur le sujet. La démarche réalisée dans le cadre de cette thèse pour

positionner les concepts et la pertinence de recourir à l'EEG alimente ainsi toutes ces questions. Principalement, le fait que la présente recherche considère l'état des connaissances en éducation d'abord, puis en psycholinguistique cognitive ensuite, pour compléter par les données issues de l'EEG souligne un aspect important que devraient respecter toutes recherches en neurosciences éducationnelles, c'est-à-dire la considération conjointe de tout ce que l'on sait actuellement dans les écrits scientifiques sur le sujet d'intérêt pour mettre en relation les comportements observables, qui attestent du fonctionnement cognitif impliqué dans une activité scolaire comme la lecture, à ses corrélats psychophysiologiques. Arriver à réaliser cela consiste en un apport nouveau et original pour les sciences de l'éducation.

CHAPITRE 2

CADRE THÉORIQUE

Afin de répondre à la première question de recherche, il convient de présenter dans ce chapitre la littérature scientifique pertinente. La première section du cadre théorique vise à décrire brièvement les différentes disciplines auxquelles cette thèse fait appel, de manière à bien situer les écrits scientifiques dont il sera question, mais surtout à clarifier l'importance de la transdisciplinarité pour traiter du sujet. La deuxième section pose les bases théoriques nécessaires à la thèse, notamment à propos de la lecture ainsi que de la place de l'identification des mots écrits dans cette compétence complexe. Les troisième, quatrième et cinquième sections contiennent les différents modèles théoriques pertinents à la réalisation de la présente étude, ainsi que leur critique en lien avec les travaux menés. La section suivante rapporte les éléments pertinents d'électroencéphalographie ainsi que les données empiriques en lien avec les difficultés de lecture, et la compensation en lecture. Finalement, la dernière partie discute des données liées à l'amorçage sémantique et à l'électrophysiologie. La considération de l'ensemble de ces éléments permet ensuite de formuler, à la dernière section de ce chapitre, des hypothèses de recherche.

2.1 Caractère transdisciplinaire de la thèse au sein des sciences de l'éducation

Il convient d'expliciter que cette thèse, tout comme le sont par définition les sciences de l'éducation (Mialaret, 2008), mobilise des connaissances issues de plusieurs disciplines scientifiques. Ces disciplines sont brièvement discutées ici.

D'abord, cette recherche s'inscrit dans la didactique disciplinaire, l'une des sciences de l'éducation. La didactique disciplinaire est reconnue comme étant un champ de recherche qui s'intéresse à l'appropriation d'un contenu disciplinaire précis par un apprenant dans un contexte scolaire ou de formation (Caillot, 1992). La didactique d'une discipline implique la considération de trois pôles, conceptualisés dans le « triangle didactique » : 1) le pôle « apprenant »; 2) le pôle « enseignant »; et 3) le pôle « objet », c'est-à-dire l'objet de la discipline visée (Reuter, 2013). Le didacticien disciplinaire étudie donc la manière selon laquelle ces trois pôles interagissent pour induire l'apprentissage d'un contenu disciplinaire chez l'élève. L'intérêt du didacticien disciplinaire peut être centré particulièrement sur un ou des pôles, mais on considère toujours l'ensemble du triangle didactique. Puisque cette recherche s'intéresse à la lecture, elle s'inscrit dans la didactique de la lecture. Plus particulièrement, comme il le sera décrit ci-contre, deux des pôles feront partie intégrante du devis de recherche, alors que le troisième

sera considéré dans la sélection des participants et sera directement touché par les résultats de recherche, en tant que retombées envisagées.

Ainsi, puisque la didactique disciplinaire tient compte de l'« objet » à apprendre, il est systématiquement nécessaire de considérer d'autres domaines de recherche pour décrire précisément cet objet. Par exemple, la didactique des mathématiques fait appel aux mathématiques; la didactique des sciences fait appel entre autres à la chimie, la physique, la biologie, etc. En ce qui concerne cette thèse, la didactique de la lecture fait nécessairement appel aux connaissances produites par la linguistique. La linguistique a pour objectif de décrire le langage et les systèmes régissant notamment les langues naturelles (Ducrot, 1972). Les connaissances produites en linguistique à propos des systèmes phonétique, phonologique, morphologique, sémantique et pragmatique des langues naturelles sont ainsi incontournables pour une considération judicieuse de l'objet à apprendre. De plus, les connaissances issues de la psycholinguistique sont primordiales pour préciser l'interaction entre le pôle « apprenant » et le pôle « objet », puisqu'elles ont pour objectif de décrire la manière dont un individu – « psycho » – traite la langue – « -linguistique » (Ying, 2013; 2016). La présente étude met effectivement l'accent sur cette interaction entre « apprenant » et « objet ».

Il convient également de noter que la présente recherche s'inscrit dans la grande famille des sciences cognitives. L'objectif des sciences cognitives est de comprendre comment fonctionne la cognition, c'est-à-dire comprendre comment fonctionnent les processus mentaux qui permettent à l'humain d'appréhender le monde qui l'entoure et d'interagir avec celui-ci. Les sciences cognitives contribuent ainsi à décrire précisément les mécanismes cognitifs impliqués dans les différentes habiletés que l'humain possède (Collins *et al.*, 2018; Bermudez, 2020). Elles sont définies par un ensemble de « disciplines contributives » (*Cognitive Science Society*), parmi lesquelles on retrouve notamment la linguistique, la psychologie, les neurosciences et l'éducation. Chacune de ces disciplines, lorsqu'elle contribue à décrire les processus mentaux ou les mécanismes cognitifs, voit son nom adjoint du qualificatif « cognitif » : « psychologie cognitive », « linguistique cognitive » et « neurosciences cognitives » sont effectivement des termes désormais utilisés pour qualifier les travaux de ces disciplines qui s'intéressent particulièrement à ces processus ou mécanismes. La didactique cognitive d'une langue est ainsi définie par Ying (2016) comme étant une discipline qui doit prendre en considération toutes les opérations mentales de l'apprenant et « chercher à faciliter leur accomplissement dans la conception d'activités pédagogiques (p. 232) ». Discutée en référence au triangle didactique, cette discipline accorde donc une importance particulière à

la dimension cognitive des élèves, dans le pôle « apprenant ». Cette dimension cognitive de l'apprenant est considérée selon l'objet à apprendre, donc elle implique particulièrement l'interaction entre les pôles « apprenant » et « objet ». La didactique cognitive prend ainsi acte des travaux en sciences cognitives et y contribue, plus particulièrement en ce qui concerne la psycholinguistique cognitive (Morais *et al.*, 2003). Finalement, l'ajout du préfixe « ortho- » à l'expression « didactique cognitive » désigne ainsi une sous-catégorie particulière de didactique cognitive, celle qui se centre sur les élèves qui présentent, ou qui sont à risque de présenter, une difficulté dans la discipline dont il est question. L'orthodidactique cognitive s'intéresse donc également à l'apprentissage comme résultant de l'interaction des 3 pôles précédemment mentionnés, avec les particularités suivantes : 1) le pôle « apprenant » cible particulièrement ceux qui présentent une difficulté ou sont à risque d'en présenter; 2) le pôle « enseignant » fait référence particulièrement à l'orthopédagogue, qui est le professionnel qui œuvre auprès des élèves en difficulté, tant dans l'évaluation que dans l'intervention, toutes deux relevant ainsi de l'orthodidactique.

Par ailleurs, la présente thèse implique également les neurosciences, plus particulièrement les neurosciences cognitives qui s'intéressent à l'étude des mécanismes neurologiques qui sous-tendent la cognition (Ward, 2015). En effet, l'électroencéphalographie (EEG) sera utilisée pour décrire le décours temporel de différents mécanismes cognitifs impliqués dans l'identification de mots. Afin d'être en mesure d'interpréter adéquatement ce type de résultats, il est nécessaire de faire appel à différents niveaux d'analyse et de lier ceux-ci entre eux. Ainsi, dans le présent cadre théorique, on conçoit que tout comportement humain est traduit de façon concomitante par plusieurs systèmes hiérarchiques, au sens de la proposition de théories unifiées de la cognition de Newell (1990). Ainsi, les comportements observables sont mis en relation avec leur corrélat au niveau cognitif (ensemble de processus mentaux qui rendent possible la cognition) ainsi qu'au niveau psychophysiologique (activation simultanée de neurones pour réaliser une fonction donnée). Cette vision « multicouche » du comportement (Mercier et Charland, 2013) fournit un cadre structurant assurant la rigueur de l'interprétation des résultats obtenus par l'imagerie cérébrale en identification de mots. Puisque ces différents niveaux sont étudiés chez des élèves qui sont en train d'apprendre la lecture en contexte scolaire, on peut indiquer que cette recherche s'inscrit dans les neurosciences éducationnelles. Ce champ, bien qu'il ne soit pas totalement nouveau (Mercier *et al.*, 2016), est encore considéré en émergence, notamment parce que plusieurs interlocuteurs discutent actuellement la définition du champ (p. ex. Bowers, 2016; Howard-Jones, 2014) Résumer les débats entourant la définition et l'utilité de ce champ pour l'éducation ne constitue pas un objectif réaliste ici, mais il convient toutefois de préciser que la présente thèse s'inscrit dans la vision selon laquelle les

neurosciences éducationnelles impliquent nécessairement la considération concomitante de plusieurs niveaux d'analyse afin d'étudier un phénomène lié à l'apprentissage. Mais parce qu'elles sont éducationnelles, ces neurosciences cognitives doivent être rigoureusement appliquées non pas dans le but d'informer la manière dont le cerveau fonctionne – ce qui est le propre des neurosciences –, ni d'informer sur les processus cognitifs eux-mêmes – ce qui est le propre de la psychologie cognitive –, mais bien dans le but d'informer la manière dont un individu apprend en contexte de formation.

Bref, ce sont toutes ces disciplines qui sont impliquées dans le cadre théorique qui suit. Celles-ci ne peuvent par contre pas être considérées « en silo » ; elles doivent plutôt être complètement intégrées pour une meilleure compréhension du phénomène étudié. Certains utilisent l'adjectif « interdisciplinaire » pour insister sur les connexions entre des modèles de recherche de différentes disciplines ainsi que des cadres de collaboration entre les gens appartenant à ces disciplines diverses (Howard-Jones, 2011; Howard-Jones et Fenton, 2012). Or, ici, il s'agit d'une réelle tentative de transdisciplinarité qui est menée à travers les travaux de la présente thèse. La transdisciplinarité représente ainsi un effort encore plus grand d'intégration des différentes disciplines. Plutôt que d'adjoindre les disciplines les unes aux autres dans une perspective de complémentarité comme c'est le cas dans l'interdisciplinarité, la transdisciplinarité est caractérisée par la transcendance des barrières traditionnelles des disciplines (Beauchamp et Beauchamp, 2013; Coch et Ansari, 2009), ceci permettant l'intégration en un seul tout cohérent de modèles théoriques communs et de méthodes communes (Knox, 2016).

2.2 Processus cognitifs de lecture : de l'identification des mots écrits à la compréhension

Puisque le présent projet s'intéresse aux difficultés d'identification des mots écrits ainsi qu'à une façon de les surmonter, c'est-à-dire la compensation, les prochaines sections visent à décrire les concepts pertinents à l'activité cognitive complexe qu'est la lecture.

Dans la perspective des sciences cognitives, on décrit la lecture comme étant l'une des habiletés les plus complexes de l'esprit humain. De nombreux auteurs, et ce depuis les premières recherches en psychologie sur le sujet, soulignent qu'une théorie de la lecture implique nécessairement la description d'un très grand nombre de fonctionnements extrêmement complexes de la cognition humaine (p. ex. Huey, 1908; Perfetti, 1985; Rumelhart, 1994), dont certains sont spécifiques aux activités langagières, alors que d'autres sont généraux et sont donc transversaux à différentes activités cognitives (Fodor, 1983). Il n'est alors pas

surprenant de constater un foisonnement de recherches et d'écrits scientifiques qui tentent de décrire cette habileté toute particulière à l'humain.

La lecture est une activité de communication; elle sert à accéder à un message transmis par le biais de la langue écrite (Milne, 2005). Elle représente donc une activité langagière, à l'instar de la communication orale (Riley et Riley, 1965). Pour le récepteur, accéder à un message transmis par le biais d'une langue orale nécessite la mise en œuvre d'une séquence d'opérations mentales sur les structures linguistiques de cette langue, permettant de se construire une représentation de la signification du message (Perfetti, 1985). Cette représentation interne, chez le récepteur, correspond ultimement à la compréhension du message véhiculé oralement par l'émetteur. En ce qui concerne la lecture, on considère qu'elle représente le volet réceptif de la communication par le biais de la langue écrite. Traditionnellement, on conçoit qu'elle débute par des informations visuelles atteignant la rétine de l'œil et culmine par une idée définie à propos du message que son auteur désire transmettre par le texte écrit (Rumelhart, 1994). De la même façon que lorsque le récepteur entend un message oral, la finalité de la lecture est d'arriver à une compréhension du message véhiculé, ici par le biais de la langue écrite.

La nature linguistique de la lecture se traduit par plusieurs caractéristiques partagées avec la langue orale, notamment concernant ces opérations mentales à mettre en œuvre sur les structures linguistiques d'une langue dans un but de compréhension (Perfetti, 1985). Par exemple, le processus d'analyse syntaxique, qui permet de traiter l'ordre des mots ainsi que leur classe, doit être mis en œuvre tant lors de la lecture d'un texte que lors de l'écoute d'un message oral (voir la section 2.3 pour davantage de détails sur ces processus communs à l'écrit et à l'oral). Cependant, comme le souligne Perfetti (1985), certaines différences existent entre la compréhension d'un message oral et la compréhension d'un message écrit; notamment le fait que le message écrit est fixe et permanent, alors que le message oral est éphémère et peut être modulé au besoin par les participants. Il attribue ces différences à deux traits distincts fondamentaux : 1) la nature du signal : le signal de l'écrit est visuo-spatial, alors que le signal de l'oral est temporo-auditif; et 2) le « design » social du contexte du message : le contexte de la communication orale est socialement interactif et pragmatique, alors que le contexte de la communication écrite renvoie à une activité le plus souvent individuelle qui n'est pas généralement pas interactive (p.6). Ainsi, les conventions qui régissent le code des messages oraux et le code des messages écrits comportent à la fois des similitudes et des distinctions. Dans le message oral, l'information est transmise de manière orale par la parole. Cette dernière inclut des phonèmes, c'est-à-dire une représentation mentale des sons d'une langue orale

donnée, qui sont prononcés les uns à la suite des autres. Un phonème ou un groupe de phonèmes forment des mots qui sont identifiés par celui qui écoute le message oral. Saussure (1916) appelle « signifiants » la forme acoustique des mots, et indique que celle-ci réfère à des concepts, appelés « signifiés », c'est-à-dire la représentation mentale de la « chose » désignée. Le code oral recourt donc à des phonèmes, qui, lorsqu'ils sont regroupés, sont associés à un concept. Les mots ainsi produits sont agencés d'une telle façon qu'ils puissent formuler des idées. Ces idées sont appelées unités de sens, et sont transmises par des propositions (voir la section 2.3 pour davantage de détails). Ces propositions sont traitées par le système cognitif de la personne qui écoute le message. Cette personne s'en fait une représentation interne. Ultimement, ces propositions sont également mises ensemble pour former un discours. Il y a donc prise en compte d'unités de sens de plus en plus larges. Toutes ces unités de sens se retrouvent ainsi dans la représentation mentale de la personne qui reçoit le message oral. Perfetti (1985), notamment, rapporte que le code lié à la langue orale correspond à un système « naturel », « facilement maîtrisé par un enfant qui vit en communauté » (p. 6). Certaines structures cognitives *innées* semblent déjà en mesure de prendre en charge la langue orale chez l'humain, et faire en sorte par exemple que les enfants en bas âge semblent acquérir cette capacité sans un quelconque enseignement (mais non sans effort), tout comme c'est le cas de la marche et de la respiration (Milne, 2005).

Le code écrit, pour sa part, transmet l'information langagière par l'entremise de symboles écrits qui sont donc perçus par le récepteur, c'est-à-dire le lecteur, par la modalité visuelle (Rayner, 1998). Dans un système d'écriture alphabétique, comme le français ou l'anglais, les symboles écrits, représentés par les lettres de l'alphabet latin, renvoient majoritairement aux unités orales que sont les phonèmes (Duncan et Seymour, 2000 entre autres; voir section 2.4.1 pour davantage de détails). Ainsi, un ensemble de lettres – ou parfois une seule – constitue alors un mot écrit, qui renvoie à un mot oral (sa forme phonologique). À partir de cet accès à la forme orale du mot, le fonctionnement qui lie le « signifié » au « signifiant » s'applique un peu de la même manière que ce qui caractérise le traitement du code oral (certaines différences demeurent, tel que mentionné précédemment). En termes d'évolution, l'utilisation d'un code écrit par les humains pour communiquer est relativement récente dans l'histoire en comparaison à la capacité à utiliser un code oral (Ducrot, 1972). Cela expliquerait pourquoi les structures internes de l'humain n'ont pas encore eu le temps de s'adapter et faire en sorte d'acquérir ces compétences naturellement et aisément comme c'est le cas pour la langue orale (Pinker, 1994). Le lecteur débutant doit donc apprendre, souvent en contexte scolaire, à maîtriser ces conventions qui font correspondre le

système écrit – les symboles écrits que sont les lettres – à un système linguistique plus large qu'il maîtrise déjà – le système linguistique qu'il mobilise à l'oral.

Pour résumer, cette description sommaire des composantes langagières de la communication orale et de la communication écrite souligne donc le fait que la lecture est une activité langagière d'abord, visant la compréhension d'un message transmis par le biais d'un code conventionnel écrit. Ce faisant, la lecture partage certaines « opérations mentales » avec la compréhension à l'oral. Par ailleurs, notamment à cause de la spécificité de son code, la lecture implique également certaines « opérations mentales » qui lui sont propres.

En sciences cognitives, les auteurs qui tentent de décrire, dans des modèles, ces « opérations mentales » réfèrent à ces dernières en termes de processus cognitifs de la lecture. Ils distinguent ainsi les processus cognitifs qui sont spécifiques à la lecture, de ceux qui sont également mobilisés dans le traitement de la langue orale. On les nomme respectivement les processus spécifiques et les processus non spécifiques, de lecture (Gough et Tunmer, 1986, Hoover et Gough, 1990, Braibant, 1994).

Les processus spécifiques de lecture sont définis comme étant ceux qui rendent possible l'identification des mots écrits à partir de l'analyse des traits visuels du matériel à lire (p. ex. Alegria, 1988; Ferrand et Grainger, 2004; Stanovich, 1988, etc.) Ainsi, l'expression « processus d'identification des mots écrits » renvoie à l'ensemble des opérations mentales réalisées par le lecteur sur un « objet visuel », les lettres constituant le mot, afin de lier sa forme physique imprimée sur la page à une représentation visuelle abstraite de cette forme stockée en mémoire. Cette opération de correspondance entre le matériel visuel et la représentation interne de ce mot permet ultimement d'associer un sens à ce mot présenté visuellement (voir Ferrand, 2007, pour un résumé). Les processus non spécifiques de lecture renvoient quant à eux à l'ensemble des opérations mentales qui permettent l'intégration syntaxique et sémantique menant à la compréhension d'unités linguistiques plus larges que le mot; c'est-à-dire la phrase, le paragraphe, le texte (Content, 1991; Perfetti, 1985).

Ainsi, Gough et Tunmer (1986) ont proposé de définir la lecture par une équation, qu'ils ont nommée la *Simple View of Reading*. Cette équation est la suivante : Lecture = décodage X compréhension.

Ici, l'expression « lecture » renvoie à la compréhension de ce qui est lu, l'expression « décodage »[2] renvoie à l'ensemble des procédures qui permettent d'identifier un mot écrit, et l'expression « compréhension » renvoie aux processus de compréhension orale, c'est-à-dire les processus non spécifiques. L'équation indique ainsi que la compréhension en lecture est le produit de la mise en œuvre des processus spécifiques d'identification des mots écrits et des processus non spécifiques tels qu'ils sont mis en œuvre pour la compréhension orale. Étant le produit de ces deux composantes indépendantes plutôt que la somme, les auteurs insistent sur le fait que la lecture ne peut exister si l'une des deux composantes est absente chez un individu. Ainsi, un lecteur qui ne fait qu'identifier les mots écrits sans comprendre ce qu'il lit peut prononcer les mots à haute voix, mais on n'affirmera pas qu'il soit en mesure de « lire ». À l'inverse, on ne peut dire qu'un individu qui ne possède que des capacités linguistiques à l'oral ne soit en mesure de lire : un individu qui comprend oralement mais qui n'identifie pas les mots écrits n'est pas en mesure de « lire ». Les deux catégories de processus sont donc considérées comme étant nécessaires à la lecture.

Cette équation demeure un modèle très général, ne se positionnant pas entre autres quant à la séquence de mise en œuvre de ces catégories de processus cognitifs (voir Kirby et Savage, 2008 et Ouellette et Beers, 2010 pour des critiques plus complètes). Ce sont les modèles de lecture dits « interactifs » qui s'inscrivent dans cette même conception de l'importance relative de chacune des catégories de processus, en plus de théoriser la séquence de mise en œuvre de ces derniers. Le principe à la base des modèles interactifs est de considérer que l'ensemble des processus cognitifs de lecture peuvent être mis en œuvre tant de manière ascendante que descendante (Mosenthal *et al.*, 1978; Stanovich, 1980).

2.3 Modèles théoriques interactifs des processus cognitifs de lecture

En sciences cognitives, plusieurs auteurs tentent de faire l'inventaire de ces processus spécifiques et non spécifiques de lecture, et de décrire précisément comment ils fonctionnent afin de permettre la compréhension du texte lu. Deux de ces modèles interactifs sont décrits ici.

[2] Il convient ici de noter que l'expression « décodage » a été remise en question par plusieurs auteurs (p. ex. Stuart et Stainhorp, 2008), puisque généralement elle ne réfère qu'à un seul mécanisme d'identification des mots écrits, alors que Gough et Tunmer (1986) semblent désigner davantage l'ensemble des mécanismes d'identification des mots écrits.

2.3.1 Modèle de construction-intégration de van Dijk et Kintsch (1983)

Un des modèles interactifs qui demeure des plus influents à ce jour est celui nommé construction-intégration élaboré par Kintsch (1988), à partir de l'ouvrage de van Dijk et Kintsch (1983). Kintsch et Rawson (2005) amènent subséquemment certaines précisions et ajouts à ce modèle, notamment concernant les processus non spécifiques. Ce modèle a été élaboré pour rendre compte de la lecture compétente.

Le modèle de construction-intégration propose qu'un lecteur se construit une représentation interne du texte lu, en se servant à la fois des informations contenues directement dans le texte ainsi que d'informations qui sont déjà emmagasinées dans sa mémoire, ses connaissances (van Dijk et Kintsch, 1983). Pour rendre compte de tout le travail cognitif à effectuer sur ces informations, les auteurs proposent que la compréhension repose sur trois niveaux de traitement. Le premier niveau est appelé le « niveau linguistique », ou le traitement de la forme du texte. Ce niveau concerne le traitement de chaque mot pris individuellement, puis le traitement syntaxique. Le deuxième niveau est appelé le « niveau conceptuel ». Il concerne tout traitement sémantique (portant sur le sens), qui vise à dégager la signification du texte. Ainsi, les significations associées aux mots sont combinées pour former des unités de signification plus larges qu'un seul mot, appelées propositions. Les propositions sont reliées entre elles, par différents moyens et différentes opérations mentales, pour former un réseau sémantique complexe. Ces propositions sont ensuite regroupées pour mener à la construction de la structure globale du texte. Le troisième niveau correspond à une compréhension approfondie d'un texte lu (Kintsch et Rawson, 2005), dans le sens où le contenu qui a été lu est finalement intégré aux connaissances et expériences préalables; c'est le « niveau intégration » (Kintsch, 1988). Pour chacun de ces trois niveaux (linguistique, conceptuel et d'intégration), les auteurs renvoient à un certain nombre de processus cognitifs précis. Sans entrer dans le détail, il semble pertinent de mentionner que, pour le niveau linguistique, les processus de « décodage » des mots écrits (processus spécifiques) ainsi que les processus d'analyse syntaxique (processus non spécifiques) sont nommés. Pour les niveaux conceptuels et d'intégration, les micro- et macroprocessus sont nommés, ainsi que la production d'inférences et l'élaboration du modèle de situation (bref, tous des processus cognitifs non spécifiques; consulter Kintsh et Rawson, 2005 et Laplante, 2011 pour des détails sur ces processus).

2.3.2 Cadre des systèmes de la lecture de Perfetti (1999) et de Perfetti et Stafura (2014)

Dès 1985 et 1988, Perfetti propose une théorie générale pour rendre compte de la compréhension en lecture qu'il appelle la théorie de l'efficience verbale. Cette théorie repose d'abord sur les contraintes cognitives s'appliquant à la lecture. En ce sens, la théorie de l'efficience verbale pose deux postulats de base concernant la manière dont les processus cognitifs sont mis en œuvre. Le premier postulat est celui selon lequel l'attention et la mémoire de travail, qui sont toutes deux sollicitées par les processus de lecture, ont des ressources limitées. Le lecteur doit composer avec une quantité limitée de ressources cognitives à dédier à l'ensemble des processus cognitifs de lecture. Ainsi, plus un processus est automatisé, moins il mobilise de ressources cognitives (attention et mémoire de travail). Ces ressources cognitives non utilisées sont alors disponibles pour être dédiées à d'autres processus, souvent les processus de plus haut niveau (p. ex. la production d'inférences ou l'élaboration du modèle de situation). On dira du lecteur pour qui plusieurs processus sont automatisés qu'il est « verbalement efficace ». Le deuxième postulat prévoit que la transmission de l'information se fait d'un processus à l'autre, le plus rapidement et le plus justement possible, de manière ascendante. Ainsi, un premier processus opère un traitement sur l'information reçue, puis envoie ces informations traitées au processus suivant, qui, à son tour, opère un traitement sur l'information reçue, et ainsi de suite. On propose que ce soit selon cette séquence ascendante que le lecteur est le plus efficace, mais on prévoit également que, dans certaines circonstances, par exemple lorsqu'un mot est inconnu ou difficilement identifiable par le lecteur, celui-ci a la possibilité de procéder à un traitement descendant.

La théorie de l'efficience verbale précise également ce qu'il en est des difficultés de lecture. Un processus déficitaire se traduit ainsi par deux éléments : 1) les informations transmises par ce processus déficitaire aux processus suivants sont inadéquates ou imprécises (les processus de lecture suivants doivent donc composer avec ces informations partielles ou incorrectes); 2) puisqu'il est déficitaire, ce processus n'est, par définition, pas automatisé, donc il prend plus de temps à opérer et nécessite le déploiement d'une quantité importante de ressources cognitives (Perfetti, 1985).

S'inscrivant dans cette théorie de l'efficience verbale, Perfetti (1999) et Perfetti et Stafura (2014) proposent le cadre des systèmes de la lecture (*Reading Systems Framework*), présenté à la figure 2.

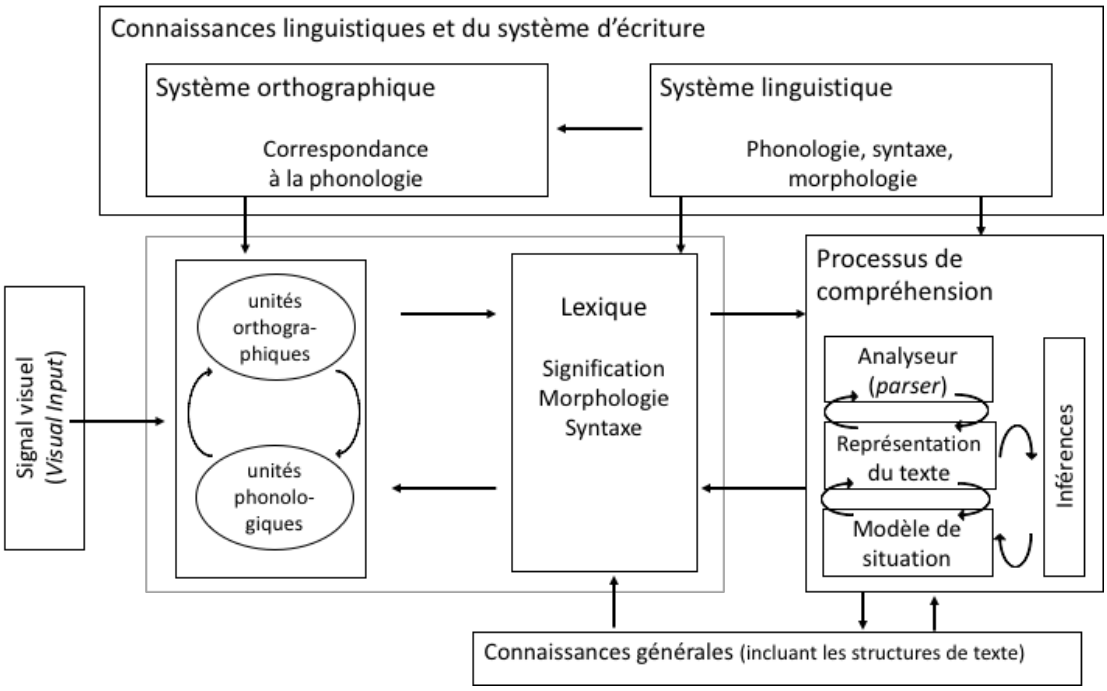


Figure 2.1 Cadre des systèmes de la lecture (Perfetti et Stafura, 2014). Traduction libre

On retrouve ainsi les mêmes processus que ceux identifiés par le modèle de construction-intégration (van Dijk et Kintsch, 1983), notamment dans l'encadré de droite indiquant les processus de compréhension, c'est-à-dire les processus non spécifiques. Ces auteurs marquent cependant une distinction entre l'identification des mots écrits et la compréhension de mots (Perfetti et Stafura, 2014). On insiste ainsi sur le fait que la compréhension de mots est permise par l'identification des mots. Sans entrer dans le détail ici, les auteurs précisent que la compréhension d'un mot lu est le produit de l'identification de mots écrits. Une fois le mot identifié, le lecteur accède donc à son sens (sa signification). Le cadre des systèmes de lecture accorde une grande importance à cette compréhension des mots isolés, puisqu'elle constitue la jonction entre les processus spécifiques et non spécifiques de lecture. Encore une fois, les précisions et davantage de détails concernant ces processus spécifiques seront traités à la section 2.4.

À la différence d'autres modèles, ce cadre rend explicite le rôle de différentes connaissances et leur influence dans la lecture. Le modèle prévoit donc trois catégories de connaissances qui seront mobilisées par les processus de lecture (Perfetti, 1999) : 1) les connaissances linguistiques : connaissances phonologiques, syntaxiques et morphologiques pour l'écrit et l'oral; 2) les connaissances orthographiques :

connaissances à propos du système qui régit les liens entre forme écrite et forme orale (davantage de précisions seront trouvées à la section 2.4); 3) les connaissances générales : toutes les connaissances sur le monde, sur les types de texte, sur des expériences de lecture précédentes, etc. Ces catégories de connaissances contribuent distinctivement à différents processus de lecture tel qu'indiqué par les flèches dans le schéma. Ainsi, par exemple, les connaissances orthographiques sont utilisées exclusivement par les processus d'identification de mots, alors que les connaissances générales sont utilisées pour l'ensemble des processus non spécifiques. Conséquemment au fait que les auteurs explicitent la place des connaissances dans leur modèle, un encadré du schéma est entièrement dédié au lexique, participant aux processus d'identification de mots. On indique que le lexique correspond aux représentations nécessaires pour lier la forme phonologique à la forme orthographique d'un mot, ainsi que pour accéder à toute autre information concernant ce mot – par exemple la classe de mots, sa morphologie, ses sens possibles, etc. (Perfetti, 1999). Encore une fois, ces éléments seront détaillés ultérieurement, dans la section portant sur l'identification de mots et ses différents mécanismes.

La spécification apportée quant à la sémantique des mots identifiés (« sélection de la signification ») ainsi que la place claire des connaissances utilisées par les différents processus font en sorte que ce modèle est approprié pour la présente recherche doctorale. Par ailleurs, il convient également de se pencher sur les modèles théoriques qui rendent compte en détails des processus spécifiques d'identification des mots écrits, étant donné que ce sont ces processus qui sont déficitaires chez les élèves concernés par cette recherche.

2.4 Modèles théoriques d'identification des mots écrits

L'identification des mots écrits est l'expression qui réfère aux processus cognitifs spécifiques à la lecture permettant de « transformer » une information de nature visuelle – une séquence de lettres – en une information sémantique – signification du mot lu. Différents modèles théoriques tentent de rendre compte des mécanismes impliqués dans l'identification des mots écrits. Aucun de ces modèles n'est parfaitement pertinent pour la présente recherche. C'est pourquoi plusieurs modèles seront présentés ici; certains modèles plus avancés scientifiquement, c'est-à-dire qu'ils ont été davantage documentés dans la littérature scientifique, qui rendent compte de la lecture experte, puis d'autres, un peu moins avancés, qui rendent compte particulièrement du développement de la lecture. Avant de présenter ces modèles, il convient de décrire brièvement les caractéristiques linguistiques de l'orthographe française, puisque, dans

cette approche psycholinguistique, les processus postulés dans les modèles varient en fonction des caractéristiques linguistiques des mots écrits.

2.4.1 Considérations préalables concernant les caractéristiques linguistiques de l'orthographe française

Les caractéristiques du système orthographique de la langue à lire influencent les mécanismes cognitifs qui sont mis en place par le lecteur d'une langue donnée pour identifier les mots écrits (Seymour *et al.*, 2003; Ziegler et Goswami, 2006). En conséquence, il convient de décrire brièvement les caractéristiques linguistiques de l'orthographe du français avant de décrire les différents mécanismes impliqués dans l'identification des mots écrits dans cette langue.

Le système d'écriture du français est un système alphabétique; les symboles écrits servent à représenter les « sons » de la langue orale, plus particulièrement les phonèmes (Desrochers *et al.*, 2011). Les phonèmes sont définis comme étant les « unités distinctives minimales » de la langue orale. En français, on recense 36 phonèmes : 16 voyelles, 3 semi-voyelles et 17 consonnes (voir l'annexe A pour la liste des phonèmes du français ainsi que les symboles de l'alphabet phonétique international utilisé dans le cadre de cette thèse pour représenter ces phonèmes). Les symboles écrits du français actuel correspondent aux symboles de l'alphabet latin; on compte ainsi 26 symboles écrits, que sont les lettres de cet alphabet. La grande disparité entre le nombre de phonèmes (40 phonèmes) et le nombre de symboles écrits (26 lettres) engendre la nécessité de recourir à différents moyens pour « pallier » l'insuffisance de caractères écrits pour transcrire l'ensemble des phonèmes de la langue orale, chacun à l'aide d'un seul symbole (Séguin et Desrochers, 2008). Ainsi, l'orthographe du français comporte des signes diacritiques, certaines lettres sont adjointes d'un accent aigu, grave ou circonflexe, ou de trémas. Elle comporte également des graphèmes complexes, c'est-à-dire que deux ou plusieurs lettres doivent être réunies pour représenter un seul phonème (p. ex. « au »-/o/, « eau »-/o/, etc.). Malgré le recours judicieux à ces moyens, le foisonnement de graphèmes complexes fait en sorte que les correspondances entre graphème et phonème (CGP) en lecture ne sont pas univoques, c'est-à-dire que certains graphèmes sont associés à plus d'un phonème : par exemple pour le graphème « s » on prononce /s/ dans « serpent » mais /z/ dans « plaisir », pour le graphème « t » on prononce /t/ dans « teinture » mais /s/ dans « relation ». Ces phénomènes rendent plus complexe le système orthographique du français. Quoi qu'il en soit de ces phénomènes plus ou moins complexes, Fayol et Jaffré (2008) désignent par l'expression « aspect phonographique de l'orthographe »

tous ces graphèmes (lettre ou groupe de lettres) qui servent à représenter les phonèmes d'une langue orale.

Par ailleurs, le système orthographique du français comporte également certains graphèmes qui ont une toute autre fonction que celle de représenter la phonologie. Ainsi, l'orthographe du français comporte des lettres ou groupes de lettres qui forment l'aspect sémiographique de l'orthographe (Fayol et Jaffré, 2008). La sémiographie de l'orthographe française se décline selon trois principales fonctions : 1) étymologique, les graphèmes ont pour fonction de rappeler l'origine plus ou moins lointaine des mots oraux ou écrits (p. ex. « doigt » ou « loup » rappellent les formes latines de ces mots); 2) distinctive, les graphèmes servent à distinguer à l'écrit deux concepts (sémantiques) différents désignés par une même forme orale – le cas des homophones (p. ex. « vers » et « vert »); 3) morphologique, les graphèmes renvoient aux morphèmes présents à l'intérieur des mots, c'est-à-dire les plus petites unités de sens (Huot, 2005; Gregory, 1999), (p.ex. le « d » de « grand » qui marque le lien avec les mots de même famille comme « grande, grandeur »; le « ette » de « fillette » qui signifie « petite »; le « s » dans « fleurs » qui marque le pluriel, etc.)

À cause des relations non univoques entre graphème et phonème dans l'aspect phonographique et à cause des graphèmes qui ont une fonction exclusivement sémiographique, on considère que le système orthographique du français est « quasi régulier » (Plaut *et al.*, 1996). Ceci indique qu'un nombre relativement important de mots peut être lu à l'aide de correspondances graphème-phonème « standards » – mots que l'on dit réguliers (comme le mot « sac ») –, alors que d'autres mots – appelés mots irréguliers ou d'exception (comme le mot « août ») – ne répondent pas à ces correspondances standards.

Dans une perspective interlangue, certains auteurs se penchent sur ces disparités de consistance entre les différents systèmes orthographiques. Dans une des plus imposantes études interlangues ayant été conduite en Europe, Seymour *et al.* (2003), grâce à la collaboration de chercheurs de 16 pays européens, ont comparé les systèmes orthographiques d'une douzaine de langues écrites. Ils placent les différentes langues étudiées pour l'identification des mots sur un continuum, partant des langues dont la consistance est très élevée (le finnois) jusqu'à des langues où la consistance est plutôt faible comme c'est le cas de l'anglais (voir le tableau 1), immédiatement précédé par le français et le danois.

Tableau 2.1 Classification théorique de certaines langues selon la complexité syllabique et l'opacité de leur orthographe (Seymour *et al.*, 2003)

		Opacité de l'orthographe				
		Transparente				**Opaque**
Structure syllabique	**Simple**	Finois	Grec Italien Espagnol	Portuguais	Français	
	Complexe		Allemand Norvégien Islandais	Néerlandais Suédois	Danois	Anglais

Un des constats issus de ces études interlangues est celui selon lequel l'apprentissage de la lecture dans une langue donnée est influencé par la consistance des relations entre graphème et phonème (Ziegler et Goswami, 2005 pour une revue des études permettant ce constat). Par exemple, on documente que la très grande majorité des élèves arrivent à identifier les mots avec un taux de succès près du score parfait dès la moitié de la première année dans des systèmes orthographiques comme le grec, le finnois, l'espagnol ou l'italien, alors que ce n'est pas du tout le cas des élèves scolarisés en français ou en anglais chez lesquels on observe de très faibles scores jusqu'au milieu de la 2^{e} année environ (Goswami *et al.*, 1998), ces scores étant plus faibles en anglais qu'en français.

2.4.2 Modèles théoriques de l'identification experte de mots écrits

Les modèles théoriques doivent tenir compte d'un certain nombre de caractéristiques de l'orthographe pour décrire les différents mécanismes de l'identification de mots. Les premiers modèles théoriques des mécanismes cognitifs qui permettent d'identifier les mots écrits ont été élaborés pour rendre compte de la lecture experte. Encore aujourd'hui, ce sont les modèles de la lecture experte qui sont les plus nombreux et les plus documentés.

2.4.2.1 Modèles à deux voies indépendantes

Dans les années 1970, est apparu le modèle à deux « voies ». L'idée d'abord émise par De Saussure quelques années auparavant a été reprise et modélisée dans différents modèles « boîtes et flèches » notamment par Forster et Chambers (1973), puis par Marshall et Newcomb (1973; voir Coltheart, 2005 pour historique détaillé de ce modèle). Le principe général de ce modèle est celui selon lequel le lecteur a

besoin de deux mécanismes distincts, appelés voies, pour être en mesure d'identifier les mots écrits; un premier mécanisme pour identifier les mots d'exception, puis un deuxième pour identifier les mots réguliers, qui répondent à l'application de correspondances graphème-phonème (CGP) « standards ». Coltheart (1978) et Coltheart, Patterson et Marshall (1980) identifient ainsi la voie lexicale pour les mots d'exception. Cette voie implique que le lecteur accède à un lexique mental (une sorte de dictionnaire interne) où sont emmagasinées les connaissances sur l'orthographe et la prononciation des mots déjà rencontrés et mémorisés par le lecteur. En accédant à ce lexique, il peut donc identifier le mot et ultimement en comprendre sa signification. L'autre voie, la voie non lexicale, n'implique pas ce lexique, mais requiert plutôt que le lecteur utilise les règles régissant les correspondances graphème-phonème « standards » pour accéder à la prononciation d'un mot régulier et ainsi accéder ultimement à sa signification (Coltheart, 1978). C'est notamment cette voie qui prend en charge la lecture de pseudomots et de non-mots puisque, par définition, ils n'existent pas dans le lexique mental du lecteur, d'où son nom (voie non lexicale). Il est à noter que plusieurs termes différents sont utilisés selon les auteurs pour désigner ces deux voies : voies lexicale et non lexicale, voies directe et indirecte, voies d'adressage et d'assemblage, etc.

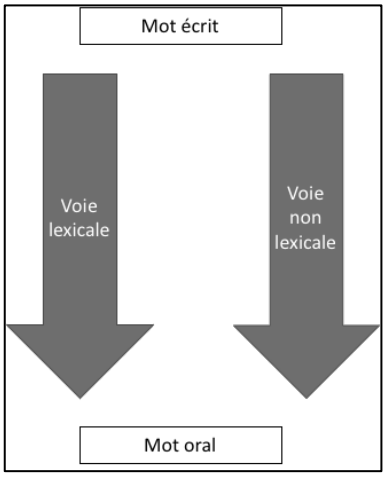


Figure 2.2 Représentation schématique du modèle à deux voies indépendantes *(Coltheart, 1978; Coltheart, Patterson et Marshall, 1980)*

Dans ces modèles, les deux « voies » constituent des modules indépendants, c'est-à-dire qu'un mécanisme est mis en œuvre de manière totalement indépendante de l'autre, et le traitement d'un mot écrit ne peut se faire dans plus d'une voie.

Il convient de noter ici que le modèle a d'abord été élaboré (Coltheart les premières publications) en tenant compte de données obtenues à partir d'adultes cérébrolésés anglophones présentant une dyslexie acquise, donc à partir de personnes qui maîtrisaient déjà l'identification de mots mais qui ont perdu cette capacité suite à une lésion cérébrale. Dans ce modèle, le rôle de la sémantique n'est pas très détaillé. L'accès à la sémantique se fait automatiquement lorsque le lecteur accède à son lexique mental dans la voie lexicale.

2.4.2.2 Modèle à deux voies en cascade (DRC)

Au cours des années 1980, les avancées en informatique ont rendu possible l'élaboration des modèles computationnels, c'est-à-dire des modèles qui peuvent être implémentés dans un programme d'ordinateur. Dans cette perspective, Coltheart *et al.* (2001), dans la poursuite de ses travaux sur le modèle classique à deux voies indépendantes[3], a créé une version computationnelle de ce modèle, ce qui a donné lieu au modèle à deux voies en cascade (*Dual-Route Cascaded model*, DRC). La figure 2.4 représente le modèle DRC.

[3] Les modèles qui ne sont pas implémentés sont dits « verbaux ». Le modèle à deux voies indépendantes ainsi que les modèles présentés en 2.4.3 pour l'apprentissage sont des modèles verbaux, alors que tous les autres modèles de la lecture experte sont également computationnels.

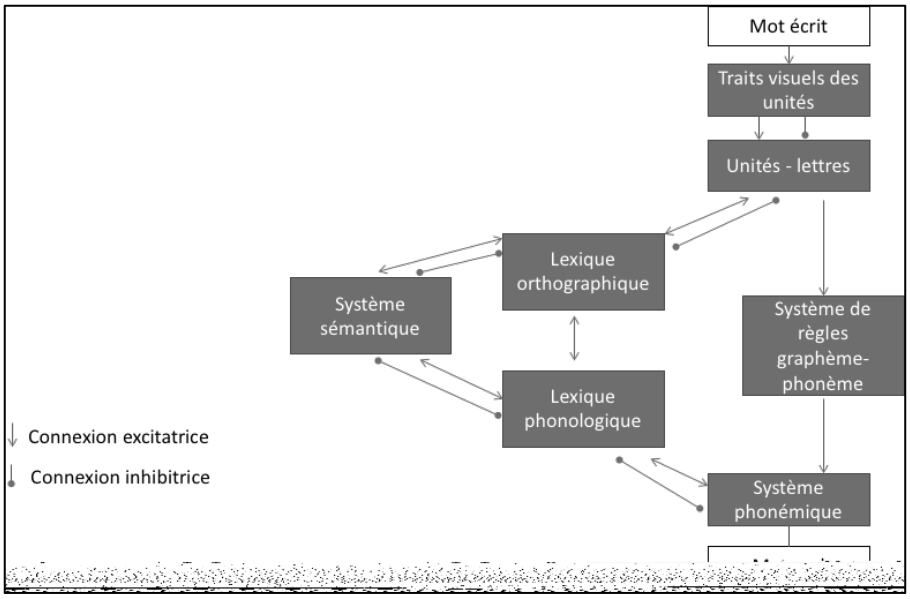


Figure 2.3 Modèle à deux voies en cascade (DRC), de Coltheart *et al.* (2001)

Le modèle DRC spécifie que l'activation d'un module à l'autre est réalisée en cascade plutôt que par seuil (Coltheart *et al.* 2001). Alors que la notion de seuil réfère au fait que le traitement réalisé par un module n'activera pas les modules suivants tant que l'activation n'aura pas atteint un certain seuil, le traitement en cascade (proposé par McClelland, 1979) implique qu'aussitôt qu'une activation est présente dans un module donné, si légère soit-elle, elle suscite également aussitôt une activation équivalente dans les modules adjacents[4]. De plus, le modèle DRC prévoit une interactivité, c'est-à-dire que le flux d'activation peut se propager dans les deux directions entre les modules adjacents (McClelland, 1993). Ceci est représenté par les différents traits entre les modules à la figure 2.4. Les deux types de traits sont associés à deux types d'interaction : l'interaction peut être inhibitrice ou excitatrice. Une interaction inhibitrice se traduit par une activation d'une unité particulière qui rendra plus difficile l'activation d'autres unités qui ne lui ressemblent pas. À l'inverse, une interaction excitatrice se traduit par le fait que l'activation d'une unité particulière augmentera l'activation d'autres unités qui lui ressemblent. Ainsi, dans le modèle, si le traitement des lettres du mot écrit conduit à l'activation de la représentation visuelle du mot entier dans le « lexique orthographique », il y aura une interaction excitatrice qui activera encore plus chaque lettre du mot dans le module des lettres puis une interaction inhibitrice qui « empêchera » l'activation des lettres

[4] L'article de 2001 sur le modèle DRC justifie ce choix. Par souci de concision, une seule justification est brièvement présentée ici, à titre d'exemple.

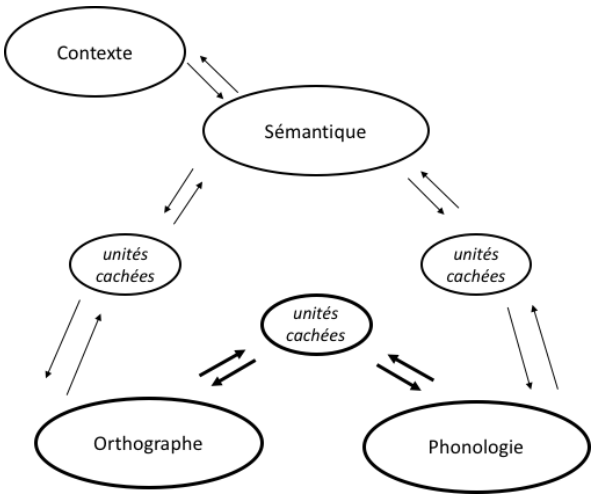


Figure 2.5 Modèle à traitement parallèle distribué (Parallel Distributed Processing, PDP), traduit et adapté de McClelland et Seidenberg (1989) et de Plaut *et al.* (1996)

Les composantes dont les traits sont présentés en lignes grasses sont les seules composantes à avoir été implémentées dans cette version du modèle, sur des mots monosyllabiques. Ainsi, peu de détails sont disponibles quant à la composante sémantique. Quoi qu'il en soit, dans chacune de ces composantes, un certain nombre d'unités « primitives » sont utilisées pour encoder l'information. Dans la composante orthographique, ces unités sont les graphèmes (Plaut *et al.*, 1996). Dans la composante phonologique, les unités « primitives » correspondent aux phonèmes de la langue à lire. La représentation d'un mot dans chaque composante prend ainsi la forme d'un patron d'activation distribuée à travers les unités « primitives » (voir la figure 2.5). Pour lier des informations de composantes distinctes – p. ex. lier la représentation orthographique à sa représentation phonologique – le modèle a recours à des unités cachées. Sans entrer dans le détail, mentionnons que ce sont les unités cachées qui permettent de gérer une plus grande complexité des informations à traiter. Par exemple, ce sont les unités cachées qui permettent de gérer des phénomènes comme les irrégularités ou le fait qu'un même graphème puisse avoir différentes réalisations phonologiques.

Le modèle PDP propose ainsi que les lettres sont d'abord traitées visuellement, activant ainsi les unités de la composante orthographique, c'est-à-dire les graphèmes qui composent le mot. Ce patron d'activation se propage ensuite dans les autres composantes (en passant par les unités cachées) avec pour résultat final un patron d'activation circulant dans les unités représentant les phonèmes du mot à prononcer, puis la production de ces phonèmes. Les flèches entre chaque élément du modèle indiquent que le flux de

l'activation est bidirectionnel. On propose ainsi deux trajets distincts pour identifier les mots : 1) correspondance directe entre la représentation orthographique du mot et sa représentation phonologique; 2) correspondance entre la représentation orthographique et la représentation phonologique qui passe par la sémantique (Powell *et al.*, 2006). Encore une fois, comme la composante sémantique n'est pas implémentée, peu de détails sont disponibles sur le fonctionnement précis de ce mécanisme. Les caractéristiques du modèle font en sorte que le mécanisme utilisé pour passer de l'information orthographique à l'information phonologique est le même pour les types d'items à identifier, que ceux-ci soient des mots irréguliers, réguliers ou des pseudomots: dans tous les cas, il s'agit de patrons d'activation distribués qui seront mis en correspondance, peu importe que l'on procède petite unité par petite unité (p. ex. « s » $\rightarrow$ /s/, « a » $\rightarrow$/a/ = /sa/) ou pour un mot complet (p. ex. « août » $\rightarrow$/u/). Certains auteurs réfèrent ainsi à ce modèle comme étant un modèle à une voie (Powell *et al.*, 2006).

Plusieurs autres modèles connexionnistes de l'identification des mots écrits existent, comme le modèle connexionniste à deux processus (CDP+ de Perry *et al.*, 2007) ou encore le modèle à activation interactive bimodale (BIAM de Grainger et Ferrand, 1994). Pour les besoins de la présente recherche, seule la présentation du modèle PDP est nécessaire, ceci afin de mettre en lumière certaines caractéristiques communes à ce modèle. La place de la sémantique est explicite dans ce modèle. Et contrairement au modèle DRC (Coltheart, 2005), la sémantique a une influence beaucoup plus importante, puisqu'elle peut intervenir à tout moment de l'identification, et dans tous les mécanismes d'identification des mots écrits. Par contre, comme cette partie du modèle n'a pas été implémentée avec le reste du modèle, peu de détails sont fournis sur le fonctionnement exact des mécanismes impliquant la sémantique.

2.4.3 Modèles théoriques développementaux de l'identification des mots écrits et le modèle à double fondation

Les modèles de l'identification des mots écrits présentés ci-dessus ont pour objectif de rendre compte de la lecture experte. À partir du milieu des années 1980, certains modèles se sont centrés sur l'apprentissage des processus d'identification des mots écrits. Un des premiers constats réalisés est celui, théorique, selon lequel rien n'indique que les mécanismes mis en œuvre par le lecteur expert soient les mêmes, ou soient mis en œuvre de la même manière, chez le lecteur débutant. Or, puisque la présente thèse concerne des élèves du primaire, il convient de s'intéresser aux modèles développementaux. Ainsi, Frith (1986) souligne une importante différence entre les mécanismes du lecteur expert et ceux du lecteur en apprentissage : la notion de changement. En effet, le lecteur qui est en train d'apprendre à identifier les mots, par

définition, doit développer ces mécanismes. Le développement des mécanismes implique que leur état ou leur fonctionnement est en perpétuel changement, pour passer de représentations et mécanismes peu développés (ou peu efficaces) à des représentations et mécanismes complètement développés (ou les plus efficaces possible). Cette différence donne donc lieu à l'élaboration de modèles qui visent à décrire la dynamique particulière qui existe pendant l'apprentissage de la lecture (Seymour, 1986; Laplante, 2011), c'est-à-dire les « modèles développementaux ». Le modèle de Frith (1986) et le modèle de Ehri (1997, 2005, 2007) sont des modèles qui impliquent une succession de phases ou de stades (Ehri, 2007). Le passage d'un stade à l'autre n'est possible que lorsque le précédent stade est suffisamment maîtrisé; le stade subséquent correspond à l'élaboration d'une nouvelle « stratégie » sur la base de ce qui a été développé au stade précédent. Bien que ces modèles soient importants, ils ne sont pas décrits en détails ici par souci de concision. Le modèle présenté ci-dessous est celui qui semble le plus avancé dans les modèles développementaux, le seul à envisager la coexistence de différents mécanismes d'identification des mots écrits plutôt que de référer à des stades ou phases hermétiques.

Le modèle à double fondation proposé initialement par Seymour (1986) et ultérieurement raffiné (Seymour, 1997; Duncan et Seymour, 2000; par exemple) fait partie des modèles qui tentent de décrire comment l'identification des mots écrits se développe chez le lecteur en apprentissage. Chaque phase est dominée par un type particulier de procédure d'identification des mots écrits. Bien qu'on propose que les lecteurs se situent à l'une ou l'autre des phases, cela ne veut pas dire pour autant qu'ils n'utilisent pas, dans certaines circonstances, les procédures caractéristiques d'une autre phase. Comme les modèles de Frith (1986) et de Ehri (1997, 2005, 2007), le modèle à double fondation est essentiellement verbal et symbolique. Ce modèle, d'abord élaboré pour l'anglais, a été soumis à la description de l'identification des mots écrits d'un certain nombre de langues européennes (p. ex. Seymour *et al.*, 2003), dont le français.

Ainsi, le modèle propose quatre mécanismes interreliés (dont l'encadré est en gras à la figure 2.9), sollicités à travers 4 phases (présentées dans la partie gauche de cette figure).

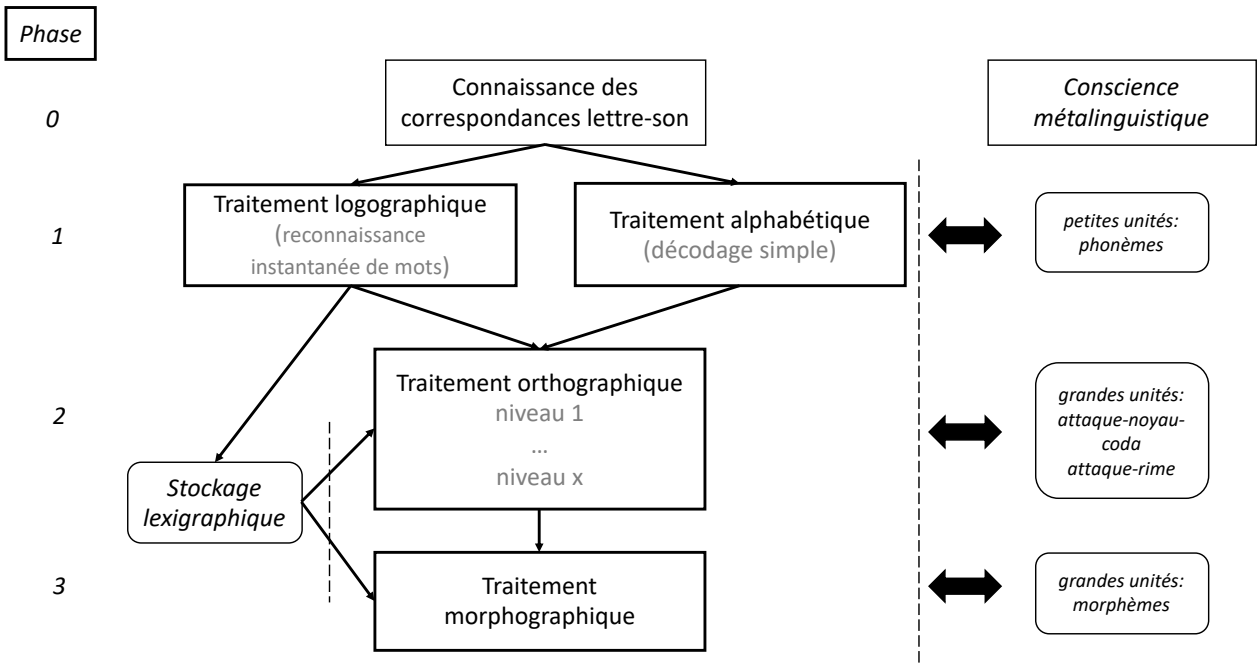


Figure 2.6 Modèle à double fondation, traduit et adapté de Seymour (2008)

Tel qu'indiqué dans la représentation schématique du modèle, le point de départ du développement de l'identification des mots écrits réside dans les connaissances et habiletés que les prélecteurs possèdent déjà à propos de la langue orale, soit la conscience métalinguistique et le stockage lexigraphique. Il s'agit en fait du seul modèle développemental qui met en évidence la contribution des habiletés métalinguistiques présentes avant le début de l'apprentissage ainsi que l'effet réciproque qu'a cet apprentissage sur le développement des habiletés métalinguistiques. Dans les langues alphabétiques, le lecteur apprend d'abord que les symboles écrits ont une valeur sonore, c'est la « compréhension du principe alphabétique » (p. ex. Ehri *et al.*, 2001 à propos du *National Reading Panel*). Les enfants apprennent ainsi les correspondances entre les lettres et leur phonème. Ceci correspond à la phase 0 du modèle, la phase de la connaissance alphabétique.

Pendant la phase 1, appelée la littératie de fondation, deux mécanismes de base sont graduellement développés de manière parallèle, soit le traitement logographique et le traitement alphabétique. Ainsi, Seymour réfère aux travaux de Ehri (1997) pour décrire le traitement logographique, soit le mécanisme de reconnaissance instantanée de mots entiers. Ce mécanisme implique qu'il existe en mémoire un bagage de mots qui ont été rencontrés suffisamment fréquemment pour qu'ils aient été emmagasinés dans leur

forme complète en mémoire. Ehri (2005) propose que ce mécanisme dépasse la seule mémorisation visuelle du mot entier, mais qu'il requiert plutôt une expérience des correspondances graphème-phonème également. Le traitement alphabétique pour sa part correspond à l'opération de conversion séquentielle, unité par unité, des graphèmes en phonèmes, puis la fusion de ces phonèmes pour obtenir la manière de prononcer le mot complet (Seymour, 2008). S'appuyant sur les règles de correspondances « standards » existant entre graphème et phonème, ce mécanisme prend efficacement en charge l'identification de mots réguliers, qui ne comportent pas d'exception. Seymour *et al.* (2003) proposent que ces deux mécanismes de base arrivent à leur maximum environ vers l'âge de 7 ans dans le cas de l'anglais (9 ans, dans Duncan et Seymour, 2000).

Une fois que ces deux mécanismes sont suffisamment développés, le lecteur en apprentissage élabore un mécanisme qui traite des phénomènes un peu plus complexes des mots écrits. On entre ainsi dans la phase 2, celle de la littératie orthographique. Partant des structures établies lors de la phase 1 à travers le traitement logographique et le traitement alphabétique, le lecteur est en mesure de dégager la complexité du système orthographique représentée dans un format « généralisable et abstrait », pour préciser ce qui est légitime ou non dans l'orthographe de la langue à lire. La prise en charge de ces phénomènes complexes est réalisée par le traitement orthographique. Seymour *et al.* (2003) proposent que le cadre orthographique développé dans la phase 3 le soit un peu à la manière décrite par Plaut *et al.* (1996), soit de manière connexionniste (voir la section 2.4.2.3 pour plus de détails). Ce faisant, certaines règles contextuelles (le « s » qui se prononce /z/ lorsqu'il est précédé et suivi de deux lettres-voyelles) sont « emmagasinées » dans le cadre orthographique. Le cadre orthographique traite également des régularités « statistiques » dégagées à partir de la structure du système (p. ex. suite de graphèmes apparaissant souvent ensemble comme « exi », position acceptée des doubles consonnes, etc.), jusqu'à la prise en compte de phénomène orthographique particulier à un seul mot (« août » fait /u/ dans ce seul mot), ce que l'on désigne par l'expression mot irrégulier.

Finalement, la dernière phase décrite par Seymour est la phase de littératie morphographique. Dans cette phase, le lecteur développe sa représentation des mots complexes, c'est-à-dire composés de plusieurs morphèmes, mots souvent multisyllabiques. Ainsi, le traitement morphographique opère sur des représentations abstraites des parties de mots que sont les morphèmes et la manière dont elles s'actualisent dans l'écrit (morphographie). Les représentations abstraites distingueront ainsi les racines des affixes, les graphèmes à rôle morphographique, etc. C'est dans cette dernière phase que la sémantique

trouve une certaine place explicite, en ce sens que les morphèmes correspondent aux plus petites unités de sens contenues dans les mots. Les significations associées aux affixes, ainsi qu'aux lettres muettes dérivables par la morphologie interviennent dans le traitement morphographique. C'est en effet la composante sémantique qui distingue les traitements orthographique et morphographique : une suite de lettres que l'on retrouve fréquemment dans la langue, si elle a une signification particulière qui lui est rattachée fait partie du traitement morphographique (p. ex. « -ette » signifie « petite », alors que « exi » ne contient aucune signification). Il s'agit ainsi de la sémantique infra-lexicale, la sémantique de petites parties de mots.

2.4.4 Théorie de la granularité psycholinguistique

Tel que mentionné à la section 2.4.1 sur les caractéristiques linguistiques de l'orthographe française, chaque langue à lire possède des caractéristiques particulières qui vont influencer les mécanismes d'identification des mots écrits que les lecteurs de cette langue doivent utiliser. Plutôt que de présenter un modèle qui décrit précisément le fonctionnement de tous les mécanismes d'identification de mots, Ziegler et Goswami (2005) proposent une théorie générale dont devraient tenir compte les modèles pour être en mesure de juger de leur applicabilité à différentes langues, comme le français, puisque c'est ce dont il s'agit dans cette thèse.

Partant du principe que l'identification des mots écrits implique des connaissances langagières que les jeunes enfants maîtrisent déjà à l'oral – notamment les connaissances phonologiques à propos de la prononciation des mots ainsi que les connaissances sémantiques à propos de la signification de ces mots –, les auteurs de la théorie de la granularité psycholinguistique (Ziegler et Goswami, 2005) insistent sur le fait que le principal enjeu du lecteur qui apprend à identifier les mots écrits est l'établissement de correspondances entre les symboles écrits et les sons. Ils appellent cet établissement de correspondances entre symboles et sons le recodage phonologique. C'est effectivement ce recodage phonologique qui permet un apprentissage adéquat, parce que c'est celui-ci qui permet à l'enfant de « décrypter » des mots qu'il a déjà entendus mais n'a jamais vus – lus. Share (1995) insiste sur la fonction d'autoapprentissage de ce mécanisme de recodage. Quoi qu'il en soit, Ziegler et Goswami (2005) centrent leur théorie, d'une part, sur ce mécanisme de recodage phonologique et, d'autre part, sur le fait que ce mécanisme, pour être efficace, doit tenir compte de la granularité des systèmes orthographique (symboles écrits et phonologique (sons). La granularité des systèmes renvoie à la notion de la taille des unités pertinentes. Ainsi, le lecteur en apprentissage doit retrouver la taille d'unité pertinente dans les systèmes

orthographique et phonologique pour chaque mot, lui permettant d'établir une correspondance qui est à la fois simple et non ambiguë. Par exemple, choisir la granularité de la lettre dans le système orthographique ne mène pas à une correspondance simple et non ambiguë dans le mot « clou » puisqu'il y a présence d'un graphème complexe. Pour ce mot, la granularité du système orthographique doit donc être sélectionnée au niveau du graphème plutôt que de la lettre (considérer les lettres « o » et « u » comme une seule unité), et doit être sélectionnée au niveau des phonèmes dans le système phonologique. Le premier défi auquel font face les enfants qui apprennent l'identification des mots écrits est donc la disponibilité de ces unités. En effet, étant donné que les enfants ont un système phonologique déjà développé par leur maîtrise de mots oraux (tant en compréhension qu'en production), on considère que leur système phonologique est déjà organisé suivant les caractéristiques phonologiques de la langue orale. L'organisation du système phonologique varie d'une langue orale à l'autre quant à la taille des unités pertinentes. En français, le niveau du phonème n'est pas nécessaire pour comprendre et produire des mots oraux. Par conséquent, le système phonologique des locuteurs du français (encore non lecteurs) n'est pas organisé autour des représentations des phonèmes de la langue, mais bien sur des unités plus larges, notamment la syllabe (Demont et Gombert, 1996). L'unité phonologique qu'est le phonème n'est donc pas encore complètement disponible pour la plupart des lecteurs qui débutent l'apprentissage de l'identification des mots écrits (voir la figure 2.10). Établir les correspondances entre graphème et phonème (CGP) nécessitera un développement cognitif supplémentaire, particulier au traitement de l'écrit, complexifiant ainsi son apprentissage.

Selon cette théorie, le deuxième défi auquel les enfants font face est celui de la consistance des correspondances entre orthographe et phonologie. Cette consistance varie d'une langue écrite à l'autre, mais peut également varier à l'intérieur d'une même langue écrite. En effet, une même langue écrite peut présenter une consistance différente entre l'identification des mots écrits (lecture) et la production des mots écrits (écriture). C'est effectivement le cas en français (voir la section 2.4.1 pour les détails); les correspondances sont très inconsistantes en production de mots écrits, mais sont beaucoup moins inconsistantes en identification des mots écrits. Selon cette théorie, la consistance des correspondances influence l'apprentissage. Apprendre à identifier des mots écrits dans une langue dont les correspondances sont totalement consistantes est beaucoup plus aisé car les correspondances rencontrées sont stables et récurrentes; un graphème, peu importe dans quel mot écrit de la langue il est rencontré, correspond toujours au même phonème. En pareil cas, le recours aux CGP peut être suffisant. À l'inverse, lorsque les correspondances sont inconsistantes, un même graphème n'a pas toujours la même

correspondance phonologique selon le mot dans lequel il apparaît (le « s » de « maison »); l'apprentissage des correspondances est donc plus difficile et surtout plus long. Comme le recours au CGP uniquement n'est pas suffisant, le système a davantage de pression pour développer un mécanisme de reconnaissance instantanée de mots entiers. Dans le cas du français particulièrement, cela veut donc dire que l'apprentissage de l'identification de mots écrits est plus aisé que l'apprentissage de la production des mots écrits, ici à cause de la consistance des correspondances entre orthographe et phonologie.

Le dernier défi auquel font face les enfants, et qui est en quelque sorte une conséquence des deux autres problèmes, est celui de la granularité. En effet, plus un système est inconsistant, plus il est nécessaire de varier la taille des unités pertinentes dans les systèmes orthographique et phonologique. Par exemple, devant le graphème « t » en début de syllabe, le lecteur doit considérer le reste de la syllabe pour savoir si ce dernier se prononce /t/ ou /s/ (s'il y a un « i »+voyelle). Ainsi, plus une langue est inconsistante, plus la granularité nécessaire à prendre en compte varie, tant dans le système orthographique que dans le système phonologique. La prise en compte de petites unités mène généralement à plus de confusions (p. ex. considérer le niveau du graphème « c » pour « seconde » mène à une identification erronée), malgré la pertinence dans d'autres cas de la prise en compte de ces petites unités. Cette caractéristique exige donc que les enfants en apprentissage soient suffisamment flexibles, pour considérer différentes granularités selon les mots du système qu'ils ont à identifier ou à produire.

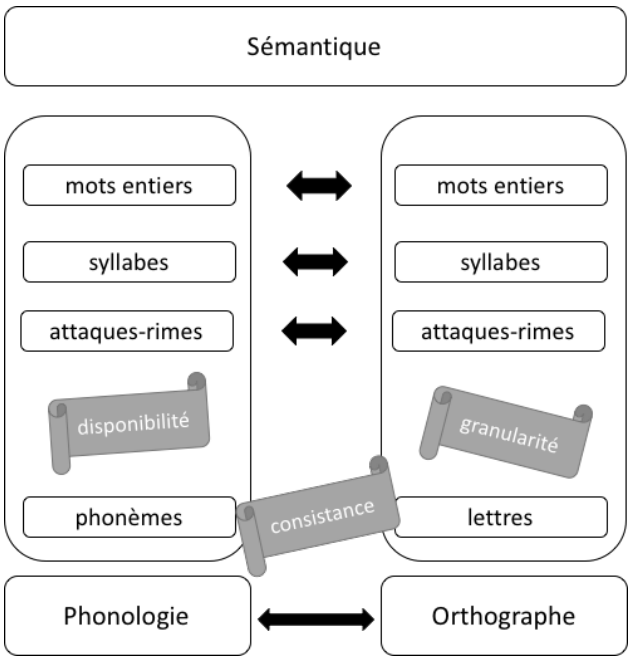


Figure 2.7 Représentation schématique des trois principaux défis de l'apprentissage de la lecture : disponibilité, consistance et granularité (traduction de Ziegler et Goswami, 2005)

Cette théorie précise que les pratiques pédagogiques mises en place pour enseigner l'identification des mots écrits ont une immense incidence sur le développement de mécanismes efficaces, surtout dans les langues dont l'orthographe est inconsistante. Ainsi, les pratiques pédagogiques optant pour l'enseignement direct qui se centrent sur les différentes granularités nécessaires à développer chez les élèves selon la langue (p. ex. mots entiers, correspondances graphème-phonème, et tout ce qui est au milieu lorsque nécessaire) ont un rôle à jouer.

En somme, comme l'orthographe française a ses propres caractéristiques, qui sont différentes de celles de l'orthographe anglaise, et comme les modèles ont surtout été élaborés pour rendre compte de l'anglais, il convient de faire preuve de beaucoup de précautions quand on tente d'appliquer les modèles théoriques décrits dans les sections précédentes au français.

2.4.5 Brève critique des modèles théoriques de l'identification des mots écrits

En résumé, le tableau 2.2 présente les principales caractéristiques des différents modèles théoriques qui ont été discutés dans cette section. Tel qu'indiqué précédemment, certains modèles se centrent sur la lecture experte, alors que le modèle à double fondation se centre sur le développement de la lecture. Bien entendu, vu le sujet de cette thèse, les modèles développementaux revêtent un caractère primordial. Par ailleurs, les modèles DRC et PDP sont les seuls modèles parmi ceux présentés à être computationnels, c'est-à-dire qu'ils ont été implémentés dans un programme d'ordinateur. Cette implémentation nécessitant de recourir à des modèles suffisamment précis et permettant de comparer les productions du modèle aux données obtenues dans des études comportementales, il va sans dire que les modèles computationnels doivent être considérés. De plus, le modèle à deux voies indépendantes ainsi que le modèle à double fondation sont des modèles symboliques, alors que les modèles DRC et PDP sont des modèles connexionnistes (le modèle à double fondation évoque la possibilité que le cadre orthographique fonctionne de manière connexionniste). Finalement, les modèles DRC, PDP accordent une place explicite à la sémantique lexicale, et le modèle à double fondation, par le traitement morphographique, accorde une place explicite à la sémantique infra-lexicale (la sémantique des morphèmes). Le modèle PDP peut également tenir compte de la sémantique infra-lexicale.

Tableau 2.2 Résumé des principales caractéristiques des modèles théoriques de l'identification des mots écrits présentés dans le cadre théorique

	Perspective développementale	Computationnel	Symbolique	Connexionniste	Rôle explicite de la sémantique lexicale	Rôle explicite de la sémantique infralexicale
Modèle à deux voies indépendantes (Coltheart, 1978)			✓			
Modèles à deux voies en cascade (DRC; Coltheart, 2005)		✓		✓	✓	
Modèle à traitement parallèle distribué (PDP; Seidenberg et McClelland, 1989)		✓		✓	✓	✓
Modèle développemental à double fondation (Seymour, 2008)	✓		✓	✓		✓

Les modèles théoriques les plus avancés et les plus puissants actuellement, en ce sens qu'ils rendent compte d'un nombre important de phénomènes comportementaux, sont les modèles computationnels et connexionnistes. Il s'agit ainsi des modèles à deux voies en cascade (DRC) et le modèle à traitement parallèle distribué (PDP). Par contre, ces modèles ne sont pas des modèles développementaux; ils ont été élaborés pour rendre compte de la lecture experte. Alors que la présente recherche s'intéresse aux lecteurs d'âge scolaire, il convient de considérer qu'il puisse y exister une dynamique particulière entre les différents mécanismes d'identification des mots écrits au cours du développement qui n'est pas complètement reflétée dans ces modèles.

Par ailleurs, tous les modèles adhèrent à l'idée selon laquelle le lecteur doit traiter trois types d'information pour identifier un mot écrit : information visuelle – symboles écrits des mots que sont les

lettres, information phonologique – phonèmes des mots; et information sémantique – signification des mots. Tous s'entendent sur le fait que l'identification de mots débute indéniablement par le traitement de l'information visuelle d'un mot, puis culmine par l'accès à (ou l'activation de) la signification de ce mot. Toutefois, plusieurs des modèles présentés jusqu'ici se sont concentrés à décrire comment le lecteur fait pour passer d'une information visuelle à une information phonologique, permettant de prononcer un mot à voix haute (Ferrand, 2007). Comme il a été présenté précédemment, la nature des informations visuelles et phonologiques pertinentes est déterminée par les caractéristiques du système de l'orthographe; puis la nature des représentations exactes et le fonctionnement des mécanismes opérant sur celles-ci sont décrits précisément dans les modèles. Cependant, tant le modèle à deux voies indépendantes que le modèle à double fondation ne précisent pas la place de la sémantique lexicale dans leur modèle. Par la formulation de ces modèles, il semble que l'on considère que la signification du mot est obtenue d'une manière automatique une fois que le traitement des informations visuelles et phonologiques est complété. Par exemple, Seymour (2008) écrit : « le traitement des symboles [du code écrit] implique une mise en correspondance directe avec le nom de la chose désignée […] avant la mise en correspondance avec la représentation sémantique » (p.4). Il considère ainsi que l'accès à la signification du mot lu constitue la finalité de l'identification de ce mot. Ce faisant, ces modèles sont peu à même d'expliquer le rôle que la sémantique lexicale peut jouer dans l'identification des mots écrits. D'un autre côté, les modèles à deux voies en cascade (DRC) et à traitement parallèle distribué (PDP), qui prévoient une place explicite à la sémantique lexicale par le « système sémantique » et la « composante sémantique » respectivement, n'ont, à ce jour, pas implémenté complètement la partie sémantique de leur modèle lors des simulations réalisées. Ainsi, même si la place de la sémantique lexicale est prévue dans leur modèle, l'influence que cette dernière peut avoir pendant l'identification des mots écrits demeure encore théorique et imprécise.

Considérant ce qui a été mentionné, il s'avère donc pertinent d'envisager que tant les modèles DRC et PDP seront considérés dans le cadre de la présente recherche. En effet, ces modèles seront considérés dans les choix méthodologiques et dans l'interprétation des résultats obtenus.

2.5 Sémantique lexicale

Malgré le fait que quelques modèles accordent une place explicite à la sémantique, aucun n'a décrit exactement quelles représentations et quels mécanismes sont impliqués dans le traitement de l'information sémantique des mots écrits. Puisque la sémantique occupe une place importante dans cette recherche doctorale, il convient de s'y pencher.

2.5.1 Caractéristiques sémantiques des mots et leur représentation interne chez le lecteur

Tel que mentionné précédemment, la sémantique des mots réfère de manière générale à la signification de ces derniers. Ainsi, le lecteur qui doit identifier le mot écrit « fille » accédera ultimement à sa signification, c'est-à-dire « un être humain de sexe féminin, jeune » (TLFi). La signification des mots, ou leurs différentes significations lorsqu'ils en ont plusieurs, sont généralement recensées dans les entrées des dictionnaires, par la discipline que l'on appelle la lexicographie (Ducrot, 1972). À cause de l'infinité de concepts désignés par les mots d'une langue, il va sans dire que l'étude de la sémantique est extrêmement complexe. En effet, le défi consiste à étudier comment les mots d'une langue renvoient à des réalités nombreuses, de nature variée et en nombre quasiment infini.

2.5.1.1 Distinction forme et sens : homonymie et synonymie

La signification d'un mot[6] est distincte de sa forme; la forme d'un mot réfère à son information visuelle, c'est-à-dire les lettres qui le composent à l'écrit, puis à son information phonologique, c'est-à-dire les phonèmes qui le composent à l'oral. La signification d'un mot correspond à son contenu (Gregory, 1999). Le contenu d'un mot, son information sémantique, réfère donc à des objets dans le monde réel qui sont des instances d'un concept, ou bien il réfère à une représentation interne d'un concept, par exemple lorsque le mot renvoie à une entité abstraite (Brousseau et Roberge, 2000).

La distinction entre forme et sens des mots amène un certain nombre de constats. D'abord, il existe des mots pour lesquels une même forme lexicale renvoie à plus d'un concept. Les liens entre forme et sens ne sont donc pas toujours univoques. Ainsi, la même forme lexicale « livre », qui se prononce /livR/, réfère à deux concepts distincts, soit LIVRE[1] - *un assemblage de pages comportant des signes destinés à être lus*, et LIVRE[2] - *unité de mesure du poids, divisée en onces*. On désigne ce phénomène par le terme homonymie : une même forme lexicale renvoie à des concepts sémantiques distincts. Dans cet exemple, « livre »[1] et « livre »[2] partagent tant la forme visuelle (même suite de lettres) que la forme phonologique (même suite de phonèmes). Il existe également des homonymes qui ne partagent que la forme phonologique : par exemple, la même forme phonologique /tõ/ réfère à un *poisson* lorsqu'il prend la forme visuelle « thon », mais réfère au déterminant possessif quand il prend la forme « ton » (ou encore, au Québec à tout le moins, à un insecte lorsqu'il prend la forme « taon »). Ce type d'homonymes est désigné par le terme

[6] Bien que certains fassent une distinction entre les termes « signification » et « sens » (p. ex. Brousseau et Roberge, 2000), ces deux expressions sont utilisées dans ce cadre théorique de manière équivalente.

homophones – mêmes phonèmes. D'autres homonymes ne partagent que la forme visuelle : « fils » est associé à la forme phonologique /fis/ quand il désigne *enfant masculin de* …, mais à la forme /fil/ quand il désigne *fibre longue et déliée d'une matière textile ... au pluriel.* Ce dernier type d'homonymes est désigné par le terme homographes – même orthographe.

Pour ajouter à l'idée de relations non univoques entre forme et sens, il existe également des cas où un même concept est désigné par des formes distinctes. Par exemple, « canapé » et « sofa » sont des formes lexicales distinctes qui renvoient au concept de *meuble rembourré constitué d'au moins un dossier.* Les deux formes lexicales « canapé » et « sofa » sont ainsi appelées synonymes, elles ont au moins un sens commun. Bien que plusieurs mots présentent des similarités de sens, rares sont les synonymes parfaits, et ce, dans toutes les langues.

2.5.1.2 Liens de sens entre les mots : hyponymie, prototype et antonymie

La distinction entre la forme d'un mot et son sens étant clarifiée, il convient de poursuivre la description des caractéristiques sémantiques des mots de la langue, puis de décrire leur représentation interne chez le lecteur.

Depuis Rosch (1975; 1978), on considère que la notion de sémantique est étroitement liée à celle de catégorisation. La catégorisation est étudiée par un grand nombre de disciplines, notamment plusieurs chercheurs qui s'intéressent à la cognition en général (Cohen et Lefebvre, 2005). On a donc étudié d'abord les caractéristiques du sens des mots en fonction de regroupements de plusieurs concepts à l'intérieur de catégories. Concernant les « objets », par exemple, on peut considérer qu'il en existe deux grandes catégories : les objets vivants et les objets manufacturés (Gregory, 1999). Chacune de ces grandes catégories se subdivise en plusieurs autres catégories et le tout se décline de manière hiérarchique (p. ex. « animal » est une sous-catégorie des « objets vivants », qui elle-même est subdivisée en catégories, comme « mammifères », etc.) La théorie des réseaux sémantiques (p. ex. Collins et Quillian, 1969) propose que les caractéristiques sémantiques des mots sont représentées suivant ces catégories hiérarchiques. Les informations sémantiques sont ainsi représentées par un vaste réseau complexe de nœuds sémantiques (ou nœuds conceptuels). C'est dans ces nœuds que sont stockées les caractéristiques sémantiques des mots. Les nœuds sémantiques sont connectés les uns aux autres suivant leur relation. Le sens d'un mot correspond ainsi à son emplacement dans le réseau, selon les relations que ce terme entretient avec les autres. Collins et Quillian (1969) proposent l'exemple de l'organisation hiérarchique suivante :

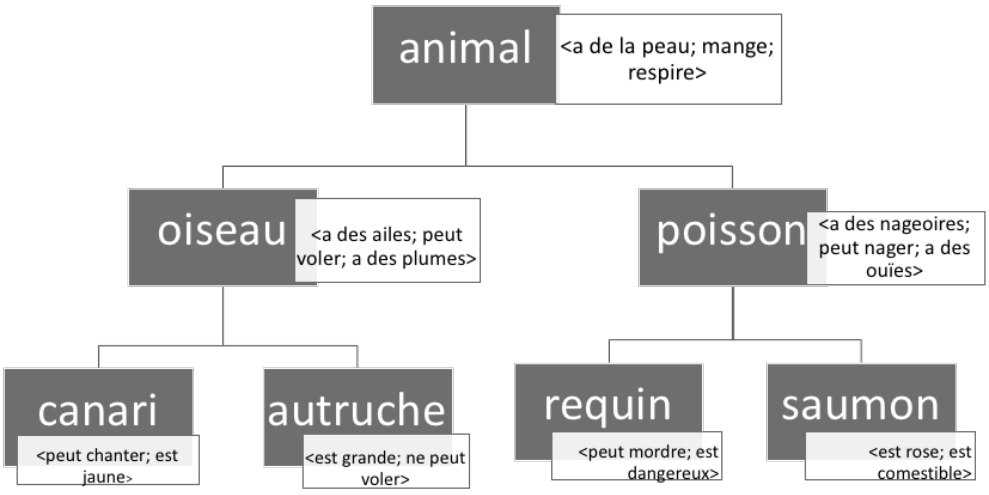


Figure 2.8 Illustration de l'organisation hiérarchique d'un réseau sémantique (Collins et Quillian, 1969, traduit par Ferrand, 2007)

Pour un concept sémantique donné, encodé par un nœud sémantique, une liste de propriétés qui s'appliquent à ce nœud sont présentées, et l'on considère que ces propriétés s'appliquent également aux nœuds qui lui sont subordonnés. Pour obtenir l'ensemble des traits sémantiques d'un mot, on doit donc inclure les traits associés directement au nœud correspondant, mais également les traits associés à l'ensemble des catégories qui lui sont superordonnées (se trouvant plus haut dans la hiérarchie).

Ceci rend bien compte du phénomène sémantique que l'on appelle l'hyponymie, c'est-à-dire le fait que certains mots entretiennent entre eux des liens hiérarchiques, étant regroupés en ensembles et en sous-ensembles. Par exemple, parmi les mots de la figure 2.11, on considère que le concept « oiseau » est hyponyme du concept « animal » et qu'il est cohyponyme de « poisson » (concept qui est également hyponyme de « animal »).

Selon la théorie des réseaux sémantiques (Collins et Quillian, 1969), pour récupérer l'ensemble des traits sémantiques d'un mot, il est nécessaire de « remonter » dans la hiérarchie pour inclure tous les traits sémantiques des concepts superordonnés. Ceci prévoit ainsi qu'il soit plus long de répondre à la question « est-ce qu'un saumon est un animal? » qu'à la question « est-ce qu'un saumon est un poisson? » Même si les données empiriques ont confirmé cela sous certaines circonstances, elles ont aussi souligné des cas pour lesquels cela ne s'appliquait pas. Ces cas ont permis de mettre en lumière le phénomène de prototype.

Ainsi, certains hyponymes d'une catégorie sont plus typiques alors que d'autres le sont moins. Par exemple, « autruche » est hyponyme de « oiseau » mais ne possède pas la caractéristique *peut voler*, alors que son cohyponyme « canari » la possède. En ce sens, « canari » est plus prototypique que « autruche ». Cela influence le temps de traitement pris par les sujets pour répondre à des questions. Ainsi, il est plus rapide de répondre à la question « est-ce qu'un canari est un oiseau » qu'à la question « est-ce qu'une autruche est un oiseau », et ce, même s'ils sont tous les deux dans la catégorie tout juste subordonnée.

Pour rendre compte de cet effet de prototype, Collins et Loftus (1975) ont ainsi abandonné les catégories purement hiérarchiques au profit d'une activation propagée (spreading-activation) dans leur modèle de la théorie des réseaux sémantiques. Ils conservent l'idée des nœuds sémantiques (ou conceptuels), mais ils modifient la manière dont ils sont connectés les uns aux autres. La connexion entre les nœuds sémantiques varie en force et en distance. Par conséquent, le lien entre « canari » et « oiseau » est plus fort et plus « proche » de « oiseau » que c'est le cas pour « autruche ». Cela leur permet de rendre compte de l'effet de prototype (un prototype est plus fortement lié et est plus proche). Cela permet également une première approche vers la capacité à rendre compte de la similarité sémantique entre des concepts; « canari » ressemble plus à « oiseau » qu'« autruche » lui ressemble.

Une autre approche de la sémantique prend plutôt comme point de départ les traits sémantiques associés à un concept plutôt que la seule notion de catégorie. La signification d'un mot est donc analysable sous la forme d'un ensemble de traits sémantiques élémentaires. La représentation de ces traits permet de rendre compte de similarités ou de distinctions entre des concepts. Par exemple, le concept « femme » contient les traits élémentaires <humain>, <adulte> et <non mâle>, alors que le concept « fillette » comporte les traits <humain>, <non adulte> et <non mâle> (Brousseau et Roberge, 2000). Cela permet ainsi de proposer une organisation des représentations sémantiques selon le partage de traits sémantiques. On postule également dans ces théories que la décomposition en traits a lieu de manière automatique et irrépressible. S'inscrivant dans cette théorie componentielle de la sémantique, Smith *et al.* (1974) proposent un modèle qui présente les traits associés aux concepts selon deux types de traits, soit les traits définitionnels et les traits spécifiques. Alors que les premiers sont obligatoires, les seconds ont un caractère facultatif – par exemple, il n'est pas obligatoire de posséder le trait <peut voler> pour faire partie du concept « oiseau » puisque les pingouins sont des oiseaux mais ne volent pas. Avec un mécanisme de comparaison de traits, Smith *et al.* (1994) rendent compte de la rapidité avec laquelle les sujets jugent de phrases comme « Un rouge-gorge est un oiseau » ou « Le cochon est un oiseau » (partage sémantique fort dans le premier cas,

ou faible dans le deuxième). Pour des phrases comme « un poulet est un oiseau » (partage sémantique ni fort ni faible), le retour aux traits définitionnels de chacun des concepts explique le temps de jugement plus long mesuré pour de telles phrases.

2.5.2 Modèle théorique cognitif de la sémantique lexicale dans l'identification des mots écrits : le modèle computationnel distribué

Ces quelques notions à propos de la distinction forme et sens, puis des représentations de la sémantique permettent de discuter de la manière dont la sémantique est impliquée dans l'identification des mots écrits. Le modèle présenté ci-dessous vise à décrire ce traitement sémantique, plus précisément quels en sont les mécanismes et leur fonctionnement.

Le modèle proposé par Plaut et Booth (2000) s'appelle le modèle computationnel distribué. Comme son nom l'indique, les représentations de la sémantique ne sont pas représentées localement – par une liste de propriétés, par exemple –, mais correspondent plutôt à un patron d'activation spécifique sur un grand nombre d'unités sémantiques (voir la section 2.4.2.3 pour les détails sur les représentations distribuées). Les unités sémantiques du modèle correspondent à des traits sémantiques (p. ex. <vivant> ou <peut voler>), tels que décrits dans les paragraphes ci-dessus. Une même unité sémantique est activée au sein du patron d'activation de tous les concepts, associés à des mots, auxquels ce trait sémantique s'applique.

Pour rendre compte de l'activation de la sémantique pendant l'identification des mots écrits, Plaut et Booth (2000) ont entraîné leur modèle computationnel à fournir la signification d'un mot présenté au système par le biais de l'écrit. Pour permettre au réseau d'« apprendre » la signification des mots, un premier aspect sémantique est encodé dans les traits sémantiques associés aux concepts présentés. Cela permet au réseau d'« apprendre » que certains mots se « ressemblent » plus que d'autres. Cette ressemblance est opérationnalisée par le nombre de traits sémantiques partagés entre deux concepts. Par exemple[7], « tulipe » et « rose » ont un grand nombre de traits sémantiques communs, alors que ce n'est pas le cas pour « tulipe » et « ruche ».

Un autre aspect sémantique est considéré dans le modèle, soit celui de la force d'association verbale. L'association verbale rend compte du fait que certains mots apparaissent souvent ensemble dans différents contextes, dans différentes circonstances; certains mots sont donc fréquemment cooccurrents

[7] Les exemples de cette section sont tirés de Bonnote et Casalis (2010).

dans des passages écrits ou oraux. Si Plaut et Booth (2000) ont choisi d'implémenter cette caractéristique dans leur modèle, c'est qu'ils relèvent plusieurs études qui documentent l'influence de cette caractéristique sur l'identification des mots écrits (voir McNamara, 2005 pour une synthèse). Cette force d'association verbale est implémentée dans le réseau par la manière dont les mots sont présentés au réseau pour son « apprentissage » des significations. Les auteurs du modèle ont ainsi présenté les mots au réseau par paires, afin de rendre compte de l'association qui peut exister entre deux concepts, sans nécessairement que ceux-ci partagent un grand nombre de traits sémantiques. Par exemple, les mots « singe » et « banane » ne partagent pas beaucoup de traits sémantiques, mais apparaissent très souvent ensemble dans différentes circonstances, donc sont fortement associés verbalement. La fréquence à laquelle deux mots ont été présentés dans la même paire au réseau reflète ainsi la force de l'association verbale qui existe entre deux concepts; les mots fortement associés comme « singe » et « banane » apparaissent plus souvent dans la même paire que le sont des mots plus faiblement associés comme « cheval » et « barrière ».

Selon ce modèle, un seul mécanisme permet de rendre compte du traitement sémantique en identification des mots écrits. Au fur et à mesure que le système cognitif obtient de l'information sur la forme du mot qu'il est en train de lire, il active certaines unités sémantiques et en inhibe d'autres jusqu'à ce que le réseau soit stabilisé sur la signification finale. Dans ce modèle, pendant l'identification des mots écrits, la modélisation de l'activation qui a cours au sein des unités sémantiques est fonction tant des traits sémantiques que de la force d'association. Ainsi, si deux concepts partagent beaucoup de traits et sont fortement associés, la configuration stable du patron d'activation est plus semblable, alors que c'est l'inverse pour deux concepts qui partagent peu de traits et sont faiblement associés. Ce faisant, on conçoit dans ce modèle qu'au cours de l'identification d'un mot écrit, le système cognitif active l'information sémantique selon les informations traitées sur la forme visuelle et la forme phonologique d'un mot. Comme l'activation sémantique a lieu de manière concomitante à l'activation de la forme, les unités sémantiques du réseau pertinentes à chaque étape du traitement de la forme seront activées, jusqu'à ce que le patron d'activation sémantique soit complètement stabilisé. Par exemple, il est fort attendu que pendant l'identification du mot « cheval », des unités sémantiques de mots dont les premières lettres sont similaires seront activées également pendant un temps, comme celles de « chevaux », « cheveux », « chevelu », « chevron », etc. ainsi que les concepts qui leur sont fortement associés.

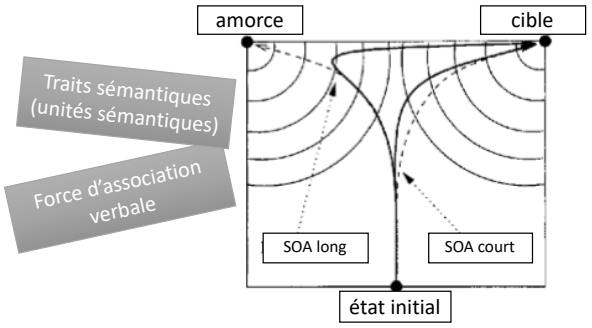


Figure 2.9 Représentation du fonctionnement du traitement sémantique dans l'amorçage sémantique selon le modèle computationnel distribué (traduction de Plaut et Booth, 2000)

2.6 Électrophysiologie de l'identification des mots écrits

Les modèles cognitifs théoriques présentés précédemment visent à décrire par quel(s) mécanisme(s) cognitif(s) l'esprit humain arrive à identifier un mot écrit. Plusieurs décennies de recherche visent également à documenter l'activité électrique émanant du cerveau pendant la lecture de mots. Ces données, obtenues à l'aide de la technique d'imagerie cérébrale de l'électroencéphalographie (EEG), fournissent de l'information à chaque milliseconde pendant l'identification de mots, ce qui permet de préciser le décours temporel des opérations mentales impliquées. Tel que mentionné à la section 2.1, cette activité électrique, si elle est considérée de manière concomitante aux données comportementales, peut apporter un éclairage intéressant à un phénomène.

Un nombre imposant de recherches portent sur l'EEG et l'identification des mots, tant oraux qu'écrits. Il convient d'abord de distinguer cette recherche doctorale des recherches en neurosciences. En effet, ces dernières ont pour objectif de documenter la structure et le fonctionnement du cerveau, plus largement du système nerveux (Baars et Gage, 2012). La présente recherche, s'inscrivant dans les neurosciences éducationnelles, a pour objectif d'utiliser ce que l'on connaît sur le fonctionnement du cerveau pour éclairer un phénomène éducatif. En ce sens, l'objectif de cette section n'est pas de faire une recension exhaustive des recherches sur les neurosciences et l'identification des mots écrits, mais bien de souligner certains éléments déjà bien documentés qui sont primordiaux aux questions de recherche.

C'est notamment la technique d'analyse des « potentiels évoqués par un événement » (*event-related potentials*; ERP) qui fournit le plus d'éléments pertinents. Cette technique consiste en la mesure de

l'activité électrique émanant du cerveau, qui est captée par un certain nombre d'électrodes (allant de quelques-unes à plus d'une centaine) apposées sur le cuir chevelu (Luck, 2005) pendant la lecture d'items qui apparaissent à l'écran. L'amplitude de l'activité électrique, mesurée à chaque milliseconde, est ensuite moyennée sur plusieurs items (p. ex. mots ou pseudomots) pour chaque électrode, et des comparaisons peuvent ensuite être réalisées sur ces amplitudes (Beres, 2017). Certaines composantes d'ERP sont très bien documentées dans la littérature scientifique, en termes de caractéristiques du signal et en termes de fonctions cognitives qui y sont associées. Une composante d'ERP correspond généralement à une déviation dans le signal électrique continu, qui est nommée en référant à une direction de l'amplitude – positif (P) ou négatif (N) – accompagnée du nombre en millisecondes (ms) où le maximum d'amplitude est généralement atteint. Les composantes N100, N2 (pour N200), P2 (pour P200), N400 et P600 sont des exemples de composantes d'ERP bien documentées dans la littérature scientifique. Étudier un phénomène cognitif, comme la lecture, à l'aide des composantes d'ERP est possible lorsque les variables qui peuvent influencer ces composantes sont relativement bien connues. Par exemple, si l'on sait comment se comporte généralement la composante N400, il est possible de comparer la composante obtenue dans différentes conditions qui visent à étudier le phénomène concerné. Pour caractériser une composante, on considère tant l'amplitude de la composante liée au sujet d'intérêt que la latence de celle-ci, c'est-à-dire le temps qui s'écoule entre la présentation du stimulus à l'écran et le moment où l'amplitude maximum de l'onde de la composante est atteinte (en anglais, ce point est appelé le *peak*).

Depuis les années '70, les études recourant à l'ERP pour documenter les composantes pertinentes sont très nombreuses en ce qui a trait à la langue orale et à la langue écrite (Beres, 2017). Pourtant, l'information que l'on peut concrètement en tirer demeure encore très fragmentée. Les synthèses réalisées sur certaines composantes particulières soulèvent souvent la question de la disparité des tâches utilisées dans les différentes études, le contrôle très variable des différentes caractéristiques linguistiques, le manque de distinction entre la langue orale et la langue écrite, le fait que les composantes soient extrêmement sensibles à des distinctions subtiles entre les tâches, etc. Hauk (2012) de même que Faisca *et al.* (2019) proposent également que les effets précoces (survenant avant 250 ms) puissent être très subtils et donc nécessiter un contrôle linguistique extrêmement fin. Par conséquent, ce qui est présenté ci-dessous se fonde sur des efforts de synthèse, trouvés dans quelques revues de littérature non systématiques.

En ce qui concerne les études sur la lecture, la composante la mieux documentée est sans nul doute la composante N400 (Barber et Kutas, 2007), une déflection négative qui arrive à son maximum environ 400 ms après la présentation d'un item à lire. Bien qu'il soit connu désormais que cette composante ne soit pas uniquement liée au langage, il demeure qu'elle a d'abord été liée à l'incongruité sémantique d'un mot à la fin d'une phrase, comme dans la phrase « Le cheval trotte dans le **ciel** », où l'incongruité est indiquée en gras. Une plus grande amplitude de la composante N400 correspond à un mot incongru présenté à la fin de la phrase (Kutas et Hillyard, 1984). Cette composante a ainsi été mise en lien avec la prédictibilité d'un mot selon le contexte qui le précède; l'amplitude de la N400 est plus faible lorsque le mot est fortement prédictible. Une plus faible amplitude de la N400 a également été documentée lorsque le deuxième mot d'une paire est sémantiquement relié au premier mot ou encore lorsqu'il lui est fortement associé (Bentin *et al.,* 1985). Ainsi, identifier un mot écrit après avoir lu un mot qui lui est sémantiquement relié entraîne une moins grande amplitude de la N400. Physiologiquement, une réduction de l'amplitude d'une composante peut être attribuable à différents éléments, qu'on ne peut distinguer actuellement : soit cela correspond à des potentiels post-synaptiques plus faibles dans les mêmes neurones, soit c'est attribuable à une activation d'une moins grande quantité de neurones, ou encore cela correspond à moins de synchronie au sein des neurones concernés.

Pour la composante N400, on a aussi associé plus récemment cette composante à la phonologie ou aux relations morphosyntaxiques dans certaines tâches particulières (p. ex. Barber et Carreiras, 2003). De ce fait, il a été suggéré, dans le domaine des langues, que la N400 est un bon « index de la facilité d'accès à l'information sémantique contenue en mémoire ainsi que de la facilité de son intégration dans le contexte local » de la tâche (Kutas et Federmeier, 2000). Ainsi, plus le mot lu est attendu selon le contexte qui précède, que ce contexte soit un mot, une phrase, un texte ou un discours, plus l'amplitude de la N400 est faible. À l'inverse, plus un item lu est inattendu selon le contexte, plus l'amplitude de la N400 est grande, reflétant notamment la difficulté à intégrer sémantiquement cet item au contexte. La composante N400 est donc sensible à plusieurs variables linguistiques, et plus les recherches se poursuivent à ce sujet, plus la portée de la N400 lors du traitement de la langue écrite semble importante (Kutas et Federmeier, 2011). On associe désormais la N400 à l'intégration de différents types d'informations accumulées pendant l'identification des mots écrits, que ce soit en lien avec la phonologie, le voisinage orthographique ou même la fréquence des mots. Les réponses de la N400 sont donc sensibles à un nombre important de variables linguistiques qui ont déjà été documentées comme influençant les temps de réponse dans des

tâches de lecture de mots/pseudomots. Certaines données particulières à cette composante en lien avec l'amorçage sémantique seront discutées à la section 2.9.

D'autres composantes sont également reconnues comme étant mobilisées dans l'identification des mots écrits. C'est le cas notamment de composantes qualifiées de précoces, c'est-à-dire des composantes dont le maximum d'amplitude survient avant 350 ms (Dien, 2009). Dans sa revue de littérature sur ces composantes précoces, Dien (2009) recense un certain nombre de variables linguistiques qui influencent les composantes d'ERP à différents délais après la présentation d'un mot à lire[8]. Ainsi, plusieurs études indiquent que la familiarité d'un mot (que ce soit une différence de sensibilité entre les mots déjà appris et les mots nouveaux ou entre les mots, les pseudomots et les items non orthographiques) influence les composantes ERP autour de 100 à 200 ms après la présentation de l'item. Les composantes ERP autour de 200 à 300 ms sont sensibles à la fréquence écrite des mots, à la lexicalité (mots/pseudomots), ainsi qu'à la classe grammaticale des mots. Les traitements phonologiques, quant à eux, semblent influencer les composantes à partir de 250 ms jusqu'à environ 450 ms après la présentation d'un item. Selon Dien (2009), tout ce qui est du ressort des représentations associées au mot complet (p. ex. la sémantique du mot) débute autour de 300 ms et se poursuit jusqu'à au moins 450 ms après la présentation de l'item, englobant ici la composante N400 dont il a été question précédemment.

Ainsi, on peut déduire que des composantes comme la N150, la P150, la N250, la P325 et la N400, qui sont toutes des composantes pertinentes à l'identification des mots écrits, voient leur amplitude et leur latence être influencées par différentes opérations mentales de ce processus cognitif. Selon Grainger et Holcomb (2009), la partie positive de la N/P150 a une distribution occipitale, plus importante dans l'hémisphère droit, et est sensible à la répétition de mots ou au chevauchement entre des éléments orthographiques communs entre deux mots (amplitude plus positive pour des mots répétés ou avec un grand chevauchement). Certaines caractéristiques visuelles des mots (sans être au niveau des représentations orthographiques) sont également pertinentes ici, comme un changement de taille de l'écriture ou de la police de caractère. La N250, quant à elle, correspond à une activation généralement plus distribuée avec une activité plus importante sur la ligne médiane et sur les sites antérieurs gauches. Cette composante semble modifiée par un grand nombre de caractéristiques orthographiques et phonologiques, ce qui

[8] Il semble primordial de noter que seules les données les plus robustes (effets obtenus par différents groupes de chercheurs) sont conservées. L'article rapporte également certaines évidences moins robustes, qui ne sont pas rapportées ici, par souci de concision.

conduit les auteurs à proposer que la N250 représente certains traitements orthographiques et phonologiques sublexicaux. La composante P325 (distribution non précisée) semble sensible à des variables qui concernent les mots complets, reflétant le fait que le traitement qui est réalisé à ce moment semble concerner les représentations visuelle (ou orthographique) et phonologique des mots entiers.

D'autres auteurs discutent de composantes similaires aussi liées à l'identification de mots écrits. Par exemple, Jucla *et al.* (2009) rapportent que la N170, en français, est une composante dont la source se situe dans la région occipito-temporale, reflétant les premiers traitements orthographiques des mots, puisqu'elle est sensible aux lettres de manière différente qu'elle l'est pour les autres symboles. Cette composante est également sensible à la fréquence et pourrait ainsi refléter l'accès lexical direct par la reconnaissance de la forme visuelle des mots. Bermudez-Margaretto *et al.* (2020), à propos de l'espagnol, indiquent que la composante P200, qui est une composante fronto-centrale, est connue pour être modulée par l'accès à la forme orthographique complète, avec des réponses plus positives pour des mots de haute fréquence écrite en comparaison à des mots de faible fréquence écrite ou à des pseudomots. Landi et Perfetti (2006) indiquent également la possibilité, moins robuste que pour la N400, que cette composante P200 puisse être influencée particulièrement par certains traitements phonologiques, pour l'anglais. Ces auteurs indiquent donc que cette composante semble être un « index de la reconnaissance de la forme (informations visuelles et phonologiques) d'un mot ».

Bien que les synthèses complètes et cohérentes soient encore attendues dans ce domaine, il semble raisonnable de résumer comme suit le décours temporel de l'identification des mots écrits, selon les études actuelles qui ont utilisé l'électroencéphalographie (EEG) :

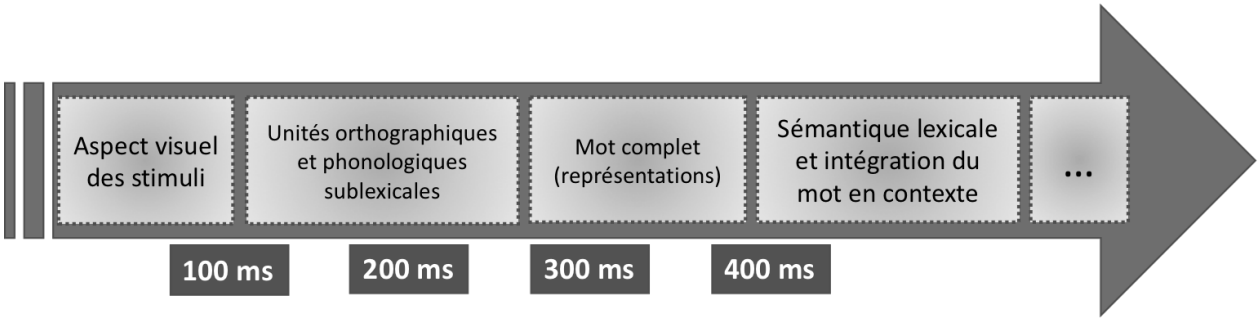


Figure 2.10 Représentation du décours temporel de l'identification des mots écrits, selon l'état actuel de la littérature scientifique dans le domaine

Bien évidemment, les éléments de cette figure ne représentent pas la profondeur de tout ce qui a été documenté pour chacune des composantes discutées dans la littérature scientifique, mais cela permet tout de même de résumer les grandes lignes des consensus actuels. Les aspects visuels du mot à lire sont ainsi traités rapidement, c'est-à-dire à l'intérieur du premier 100 ms. Ensuite, ce sont les unités orthographiques et phonologiques sublexicales qui sont traitées entre 100 et 300 ms, pour donner lieu à l'activation de la représentation du mot complet autour de 300 ms. Finalement, vers 400 ms, la sémantique lexicale du mot est activée, et l'intégration du mot au contexte est réalisée. Cette manière de résumer les connaissances issues de la littérature scientifique actuelle sous-estime, certes, le caractère hautement interactif des différentes composantes de l'identification des mots écrits, mais cet aspect sera abordé plus en détails à la section 2.9.

Bien qu'il y ait un effort commun de la part des chercheurs du domaine pour mieux articuler les données d'ERP avec les modèles théoriques de l'identification des mots écrits, il semble pertinent de mentionner qu'on n'y est toujours pas arrivé. En effet, les données actuelles les plus robustes sur les composantes de l'ERP sont encore centrées sur des phénomènes isolés; on observe souvent l'influence d'une seule variable linguistique à la fois. Ainsi, tous les modèles d'identification de mots proposés, qu'ils soient développementaux, symboliques, connexionnistes, etc., peuvent être utilisés pour discuter des données d'ERP. Comme le soulignent Taylor *et al.* (2013), les données issues des techniques d'imagerie cérébrale ne permettent pas actuellement de valider ou d'invalider un modèle en particulier, puisque tous les modèles théoriques de l'identification des mots écrits peuvent être appuyés par les données actuelles.

2.7 Les difficultés d'identification des mots écrits

Tel que mentionné à la section 2.3 concernant les processus cognitifs de lecture, l'apprentissage de la lecture chez les élèves qui ne présentent pas de difficulté se manifeste au niveau cognitif par une mobilisation distincte des processus cognitifs spécifiques et non spécifiques au fur et à mesure de cet apprentissage. Ce que comprend le lecteur en début d'apprentissage d'un texte qu'il lit dépend davantage de ses processus spécifiques, alors que pour le lecteur plus avancé, cette compréhension de texte est davantage liée à ses capacités au niveau des processus non spécifiques de compréhension orale (Catts *et al.* 2005; Francis *et al.*, 2005). Ainsi, lorsque les processus spécifiques d'identification des mots écrits sont automatisés, c'est-à-dire qu'ils opèrent de manière exacte et rapide (Laberge et Samuel, 1974 ; Savage et Kirby, 2008), davantage de ressources cognitives peuvent être dédiées à la mise en œuvre plus efficace des processus non spécifiques menant à la compréhension du texte lu (Perfetti, 1985; Sprenger-Charolles,

2016; Stanovich, 1980). En français, tenant compte des caractéristiques linguistiques qui sont propres à cette langue, on indique que les processus spécifiques d'identification des mots écrits ne sont généralement automatisés qu'à la fin de la deuxième année d'enseignement formel de la lecture, soit vers 8 ans (Seymour *et al.*, 2003; Sprenger-Charolles, 2016; Ziegler et Goswami, 2005). Par conséquent, il n'est pas attendu que les élèves de moins de 9 ans aient complètement automatisé l'identification des mots écrits.

Bien entendu, certains élèves présentent des difficultés au cours de l'apprentissage de la lecture. La nature des difficultés d'apprentissage de la lecture rencontrées varie d'un élève à l'autre, de même que les manifestations de ces difficultés (Braibant, 1994). Faire appel aux modèles cognitifs permet de distinguer différents types de difficulté de lecture.

2.7.1 Profils de lecteurs en difficulté de lecture

Une première approche permettant de les distinguer sommairement consiste à utiliser les composantes du modèle *Simple View of Reading* (SVR), de Gough et Tunmer (1986). Tel qu'indiqué précédemment, cette équation réfère à deux composantes essentielles qui interagissent pour permettre au lecteur de comprendre le texte écrit : d'une part l'identification des mots écrits, et d'autre part les processus de compréhension langagière tels qu'ils sont mobilisés à l'oral. En recourant à ces deux éléments, trois profils d'élèves en difficulté de lecture sont dégagés (Catts *et al.*, 2003; Catts *et al.*, 2006; Kirby et Savage, 2008; Stuart *et al.*, 2008; pour n'indiquer que ceux-ci), tel qu'illustré à la figure 2.12[9]. Le recours à ces deux éléments du modèle SVR forme un espace à deux dimensions dans lequel les élèves sont distribués selon l'éloignement de leur performance par rapport à une « moyenne » représentée par le point au croisement des deux axes.

[9] Il est à noter que certains auteurs ne distinguent pas la compréhension orale et la compréhension écrite. Les études auxquelles on réfère ici sont celles qui mesurent bien la compréhension langagière à l'oral.

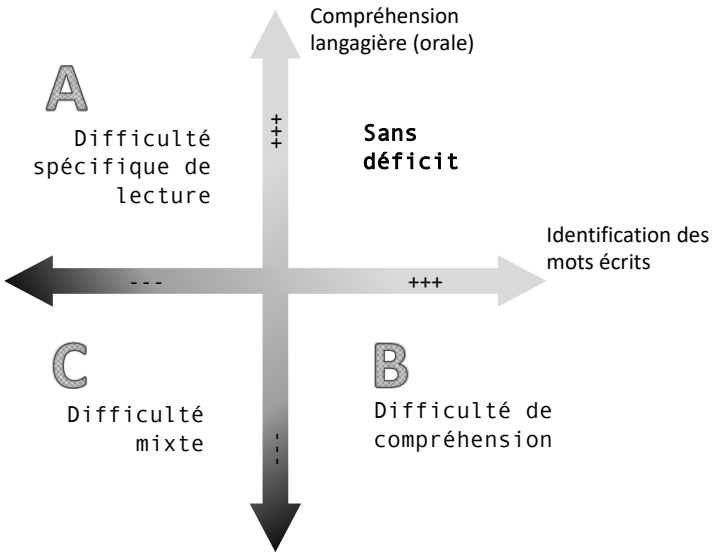


Figure 2.11 Profils de difficulté de lecture selon le modèle Simple View of Reading (traduit et adapté de Stuart *et al.*, 2008; Tunmer et Greany, 2010)

Ainsi, les élèves qui performent bien en identification des mots écrits et en compréhension orale sont situés dans le quadrant supérieur droit de la figure et font partie d'un groupe d'élèves sans difficulté. Le quadrant supérieur gauche correspond à un premier profil de difficulté (A) : ces élèves présentent des performances faibles en identification des mots écrits, mais aucun déficit n'est constaté en compréhension orale. Le second profil est présenté dans le quadrant inférieur droit (B) et correspond aux élèves pour lesquels la performance en identification des mots écrits est bonne, mais pour lesquels un déficit est constaté en compréhension orale. Puis, le dernier profil de difficulté est présenté dans le quadrant inférieur gauche (C) et correspond à des élèves dont la performance est faible tant en identification des mots écrits qu'en compréhension orale. Plusieurs recherches documentent l'existence empirique de ces profils parmi les élèves en difficulté de lecture (Catts *et al.*, 2003; voir Stuart *et al.*, 2008 pour une revue des principaux résultats par profil). Tel que le soulignent Catts *et al.* (2006), bien que le recours aux éléments du modèle *Simple View of Reading* mène à distinguer 3 « catégories » de difficultés de lecture, il importe de considérer que la performance d'un élève à chacun de ces éléments se situe sur un continuum. Ainsi, à l'intérieur d'un même quadrant, bien que l'ensemble des élèves représentés éprouvent des difficultés de nature similaire, cette difficulté peut être plus ou moins importante, et ce, pour chaque élément, résultant en des profils similaires mais non identiques.

Considérant le caractère continu de la performance des élèves sur chaque élément du modèle SVR, on peut envisager que les performances correspondant aux extrêmes soient associées à des « troubles », qui, par définition, ont une origine neurologique (Lyon *et al.*, 2003). Ainsi, parmi les élèves du quadrant A, qui présentent un déficit d'identification des mots écrits, on retrouve une certaine proportion d'élèves pour lesquels cette difficulté est attribuable à un désordre neurodéveloppemental appelé trouble spécifique d'apprentissage de la lecture (DSM-V), ou encore dyslexie[10] (Lyon *et al.*, 2003). Dans les quadrants B et C, on pourrait retrouver certains élèves qui présentent un trouble développemental du langage oral ou d'autres problématiques telles que le syndrome de Williams, l'autisme ou le syndrome de l'hémisphère droit (Nation, 2007). Puisque cette thèse s'intéresse aux élèves qui compensent une difficulté d'identification des mots écrits, il est possible dès maintenant de centrer la suite de ce cadre théorique sur les profils des deux quadrants de gauche. Plus encore, pour des raisons qui seront expliquées sous peu, ce sont les élèves du quadrant A qui sont concernés par la présente recherche. Pour reprendre l'expression de Laplante (2011), tous les élèves de ce quadrant sont désignés par le terme « élèves qui présentent une difficulté spécifique de lecture ». Cela met en lumière le fait que leur compétence orale est intacte, alors que leurs processus spécifiques d'identification de mots sont déficitaires, et ce, peu importe la gravité de ce déficit. L'existence de ce profil d'élèves est appuyée par plusieurs études, notamment celles qui précisent les manifestations des troubles d'apprentissage (p. ex. Seymour et Evans, 1999; Vellutino *et al.*, 2004), et celles qui indiquent que ce sont les processus spécifiques d'identification des mots écrits qui semblent la source principale des difficultés de lecture (sans nécessairement être des troubles), autant chez l'enfant que chez l'adulte (Desrochers, 2009; Perfetti, 1985; Siegel et Faux, 1989).

2.7.2 Manifestations des difficultés d'identification des mots écrits

Les élèves qui rencontrent uniquement une difficulté d'identification des mots écrits, même s'ils font partie d'un même quadrant dans la figure précédente, ne forment pas pour autant une catégorie homogène de lecteurs en difficulté. Au contraire, tout comme les modèles cognitifs décrits en 2.4. en rendent compte, il existe plusieurs mécanismes distincts dans l'identification des mots écrits. Ce faisant, chacun de ces mécanismes peut être altéré, à des degrés divers. Chaque modèle théorique possède sa manière de classer les difficultés d'identification des mots écrits, de les conceptualiser. Malgré cela, tous

[10] L'expression « trouble spécifique d'apprentissage de la lecture » sera préférée dans le cadre de cette thèse, par soucis de cohérence avec la littérature scientifique actuelle.

s'entendent sur le fait qu'une majorité des élèves qui rencontrent des difficultés d'identification des mots écrits présentent plus particulièrement des difficultés de décodage (Shankweiler *et al.,* 1999), c'est-à-dire de la difficulté à identifier les mots en recourant de manière séquentielle aux correspondances graphème-phonème CGP[11].

2.7.2.1 Importance des difficultés de décodage

La variable linguistique souvent utilisée pour mesurer les habiletés de décodage est la lexicalité. Celle-ci rend compte du fait que les items à identifier peuvent être des mots ou des pseudomots. L'effet de lexicalité fait référence au phénomène selon lequel les mots sont généralement mieux identifiés et reconnus plus rapidement que les pseudomots. On a documenté qu'une proportion importante d'élèves en difficulté d'identification de mots ont une lacune particulière à la lecture de pseudomots en comparaison à la lecture de mots existants; pour eux, la lecture de pseudomots est vraiment plus faible que la lecture de mots existants. Ce phénomène a été documenté surtout pour l'anglais (p. ex. Adelman *et al.*, 2014; Caravolas, 2018; Rack *et al.*, 1992; Van den Broeck *et al.*, 2010, pour ne nommer que ceux-ci) mais également pour le français (Delahaie *et al.*, 2007; INSERM, 2007; Sprenger-Charolles *et al.*, 2006).

Outre ces études sur l'effet de lexicalité particulièrement présent chez les élèves en difficulté d'identification des mots écrits, les chercheurs soulignent souvent le caractère primordial de la phonologie dans cette identification. En effet, suivant les propositions de Share (1995), plusieurs résultats de recherches appuient le fait que les enfants utilisent le recodage phonologique comme mécanisme d'auto-apprentissage de la lecture. Selon Share (1995), « chaque mot inconnu décodé de manière correcte fournit une opportunité d'acquérir l'information orthographique spécifique à un mot, ce qui est la fondation du processus d'identification des mots écrits (…) De cette manière, le recodage phonologique agit comme mécanisme d'auto-apprentissage (…) permettant à l'enfant de développer de manière indépendante ses connaissances orthographiques (p. 155, traduction libre) ». Ainsi, chez tous les élèves, il semble que la maîtrise de ce mécanisme de décodage constitue un outil très puissant et important pour l'apprentissage de l'identification des mots écrits (voir Sprenger-Charolles, 2016 pour un résumé des principaux résultats de recherche en français). Par ailleurs, et sans entrer dans le détail par soucis de concision, les données de recherche à propos d'interventions en début d'apprentissage qui visent l'enseignement-apprentissage de

[11] Il convient de noter qu'afin d'éliminer toute ambiguïté liée à la terminologie d'un modèle plutôt qu'un autre, le terme « décodage » est retenu dans la présente thèse pour indiquer la procédure de traitement qui implique le recours aux corespondances graphème-phonème.

la conscience phonémique (l'un des aspects du traitement de la phonologie) contribuent significativement à améliorer la lecture chez tous les types d'élèves (Brodeur *et al.*, 2009; National Reading Panel, 2000).

L'importance de la phonologie dans l'explication des difficultés d'identification de mots écrits est également fortement appuyée par la recherche entourant le trouble spécifique d'apprentissage de la lecture-écriture. Ce trouble est défini par Lyon *et al.* (2003) comme étant un « trouble spécifique d'apprentissage qui est d'origine neurobiologique » (p.2) et qui est caractérisé par des difficultés avec « la reconnaissance de mots », le décodage et l'orthographe. Concernant ce trouble, l'hypothèse la plus consensuelle actuellement est celle selon laquelle ce trouble serait attribuable à un déficit phonologique (Bruck, 1992; Fletcher *et al.*, 1994; Liberman et Shankweiler, 1991; Morris *et al.*, 1998; Stanovich et Siegel, 1994; voir Ramus, 2003, pour une revue des différentes hypothèses liées à la dyslexie et leur poids relatif). Ce déficit phonologique se manifeste particulièrement par une difficulté à lire des pseudomots (Herrmann *et al.*, 2006; Rack *et al.*, 1992). Ainsi, sachant que les difficultés d'identification des mots écrits semblent trouver souvent leur source dans une difficulté particulière à décoder les pseudomots, sachant qu'il en est de même pour le trouble spécifique d'apprentissage de la lecture-écriture, et connaissant le rôle important de la phonologie pour l'apprentissage de la lecture, il semble que le décodage soit un élément important à considérer dans les difficultés d'identification des mots écrits.

2.7.2.2 Précision et latence de l'identification des mots écrits

Deux mesures comportementales distinctes sont généralement utilisées pour évaluer l'identification des mots écrits : les mesures de précision de l'identification des mots écrits ainsi que les mesures de vitesse (Ferrand, 2007). Le terme « fluence » est utilisé pour désigner un ratio qui tient compte à la fois de la précision et de la vitesse. De manière générale, on calculera le nombre d'items correctement lus dans une unité de temps, par exemple en une minute.

Ainsi, en considérant ces deux mesures, certains élèves peuvent être en difficulté relativement à la précision de l'identification des mots écrits seulement, alors que d'autres peuvent l'être uniquement quant à la vitesse à laquelle l'identification est réalisée. Certains autres sont en difficulté tant au niveau de la précision qu'au niveau de la vitesse. En effet, il arrive que la seule manifestation d'une difficulté d'identification de mots se trouve dans la vitesse d'identification des mots; les élèves ont une bonne précision, mais prennent un temps anormalement long pour répondre. C'est le cas de certains élèves dyslexiques. La définition du trouble spécifique d'apprentissage de la lecture-écriture, aussi appelé

dyslexie, de Lyon, Shaywitz et Shaywitz (2003), fait ainsi directement référence à ce phénomène : « … La dyslexie est caractérisée par des difficultés avec l'identification précise et/ou fluide de mots […] » (p. 2). Cette définition indique que les élèves qui ont une bonne précision dans l'identification de mots mais qui ne le font pas de manière « fluide » sont tout de même éligibles à faire partie du groupe de personnes ayant un trouble spécifique d'apprentissage de la lecture. De plus, le fait que la lenteur de l'identification de mots soit parfois le seul témoin de difficultés est appuyé par plusieurs résultats de recherche qui mettent en évidence des phénomènes particuliers aux langues transparentes (Sprenger-Charolles *et al.*, 2000). Par exemple, certaines études soulignent que, dans les langues transparentes comme l'espagnol et même le français (en lecture seulement), le déficit phonologique associé à ce trouble d'apprentissage ne devient apparent que lorsque l'on considère les temps de réponse (Paulesi, 2001; Sprenger-Charolles *et al.*, 2010; Wimmer, 1996; Ziegler et Goswami, 2005). La lecture de pseudomots étant plus aisée dans les langues transparentes, les chercheurs proposent que les élèves qui ont un trouble spécifique d'apprentissage de la lecture peuvent arriver à une maîtrise relativement acceptable des correspondances graphème-phonème (CGP), ce qui se manifeste par des réponses précises. La performance mesurée en termes de précision atteint donc très souvent un plafond même chez des élèves en difficulté. D'un autre côté, ces élèves, même s'ils ont une bonne précision en lecture, témoignent d'une lenteur en décodage de pseudomots. La considération de la vitesse d'identification, se traduisant par la prise en compte des temps de réponse, permet également d'éviter l'effet plafond obtenu sur les scores. Bref, cela souligne donc l'importance de bien distinguer la performance des élèves en difficulté au niveau de la précision des réponses ainsi qu'au niveau de la fluence de celles-ci, c'est-à-dire le temps pris pour identifier les mots écrits.

Plusieurs arguments avancés dans cette section concernent les élèves qui présentent un trouble spécifique d'apprentissage de la lecture. Pourtant, le quadrant A inclut tous les élèves qui rencontrent des difficultés d'identification de mots sans déficit au niveau de la compréhension langagière (à l'oral), que ce soit une difficulté qui provient d'une particularité cérébrale, comme c'est le cas des troubles d'apprentissage, ou que ce soit une difficulté ne provenant pas d'une telle anomalie. Ainsi, on considère dans cette perspective que les élèves qui présentent un trouble d'apprentissage n'ont pas un comportement qualitativement différent du comportement de ceux qui ne présentent pas ce trouble. Ceci signifie que les troubles d'apprentissage sont considérés comme étant à l'extrémité faible d'une courbe normale portant sur l'habileté à identifier les mots écrits (Gilger *et al.*, 1996, cité dans Shaywitz et Shaywitz, 2005).

Quoi qu'il en soit de l'origine des difficultés d'identification des mots écrits, plusieurs études mentionnées ci-haut documentent que l'impact de ces difficultés va jusqu'à empêcher l'élève de bien comprendre un texte lu de manière autonome. En effet, tel qu'il a été présenté à la section 2.3, dans les modèles interactifs de la lecture, un processus qui est déficitaire est, par définition, non automatisé. Par conséquent, ce processus nécessite beaucoup de ressources cognitives, ce qui empêche la mise en œuvre efficace des autres processus (Perfetti, 1985). Dans le cas des lecteurs en difficulté du quadrant A, ce sont les processus spécifiques d'identification des mots écrits qui nécessitent une grande part des ressources attentionnelles et cognitives, n'en laissant pas suffisamment pour la mise en œuvre efficace des processus non spécifiques. Cela signifie que, même si les seuls processus à être déficitaires sont les processus spécifiques, il est fort probable que la mise en œuvre de tous les processus non spécifiques ne soit pas optimale non plus, ce que Laplante (2021) appelle l'effet domino. Quoi qu'il en soit, ces éléments indiquent que la plupart des élèves en difficulté d'identification des mots écrits ont une compréhension de l'écrit qui est affectée, même si leur compréhension orale n'est pas du tout déficitaire, à cause de cet effet domino enclenché par des lacunes sur le plan des processus spécifiques. Dans la plupart des cas, une difficulté au niveau de l'identification des mots écrits explique à elle seule les difficultés en compréhension écrite (Aaron *et al.*, 1999; Catts *et al.*, 2005; Kirby et Savage, 2008).

2.7.3 Profil atypique de lecteurs en difficulté

D'autres études documentent des cas d'élèves en difficulté dont le profil correspond au quadrant A, c'est-à-dire qu'ils présentent bel et bien une difficulté d'identification des mots écrits (sans difficulté de compréhension orale), mais pour lesquels la compréhension écrite est relativement normale, contrairement à ce qui est attendu selon les postulats du SVR (Gough et Tunmer, 1986). Par exemple, Braibant (1994), dans son étude sur une population d'enfants de la 2^{e} année du primaire, recense qu'environ 14% des élèves de l'échantillon ont une performance se situant sous le 30^{e} percentile en identification des mots écrits présentés de manière isolée, alors qu'ils sont au-delà du 50^{e} percentile en compréhension écrite de phrases. Plus encore, il dénote les performances plutôt remarquables de 5% des élèves de son échantillon, qui, malgré une performance en identification des mots écrits sous le 30^{e} percentile, se situe au-dessus du 70^{e} percentile en compréhension écrite. Des résultats similaires sont obtenus par Aaron (1989), Conners et Olson (1990) et Yuill et Oakhill (1991) pour ne nommer que ceux-ci.

Ces performances inattendues en compréhension écrite, en dépit d'un déficit sur le plan de l'identification des mots écrits, sont également documentées de manière robuste chez des adultes qui présentent un

trouble d'apprentissage, mais qui arrivent tout de même à réaliser des tâches de lecture complexes, adultes souvent appelés « dyslexiques résilients » dans la littérature scientifique. Dans ces études, on tente de documenter la prévalence de cette catégorie atypique de lecteurs en difficulté en mesurant la compréhension écrite de texte ainsi que la reconnaissance de mots écrits et le décodage de pseudomots. Jackson et Doellinger (2002) établissent à 3% de leur échantillon d'étudiants universitaires la prévalence de « dyslexiques résilients », alors que Binder et Lee (2012) l'établissent plutôt à 14% en éducation « de base » aux adultes.

L'idée que certaines forces des élèves pourraient venir amoindrir les effets délétères des difficultés d'identification des mots écrits se retrouve également dans certaines études, notamment celle de van Viersen *et al.* (2016). Cette étude porte sur les élèves qui présentent un trouble spécifique d'apprentissage de la lecture de manière concomitante à une douance. Ces élèves, doublement exceptionnels (*twice-exceptionality*), ont mieux réussi que les élèves ayant uniquement un trouble d'apprentissage, et ont un peu moins bien réussi que les élèves normo-lecteurs dans tous les tests de lecture réalisés, malgré des déficits au niveau de la conscience phonémique et de la dénomination rapide similaires, voire pire, que les élèves ayant un trouble spécifique d'apprentissage sans douance. Les habiletés de lecture mesurées comprennent la fluence de reconnaissance de mots écrits, la fluence du décodage de pseudomots ainsi que la fluidité en lecture orale de texte. En plus de ces mesures, des habiletés de langage oral (grammaire) et des habiletés cognitives liées à la lecture (mémoire de travail) ont été mesurées. On a ainsi documenté dans cette étude que des forces particulières sont présentes chez ces élèves doublement exceptionnels au niveau de la mémoire verbale à court terme, de la mémoire de travail et au niveau des habiletés langagières à l'oral. Welcome *et al.* (2009), quant à eux, ont documenté une performance significativement plus faible chez les lecteurs résilients adultes au niveau des tâches de reconnaissance de mots isolés, mais pas dans des tâches impliquant l'accès à la sémantique, mettant l'accent sur le rôle que la sémantique pourrait jouer dans le fait de surmonter (ou compenser) ses difficultés à identifier les mots écrits.

En somme, il semble que les conséquences des difficultés d'identification des mots écrits sur la compréhension écrite soient variables d'un individu en difficulté à l'autre. Alors que la plupart des lecteurs voient leur compréhension écrite fortement altérée par une difficulté d'identification des mots écrits, certains autres – dans une moindre proportion évidemment – n'ont que très peu de conséquences sur leur capacité à comprendre un texte écrit. Ce sont particulièrement les élèves de ce profil qui font l'objet de la présente recherche; ils sont désignés désormais par l'expression « lecteur compensateur ».

2.8 Compensation d'une difficulté d'identification des mots écrits

Les éléments empiriques rapportés précédemment documentent le fait que, généralement, un déficit d'identification des mots écrits entraîne des difficultés de compréhension écrite. Ils documentent également que pour un certain sous-groupe d'élèves en difficulté d'identification des mots écrits, ceci ne semble pas être le cas, puisque leur compréhension écrite d'un texte se situe dans la normale. Cette dernière catégorie d'élèves représente un profil atypique de difficultés de lecture, qui sont non apparentes. Ces élèves arrivent somme toute à compenser un déficit d'identification de mots écrits de manière suffisamment efficace pour que leur compréhension écrite se situe dans la normale. Il convient ainsi de se pencher sur ce phénomène de compensation d'un déficit d'identification des mots écrits en lecture.

2.8.1 Effet de contexte

Un des phénomènes les plus robustes pour documenter l'existence de la compensation comprend un grand nombre d'études sur l'« effet de contexte », ou l' « effet facilitateur du contexte ». À la fin des années 1970, deux équipes de chercheurs ont mis en évidence de manière parallèle l'effet du « contexte » sur les performances à identifier des mots écrits. Perfetti *et al.* (1979), et West et Stanovich (1978; 1979) ont en effet publié les études princeps de ce phénomène. Leurs résultats sont les suivants : les lecteurs débutants et ceux qui présentent des difficultés sont ceux pour lesquels la présence d'un « contexte » contribue à une amélioration importante de leur performance en identification de mots; alors que les lecteurs plus avancés ou sans difficulté ne voient pas leurs performances augmenter significativement en présence de ce « contexte ».

Toutes les études sur l'effet de contexte comparent les performances des sujets en identification de mots écrits présentés de manière isolée à leur performance en identification de mots écrits présentés en contexte. Le contexte proposé varie d'une étude à l'autre; parfois le contexte correspond à un texte (p. ex. Perfetti *et al.*, 1979; Schwantes, 1981; Schwantes, 1982), à un court paragraphe (p. ex. Schwantes *et al.*, 1980; West, *et al.*, 1983) à une phrase (p. ex. Nation et Snowling, 1998; Riggs *et al.*, 1984; Stanovich *et al.*, 1985; Stanovich *et al.*, 1981; West et Stanovich, 1978;) ou encore à un mot (p. ex. Becker, 1982; Schvaneveldt *et al.*, 1977; Simpson et Forster, 1985; Simpson et Lorsbach, 1983; Simpson *et al.*, 1983)[12]. La tâche proposée aux sujets est, en règle générale, soit une tâche de lecture orale des items-cibles, soit une tâche de décision lexicale (indiquer si l'item correspond ou non à un mot de la langue). Les données

[12] Il y a aussi Sprenger-Charolles (1989) qui présente un contexte en image, mais peu d'études en font ainsi.

considérées sont dans la très grande majorité les temps de réponse, c'est-à-dire le temps qui s'écoule entre le début de la présentation d'un stimulus et le début de la réponse du participant, ainsi que la précision des réponses (souvent calculée en taux de réussite).

Ainsi, à titre d'exemple, est rapportée ici plus en détail l'étude de West et Stanovich (1978), dont le contexte correspond à une phrase, en langue anglaise. Ces auteurs ont proposé à 48 élèves de 4e année (en moyenne 9 ans), 48 élèves de 6e année (en moyenne 11 ans) et 48 élèves de collège (en moyenne 20 ans) une tâche de lecture de mots. Les 54 items à lire oralement (cibles) étaient présentés dans trois conditions : 1) à la suite de la présentation d'une phrase écrite se terminant par la préposition « the » dont le contexte sémantique est congruent à la cible (p. ex. « Le chien court après le chat. »); 2) à la suite de la présentation d'une phrase écrite se terminant par la préposition « the » dont le contexte sémantique n'est pas congruent avec la cible (p. ex. « La fille s'assoit sur le chat. »; 3) sans présentation d'une phrase écrite, seulement avec la préposition « the ». Les résultats indiquent que les temps de réponse pour la lecture orale des mots-cibles sont plus rapides pour les mots-cibles précédés d'un contexte congruent (en comparaison à l'absence de contexte) pour les élèves de 4e année et de 6e année, mais que c'est le cas dans une moindre mesure pour les adultes. Des résultats similaires sont obtenus en considérant les taux d'erreurs. Qui plus est, le temps de réponse (et le taux d'erreur) augmente significativement lorsque le contexte proposé est non congruent, mais seulement pour les élèves du primaire, pas du tout chez l'adulte. On note ainsi que plus les lecteurs avancent en âge et en expertise, moins ils semblent utiliser le contexte comme support à l'identification de mots. De plus, à l'intérieur des groupes de même niveau scolaire, les auteurs rapportent une très forte corrélation négative entre les résultats obtenus à un ensemble de tests standardisés de lecture-écriture (score composite incluant l'identification des mots écrits, la compréhension écrite de phrases et la production des mots écrits) et la facilitation de l'identification des mots écrits par le contexte, et ce, chez l'ensemble des sujets; même les sujets adultes en difficulté recourent davantage au contexte que les adultes sans difficulté.

Lorsque le contexte fourni correspond à un mot, il s'agit d'un mot qui entretient ou non un lien sémantique avec le mot-cible à identifier. Ce paradigme s'appelle l'amorçage sémantique et sera détaillé à la section 2.9. Par exemple, l'étude de Schvaneveldt *et al.* (1977) présente les items-cibles dans 4 conditions : 1) l'item-cible est un mot précédé d'un mot qui lui est sémantiquement relié (p. ex. chien-CHAT); 2) l'item-cible est un mot précédé d'un mot qui ne lui est pas sémantiquement relié (p. ex. table-CHAT); 3) l'item-cible est un pseudomot précédé d'un mot (p. ex. table-TATO; 4) l'item-cible est un pseudomot précédé

d'un autre pseudomot (p. ex. bable – TATO). Contrairement à ce qui a été présenté précédemment, le contexte fourni ici concerne principalement les liens sémantiques (ou l'absence de lien) entre le premier et le deuxième mot. Dans la tâche élaborée par Schnaneveldt et ses collègues (1977), on demande aux participants d'appuyer sur un bouton lorsque l'item est un mot et sur un autre bouton lorsque l'item n'est pas un mot, ce que l'on appelle une tâche de décision lexicale. Le temps de réponse est mesuré à partir du moment où le mot-cible apparaît, jusqu'à ce que le sujet ait fourni sa réponse. La tâche a été réalisée par 24 sujets de 2e année (âge moyen de 7,6 ans) et par 24 sujets de 4e année (âge moyen de 9,5 ans). Le lien sémantique retenu par Schvaneveldt et ses collègues concerne la force d'association des items (voir section 2.5.2 de ce cadre théorique), et correspond à des mots fortement associés pour les enfants les plus jeunes (normes pour enfants de 7 ans)[13]. De cette façon, les auteurs assurent que tant les plus jeunes sujets de l'échantillon que les plus vieux ont les connaissances sémantiques nécessaires pour apprécier le lien entre les mots. Les résultats obtenus indiquent que les élèves de 2e et 4e années sont sensiblement similaires sur le gain obtenu grâce au contexte, avec des performances légèrement plus hétérogènes chez les élèves les plus jeunes. À l'intérieur de chaque groupe d'âge, il a aussi été montré que ce sont les élèves les plus en difficulté sur le plan de l'identification des mots écrits qui bénéficient le plus du contexte sémantique fourni par un mot. Encore récemment, ce paradigme est utilisé pour étudier l'identification des mots écrits. Andrews et Bond (2009) arrivent à une conclusion similaire en comparant la sensibilité au contexte chez des adultes; ce sont les adultes forts en orthographe (qui ont donc de bonnes représentations internes des mots écrits) qui sont les moins sensibles au contexte. Van der Kleij *et al.*, en 2019, ont démontré dans leur étude que les élèves dyslexiques néerlandais de la 6e année du primaire sont ceux qui sont les plus sensibles au lien sémantique entre les mots, c'est-à-dire que l'effet du lien sémantique est plus fort chez les élèves dyslexiques que chez des normolecteurs.

L'explication fournie par les auteurs pour expliquer cet effet facilitateur du contexte, reproduit dans de multiples études, dans de multiples langues, auprès de multiples populations, a trait à l'automatisation des processus d'identification des mots écrits. On indique ainsi que plus les processus d'identification des mots écrits sont automatisés, moins le recours au contexte est nécessaire, que ce contexte corresponde à un texte, quelques phrases, une phrase ou un seul mot qui est sémantiquement relié au suivant.

[13] Il convient de noter que dans d'autres études, on présente des antonymes ou des mots qui partagent un certain nombre de traits. Les caractéristiques sémantiques choisies sont donc variables mais mènent à des résultats similaires.

Afin de vérifier le lien entre l'automatisation des processus d'identification des mots écrits et une moindre dépendance au contexte pour identifier les mots écrits, Perfetti (1985) rapporte que l'effet de contexte est reproductible chez des sujets normaux qui doivent identifier des mots écrits dont la présentation visuelle est dégradée. Ces résultats indiquent que lorsque les processus d'identification des mots écrits ne peuvent être mis en œuvre de manière automatisée, même chez des lecteurs sans difficulté, cette plus grande dépendance au contexte apparaît. Plus récemment, Benuck et Peverly (2004) ont mis en lumière une plus grande dépendance au contexte lorsque des locuteurs bilingues de l'anglais et de l'hébreu doivent identifier des mots écrits en hébreu que lorsqu'ils doivent identifier des mots anglais, à cause des caractéristiques des mots hébreux qui sont phonologiquement ambigus vu les voyelles non spécifiées. Un peu de la même manière, une partie de l'étude de Ricketts *et al.* (2016) porte sur la lecture de mots anglais fortement irréguliers chez des lecteurs en première année de l'enseignement formel de la lecture. Les résultats indiquent que les lecteurs en début d'apprentissage recourent davantage au contexte fourni par une phrase pour identifier justement et rapidement des mots irréguliers en comparaison aux mots réguliers. Ainsi, tant à cause de stimuli visuels dégradés qu'à cause de mots qui comportent un défi particulier (p. ex. irrégularité), il semble que moins l'automatisation de l'identification des mots écrits est grande, plus il y a dépendance au contexte pour soutenir cette identification des mots écrits, et ce, peu importe la cause de cette absence d'automatisation. En d'autres termes, qu'il s'agisse d'un processus cognitif ou un mécanisme déficitaire, ou encore une difficulté propre au contexte de lecture ou au mot à lire, dès que l'identification des mots écrits ne se produit pas de manière automatisée, le contexte a une plus grande influence.

Pour terminer avec l'effet facilitateur du contexte, il convient de souligner, tel que l'indique Stanovich (2000) dans une recension exhaustive de l'effet de contexte, qu'il est important d'assurer que les élèves ont les connaissances nécessaires pour utiliser le contexte proposé. En effet, une condition importante pour être en mesure de recourir au contexte est d'être en mesure de traiter adéquatement ce contexte. L'étude de Nation et Snowling (1998) met exactement en lumière cette nécessité. Les chercheuses ont proposé une tâche de lecture orale de mots en fournissant ou non un contexte correspondant à une phrase. En comparant la performance d'élèves de 7 à 10 ans qui ont une difficulté d'identification des mots écrits à celle des élèves du même âge chronologique qui présentent une difficulté en compréhension orale, ces auteures arrivent à reproduire les données d'effet de contexte, mais seulement pour les élèves chez lesquels la compréhension orale est intacte. Leur conclusion est donc à l'effet que le contexte peut être

utilisé pour faciliter l'identification des mots écrits, à condition que les processus de compréhension orale ne soient pas déficitaires.

Par conséquent, il apparaît que les élèves du quadrant A de la figure 2.11 correspondent aux candidats idéaux pour étudier la compensation. En effet, ces élèves présentent des processus d'identification des mots écrits déficitaires et n'ont pas de lacune au niveau de la compréhension orale, c'est-à-dire pour lesquels on peut attester que leur compréhension orale est intacte.

2.8.2 Le modèle interactif-compensatoire (Stanovich, 1980)

Avec l'avènement des modèles interactifs plutôt que les modèles strictement ascendants ou strictement descendants, des explications à ce phénomène d'« effet de contexte », davantage présent chez les élèves en difficulté sur le plan de l'identification des mots écrits, sont devenues possibles. En effet, tel que le souligne Stanovich (1980), ce n'est pas obligatoire que les processus de plus bas niveau (p. ex. identification de mots écrits) aient complété entièrement leur traitement avant que les processus de plus haut niveau ne soient mis en œuvre (p. ex. analyse syntaxique). Stanovich (1980), s'inspirant des propositions de Rumelhart (1977), propose également que le recours à différentes sources de connaissances, notamment les connaissances sémantiques (il réfère également à d'autres connaissances, comme les connaissances orthographiques, syntaxiques, etc.) puissent être mobilisées à n'importe quel moment. Ainsi, ce principe général d'interactivité de la mise en œuvre des processus cognitifs et des connaissances pertinentes permet à Stanovich (1980; cité dans Stanovich, 2000) de proposer le modèle interactif-compensatoire. Dans ce modèle, on propose une hypothèse de compensation, selon laquelle un processus cognitif de n'importe quel niveau peut compenser un déficit affectant un processus à n'importe quel autre niveau (Stanovich, 2000, p.36). Ce faisant, cette hypothèse de compensation indique que les processus cognitifs de plus haut niveau (ainsi que les connaissances s'y rattachant, au sens de Perfetti et Stafura, 2014) pourraient théoriquement compenser un déficit quant à un processus de plus bas niveau. Dans le cas concerné par la présente recherche, cela signifie qu'il existe, depuis les travaux de Stanovich (1980), la possibilité théorique qu'un élève qui présente un déficit d'identification des mots écrits puisse recourir à des connaissances et des processus de plus haut niveau pour compenser ledit déficit.

Plus encore, le modèle interactif-compensatoire réfère aux travaux de Posner et Snyder (1975; dans Stanovich, 2000) pour discuter de l'effet de contexte sémantique, tout comme le fait Neely (1976, 1977). En résumé, appliquée à l'effet de contexte en lecture, la théorie de Posner-Snyder sur l'anticipation stipule

que le contexte sémantique affecte l'identification des mots écrits de deux manières distinctes, correspondant à deux mécanismes distincts (voir aussi Meyer et Schvaneveldt, 1971). Le premier mécanisme correspond à une activation automatique d'informations emmagasinées en mémoire, c'est-à-dire ici une activation automatique des informations sémantiques « proches » de celles qui sont associées à un mot lu. Cette activation d'informations sémantiques liées au mot lu en premier facilite ainsi l'identification d'un mot écrit présenté subséquemment puisque l'information sémantique de ce mot est préactivée dans le cas d'un mot lié, d'où l'effet facilitateur de la sémantique. Ce mécanisme d'activation automatique agit rapidement, n'est pas possible à contrôler consciemment, et ne mobilise pas (ou peu) les capacités attentionnelles. Le mécanisme d'attention consciente, quant à lui, est différent. Il répond au contexte précédant le mot-cible en dirigeant consciemment le traitement de l'information vers le mot attendu, le mot anticipé. Ce mécanisme est donc plus lent à opérer, est consciemment contrôlé et mobilise les capacités attentionnelles. C'est la direction consciente de l'attention sur le contexte qui permet un effet facilitateur de la sémantique puisque le système cognitif est centré sur ce qui est attendu, anticipé. Ce deuxième mécanisme crée également un effet d'interférence, c'est-à-dire une moins bonne performance lorsque les mots ne sont pas liés sémantiquement, puisque dans ces cas, on doit cognitivement ramener le processeur initialement envoyé dans une mauvaise direction pour traiter le mot-cible non lié. La théorie de Posner-Snyder (1975) fournit ainsi une première tentative d'explication sur le fonctionnement précis du mécanisme de la compensation en lecture : quand l'identification des mots écrits isolés est rapide, seule l'activation automatique est responsable de l'effet facilitateur; mais quand l'identification des mots écrits est lente, le mécanisme d'attention consciente a le temps d'opérer et de fournir une autre source de facilitation contextuelle.

En somme, le modèle interactif-compensatoire propose que l'identification des mots écrits puisse être, dans le cas où il y a un déficit à ce niveau, soutenue par le traitement des informations contenues dans un « contexte ». Un mécanisme compensatoire intervient donc à ce moment, correspondant à une plus grande utilisation du contexte, de manière automatisée et non consciente, ou de manière consciente. Puisque le mécanisme d'attention consciente nécessite des ressources attentionnelles, il pourrait être possible que les élèves qui y recourent aient moins de ressources cognitives disponibles pour la mise en œuvre des processus de plus haut niveau. En effet, on considère qu'il y a une réallocation vers l'identification des mots écrits des ressources attentionnelles et des ressources liées à la mémoire, alors que ces ressources seraient habituellement dirigées vers l'élaboration du modèle de situation (afin de mettre en œuvre le mécanisme compensatoire). Il y a donc une forme de troc (en anglais, un *trade-off*)

entre les processus; soit on alloue les ressources cognitives mentionnées ci-haut aux processus spécifiques d'identification des mots écrits, soit on en alloue aux processus non spécifiques de compréhension langagière. Cette dernière idée est beaucoup mieux décrite dans le modèle présenté à la prochaine section.

2.8.3 Le modèle d'encodage compensatoire de Walczyk (2000)

Outre le modèle interactif-compensatoire de Stanovich (1980), un auteur a récemment formulé une proposition pour une extension au modèle précédent, le *compensatory-encoding model* (CEM) ou, en français, le modèle d'encodage compensatoire (Walczyk, 2000). Cette proposition reprend des éléments d'autres modèles des processus cognitifs en lecture et ajoute certains éléments spécifiques à la compensation.

Reprenant dans sa totalité la théorie de l'efficacité verbale proposée par Perfetti (1985) qui est présentée à la section 2.3, le CEM vise à préciser quels sont les mécanismes mis en place par certains lecteurs pour compenser une difficulté de lecture, quelle qu'elle soit. Pour Walczyk (2000), les élèves en difficulté de lecture ont au moins un processus de lecture déficitaire. Il précise que quand ce processus déficitaire fournit une information incomplète, erronée et/ou lente à la mémoire de travail, cela entraîne une « confusion ». Le CEM propose donc que les lecteurs en difficulté tentent par tous les moyens à leur disposition de surmonter cette confusion, pour : 1) arriver à une compréhension écrite littérale la plus complète possible du texte à lire; 2) diminuer le plus possible la distraction causée par le processus déficitaire, et ainsi ne pas trop décentrer l'attention de l'élaboration du modèle de situation. C'est en référence à ces principes que Walczyk, Marsiglia, Johns et Bryan (2004) proposent une définition de la compensation en lecture : la compensation désigne « les actions du lecteur, intentionnelles ou non » (p. 49), qui lui permettent, soit d'assister un processus déficitaire en lui allouant suffisamment de ressources cognitives pour qu'il puisse résoudre son problème; soit d'éviter complètement ce processus et fournir l'information adéquate en passant par un tout autre processus. Tout comme dans le modèle précédent, on indique ici que la compensation en lecture n'est pas toujours consciente. De plus, cette définition souligne qu'il existe deux manières distinctes de compenser. La première manière implique de s'arrêter lorsqu'une difficulté est rencontrée – par exemple quand il y a un mot que l'on n'arrive pas à identifier – et d'allouer plus d'attention et utiliser davantage de mémoire de travail pour ce processus. En faisant cela, le lecteur souhaite que le processus déficitaire arrive à fournir une meilleure réponse. L'autre manière de compenser sa difficulté est de recourir à un tout autre processus. Par exemple, devant un mot qu'il n'arrive pas à identifier, le lecteur pourrait recourir à un processus de plus haut niveau, comme l'inférence ou les

c'est-à-dire qu'ils n'apparaissent pas nécessairement ensemble dans le discours[14]. En comparaison à des paires d'items qui ne sont pas liés sémantiquement, les paires d'items qui partagent plusieurs traits entraînent des temps de réponse significativement moins longs chez les adultes lecteurs experts (Ferrand, p. 373). Différents types de relations sémantiques ont été étudiés à travers le temps, et ont entraîné ces résultats. Lucas (2000) recense un plus fort effet d'amorçage pour les relations fonctionnelles (p. ex. voiture-conduire) que pour les autres relations sémantiques, telles que les synonymes, les antonymes, les paires coordonnées et les relations de script (théâtre-pièce), qui, elles, témoignent toute d'un effet d'amorçage sémantique similaire. D'autres aspects sémantiques des mots comme l'imageabilité d'un concept, la richesse sémantique (nombre de traits sémantiques) ou la concrétude, ont également été étudiés et ont donné lieu à certaines nuances relativement à la plus ou moins grande force de l'effet d'amorçage sémantique. Comme l'objectif de cette section n'est pas d'être exhaustif, ces résultats ne sont pas rapportés en détails ici. Bien que la méta-analyse de Lucas (2000) mène à des résultats légèrement différents, Hutchison (2003) et McNamara (2005) démontrent qu'il existe également un effet d'amorçage sémantique purement associatif, même en l'absence de relation sémantique. C'est notamment le cas en amorçage médié, technique qui consiste à présenter deux items qui ne sont pas directement liés, mais qui le sont par l'entremise d'un autre mot (qui n'est pas présenté), p. ex. « lion » est lié à « rayures » par l'entremise du mot « tigre ». Quoi qu'il en soit, on documente surtout que l'effet d'amorçage sémantique est encore plus important lorsque les deux mots de la paire sont fortement associés verbalement en plus de partager un nombre important de traits sémantiques. Les auteurs désignent ce phénomène par le terme « associative boost » - une forme de renforcement d'un lien sémantique par l'association verbale. Ainsi, il semble que tant le partage de traits sémantiques que la force d'association ont un rôle à jouer dans cet effet de l'amorçage sémantique (Ferrand, 2007).

De plus, tel que le propose la théorie générale de Posner et Snyder (1975), l'amorçage sémantique semble être le résultat de processus autant stratégiques (volontaires) qu'automatiques (involontaires; voir la synthèse de McNamara, 2005). Les processus stratégiques correspondent à des stratégies conscientes et contrôlées, qui agissent lentement, qui nécessitent une intention, et que les lecteurs appliquent volontairement pendant la lecture. Les processus automatiques sont généralement rapides, procèdent sans que le lecteur en soit conscient, et sans intention de sa part; les lecteurs n'appliquent donc pas ces stratégies de manière volontaire. Certaines données obtenues avec l'amorçage sémantique ne peuvent

[14] Les exemples proviennent de l'article de Ferrand et New (2003).

pas être expliquées uniquement par des processus stratégiques. En effet, si une seule paire de mots liés est présentée à travers une longue liste, l'amorçage sémantique a un effet, et ce, même si le lecteur ne peut pas consciemment se douter que certains mots peuvent éventuellement être liés sémantiquement. Aussi, si l'amorce est présentée en vision périphérique et que les lecteurs sont avertis de ne pas y porter attention, l'amorçage sémantique a tout de même lieu. Ces éléments sont difficilement explicables par la mise en œuvre consciente par le lecteur de processus stratégiques; cela met donc en évidence le fait que l'effet d'amorçage sémantique peut être le fruit uniquement de processus automatiques.

Par ailleurs, deux processus stratégiques sont documentés : 1) l'anticipation sémantique et 2) le pairage sémantique. L'anticipation correspond à la génération active de candidats potentiels du mot-cible qui s'en vient. Le deuxième processus correspond à la recherche – après la présentation de la cible – d'un lien sémantique entre les deux mots. Les processus stratégiques sont généralement reconnus comme créant un effet d'interférence dans l'amorçage (Neely, 1976). Par exemple, avec l'anticipation, le lecteur active des candidats potentiels qui sont liés sémantiquement à l'amorce. Lorsque la cible ne fait pas partie de ces pré-activations parce que celle-ci n'est pas reliée sémantiquement à l'amorce, les réponses sont plus longues en comparaison à une situation où l'anticipation n'est pas du tout possible (p. ex. en recourant à une condition neutre où l'amorce est une séquence de « x ») [15]. Cette interférence est petite, voire inexistante, quand le temps de présentation entre l'amorce et le mot-cible (*onset stimulus asynchrony*, ou SOA) est court, soit un délai de moins de 300 ms. Par conséquent, on peut affirmer que les processus stratégiques d'anticipation et de pairage sémantique n'ont pas le temps d'être mis en œuvre lorsque le délai entre les mots est court. Dans ces cas, où le SAO est court (moins de 200 ms), c'est plutôt uniquement un effet de facilitation que l'on retrouve fréquemment (Neely, 1991), c'est-à-dire que la présentation d'un mot-cible lié sémantiquement à l'amorce rend plus rapide l'identification de ce mot-cible, en comparaison à une amorce neutre (une séquence de « x »). Dans sa synthèse, Ferrand (2007) résume les études portant sur le partage de traits entre l'amorce et la cible en indiquant qu'à des SOA de 500 ms et plus, les effets sémantiques sont le résultat d'une combinaison de facilitation (venant des paires d'items reliés) et d'interférence (venant des paires d'items non reliés) lors du recours à une condition neutre. Lorsque le SOA est de moins de 250 ms, l'effet de facilitation demeure, mais l'effet d'interférence disparait. Il en va

[15] Il convient de noter ici que, chez les lecteurs experts, seuls les temps de réponse sont discutés, parce que les scores de réussite dans ces tâches sont trop élevés, créant ainsi un effet plafond. Chez le lecteur en apprentissage, il pourrait en être autrement et l'on pourrait discuter ces effets de facilitation et d'interférence en termes de précision de l'identification des mots écrits.

de même pour les amorces et les cibles fortement associées verbalement, à l'exception du fait que l'interférence demeure toujours stable, peu importe le SOA (l'interférence ne diminue pas, ni ne disparait).

Au-delà des processus automatiques documentés dans certaines circonstances d'amorçage sémantique, certaines études ont également démontré qu'il est possible de mesurer un important effet d'amorçage sémantique dans des conditions de présentation subliminale de l'amorce, c'est-à-dire en présentant l'amorce de manière très brève et intensément masquée. De cette manière, les participants rapportent qu'ils n'ont pas conscience de la présence (ou de l'absence) d'une amorce ou ils ne reconnaissent pas l'identité de celle-ci. Malgré cette présentation subliminale de l'amorce, plusieurs études documentent un effet important lorsque le mot-cible est précédé d'une amorce liée sémantiquement (p. ex. Durante et Hirshman, 1994; Gaillard *et al.,* 2006; Marcel, 1983).

En ce qui concerne les données d'amorçage sémantique chez les enfants, elles ont déjà été principalement rapportées dans la section portant sur le robuste « effet de contexte ». Ainsi, il a été documenté que ce sont les enfants les plus jeunes ou les plus en difficulté de lecture qui bénéficient le plus d'un lien sémantique entre l'amorce et le mot-cible (Stanovich, 2000). Les effets d'amorçage sémantique sont donc de plus grande ampleur chez l'enfant que chez l'adulte (McCauly *et al.* 1976; Schvaneveldt *et al.*, 1977; Simpson et Lorsbach, 1983); les données tendent à démontrer que l'effet de la sémantique diminue au fur et à mesure de l'apprentissage de la lecture. Par ailleurs, on n'arrive pas toujours à mettre en évidence cet effet de l'amorçage sémantique lorsque le SOA est trop court (p.ex. Nievas and Justicia, 2004). Cela semble signifier que les enfants plus jeunes, pour lesquels les processus d'identification des mots écrits ne sont sans doute pas automatisés, pourraient ne pas être en mesure de bénéficier du lien sémantique s'ils n'ont pas le temps de traiter l'entièreté du mot-amorce avant la présentation du mot-cible. Ainsi, certains chercheurs n'ont pas trouvé d'effet d'amorçage sémantique dans une décision lexicale avec des SOA autour de 250 ms (p. ex. Nievas et Justicia, 2004) chez des enfants de 10 ans. Holderbaum et de Salles (2011) n'ont pas trouvé cet effet sous 300 ms, chez des élèves de 3^{e} année, en espagnol.

2.9.2 Amorçage sémantique et données électrophysiologiques

D'un point de vue méthodologique, les données comportementales (précision des réponses et temps pris) sur l'amorçage sémantique permettent de considérer des différences de gain obtenu en recourant à la sémantique. En effet, un groupe qui réussit significativement mieux l'identification de mots écrits lorsque ceux-ci sont précédés d'un mot sémantiquement relié est reconnu comme étant plus « dépendant » de

cette sémantique pour identifier les mots. Cette plus grande dépendance se traduit également dans les temps de réponse. Ainsi, les mots précédés d'un item sémantiquement relié sont identifiés plus rapidement que ceux qui ne sont pas précédés d'un tel item, particulièrement chez un groupe qui bénéficie davantage de la sémantique pour les raisons mentionnées plus haut (notamment le manque d'automatisation des processus spécifiques). Les temps de réponse dont il est question ici concernent le moment où le participant appuie sur un bouton pour signaler sa réponse; l'item est-il un vrai mot? – l'item appartient-il à la catégorie recherchée? Quoi qu'il en soit, cette pression de bouton est le résultat d'une prise de décision par le participant : celui-ci identifie l'item d'abord, puis choisit si cet item fait partie de la catégorie recherchée, pour finalement exécuter le geste moteur inhérent à la pression du bouton. Les temps de réponse mesurés par la pression d'un bouton dans les tâches d'amorçage sémantique incluent ainsi une panoplie de processus liés à la prise de décision et au mouvement. Les données électrophysiologiques des composantes bien connues de l'identification de mots peuvent aider à désambiguïser les différents processus, en ciblant le premier moment d'accès à une catégorie de représentation (p. ex. mot fréquent), et ce, avant même que le participant ait fourni sa réponse par la pression du bouton. En effet, les composantes d'ERP mentionnées précédemment culminent à partir de 150 ms, jusqu'à 400-500 ms après la présentation d'un item à identifier, et reflètent une caractéristique bien précise qui est manipulée. La pression du bouton, quant à elle, est mesurée à environ à 600-800 ms chez des adultes, et inclut toutes les étapes inhérentes au processus de décision et de mouvement. La prise de décision est par ailleurs nécessairement distincte entre les deux catégories d'items d'une tâche. Luka et Van Petten (2014) ont démontré que, chez l'adulte, les mesures de temps de réponse sont moins sensibles que les mesures d'ERP. En effet, devant des items fortement reliés sémantiquement ou faiblement reliés sémantiquement, les temps de réponse ne divergeaient pas significativement, alors que les données d'ERP mettaient en lumière cette nuance. Quoi qu'il en soit, il semble pertinent de considérer ici l'activité électrophysiologique, notamment les composantes liées à la sémantique.

La composante N400, qui est une composante qui concerne notamment l'activation de l'information sémantique d'un mot lu, emmagasinée en mémoire à long terme, a évidemment été énormément étudiée en lien avec l'amorçage sémantique depuis les années 1970. D'abord largement étudiée chez l'adulte, plusieurs études concernent désormais les enfants. Tel que mentionné dans une section précédente, cette composante est probablement l'une des plus robustes et des mieux documentées des composantes de l'approche par potentiels évoqués (ERP). Les premières études sur la N400 se sont centrées sur la détection d'une anomalie sémantique présentée à la fin d'une phrase, chez des lecteurs adultes; par exemple « Le

cheval trotte dans le nuage. » (Hillyard et Kutas, 1983; Kutas et Hillyard, 1980). On a ensuite documenté à quelles manipulations de variables linguistiques cette composante était sensible, et on a tenté de documenter les résultats comportementaux et électrophysiologiques associés à ces manipulations (Kutas et Federmeier, 2011). On a ainsi découvert que la réponse électrophysiologique mesurée par la N400 était très fortement corrélée à des mesures d'anticipation du mot dans la phrase, aussi appelées *cloze probability* (Debruille, 2007); plus un mot est attendu dans le contexte de la phrase, moins l'amplitude de la N400 est importante. À l'inverse, moins un mot est attendu (p. ex. « nuage » dans la phrase précédente), plus l'amplitude de la N400 est importante. Dans une perspective développementale, il semble que la composante N400 apparaisse à l'oral chez des enfants très jeunes; on rapporte l'effet d'incongruence à la fin des phrases orales à 19 mois environ (Friedrich et Friederici, 2006).

À la suite de ces recherches, les études s'inscrivant dans ce champ ont découvert que la N400 n'était pas uniquement présente en lecture de phrase, mais qu'il était aussi possible de la mesurer pour un item non attendu dans une liste de mots ou encore dans des expérimentations s'inscrivant dans des paradigmes d'amorçage lexical, notamment dans l'amorçage sémantique. Dans ce paradigme, il a été documenté que les paires d'items reliés sémantiquement entraînent une N400 de moins grande amplitude que les paires d'items non reliés, qui, elles, entraînent une réponse de la N400 dont l'amplitude est plus importante (voir Kutas et van Petten, 1988 pour une revue systématique des données empiriques chez l'adulte). Kutas et Federmeier (2009) soulignent que l'amplitude de la N400 varie selon les mêmes variables linguistiques qui sont documentées comme influençant les résultats comportementaux de temps de réponse. Ainsi, il semble que l'information fournie par le contexte serve à préactiver des traits sémantiques des mots potentiels (Kutas et Federmeier, 2000). Ces mêmes constats ont été dégagés chez l'enfant à l'oral (p. ex. Atchley *et al.*, 2006; Hahne *et al.*, 2004; Juottonen *et al.*, 1996) et à l'écrit (p. ex. Ackermann *et al.*, 1994; Coch et Holcomb, 2003; Schulz, 2008).

Les latences de la N400, c'est-à-dire le délai qui existe entre le moment où le mot-cible est présenté et le moment où la composante N400 atteint son maximum d'amplitude négative, sont relativement stables à travers les différentes manipulations réalisées dans les tâches d'amorçage sémantique chez l'adulte, pour des raisons qui sont encore mal connues (Kutas et Federmeier, 2011). La topographie est, elle aussi, difficile à documenter précisément, puisque plusieurs phénomènes semblent se chevaucher à ce moment dans le décours temporel, en plus du fait que l'activité électrique semble sensible aux différentes tâches à

réaliser ou encore à la modalité (auditive ou visuelle). Cela étant, la N400 en amorçage sémantique semble généralement être localisée en région centro-pariétale ou fronto-centrale (Mehlhase *et al.*, 2020).

La littérature scientifique sur la N400 dans l'amorçage sémantique et les élèves en difficulté souligne une amplitude généralement moins grande de la N400 sur les items qui ne sont pas sémantiquement reliés, en comparaison aux élèves sans difficulté. Ainsi, Kemény *et al.* (2018) indiquent que ce sont les élèves (dont l'âge moyen est de 9 ans) qui sont faibles en orthographe qui ont les amplitudes de la N400 les moins négatives. Hasko *et al.* (2013) rapportent une absence de différence entre l'amplitude de la N400 provoquée par les mots et les pseudomots chez les élèves dyslexiques en comparaison aux contrôles. Il en est de même pour Ackermann *et al.* (1994). Ces données sont plus ou moins consistantes à travers les études, puisque certaines d'entre elles ne trouvent pas cette différence d'amplitude entre élèves en difficulté et les élèves contrôles (p. ex. Landi et Perfetti, Bakos *et al.,* 2018; Silva-Pereyra *et al.,* 2003, etc.) Il apparaît que la grande variété des tâches utilisées, de même que le manque de différenciation entre tâches orales et écrites, en plus du contrôle très variable des caractéristiques linguistiques des items à travers les études, peuvent toujours expliquer cette inconsistance des données.

Par ailleurs, une différence de latence de la N400 semble caractériser plus systématiquement les élèves qui ont des difficultés d'identification de mots écrits. En effet, Schulz et coll. (2009) documentent une composante N400 plus tardive chez des élèves du primaire qui présentent un trouble spécifique d'apprentissage de la lecture-écriture en comparaison à des élèves typiques de même âge chronologique. De la même manière, Jednórog *et al.* (2010) rapportent une N400 d'amplitude similaire à celle des contrôles mais dont la latence est significativement plus grande, chez des élèves dyslexiques d'environ 10 ans. Par ailleurs, dans une tâche centrée sur le traitement de l'orthographe d'un mot (donc pas dans l'amorçage sémantique), Mehlhase *et al.* (2020) documentent chez des élèves d'environ 10 ans une composante N400 de même amplitude que celle des élèves sans difficulté, dans une fenêtre s'étalant de 280 à 440 ms. Pour expliquer cette absence d'effet de la N400, les auteurs suggèrent que cette sensibilité arrive probablement plus tard que cela pour les élèves en difficulté.

Quoi qu'il en soit, un relatif consensus existe actuellement concernant le fait que la N400 reflète sans doute une activation sémantique post identification du mot (Hauk *et al.* 2012), comme une forme de réanalyse de la sémantique (ou d'autres caractéristiques) une fois que le mot a été complètement identifié (Pulvermüller *et al.*, 2009). Cette idée est appuyée sur le fait qu'un nombre important de variables

linguistiques sont entrelacées à ce moment du décours temporel; les mots mais aussi les pseudomots ainsi que des chaînes de caractères semblent pouvoir influencer la N400 dans certaines circonstances (voir Barber et Kutas, 2007 pour une revue). Ceci implique donc que les caractéristiques phonologiques et orthographiques des mots influencent également la composante N400, de la même manière que la sémantique peut le faire. La composante N400 est ainsi associée à la sémantique lexicale, mais particulièrement à l'aisance avec laquelle on peut récupérer l'information sémantique du mot complet en mémoire à long terme selon le contexte de l'identification du mot (Swaab *et al.*, 2012). Plus encore, certains associent plutôt la N400 à l'intégration des informations phonologiques et orthographiques en une représentation lexicale ou sémantique (Coch *et al*, 2012, cité dans Mehlhase *et al.*, 2020).

Ces données de recherche à propos de la N400 mettent l'accent sur l'interactivité des mécanismes d'identification des mots écrits plutôt que sur la sérialité d'informations modulaires. On conçoit ainsi que les informations visuelles (orthographiques), phonologiques et sémantiques peuvent être activées quasi-simultanément très tôt dans le décours temporel (voir la section 2.4 pour la discussion des modèles théoriques). Par conséquent, plusieurs chercheurs se sont penchés sur la possibilité d'une activation précoce de la sémantique pendant l'identification de mots écrits, soit bien avant la fenêtre temporelle associée classiquement à la composante N400, qui se situe entre 300 et 500 ms.

Ainsi, plusieurs résultats de recherche appuient la possibilité d'une activation précoce de la sémantique lors de l'identification des mots écrits. Dans l'état actuel des connaissances scientifiques, toutes ces études, plus récentes, ont été réalisées chez l'adulte. Demeure encore à vérifier si ces constats sont les mêmes dans une population d'enfants. Néanmoins, il convient d'évoquer ici les données qui appuient cette possibilité. En situation de lecture de mots isolés, Loeches *et al.* (2001) ont documenté un effet de catégorisation sémantique environ 250 ms après la présentation des stimuli. Del Prado Martin et ses collaborateurs (2006) ont mis en évidence une distinction sémantique environ 200 ms après la présentation des stimuli. Zheng (2008) trouve aussi des effets de variables sémantiques à l'intérieur du premier 200 ms; Wirth (2008), Cavalli *et al.* (2016) et Hauk *et al.* (2012) également. Segalowitz et Zheng (2009) documentent l'effet de facilitation sémantique environ 168 ms après la présentation des stimuli, etc. Ces études, bien qu'elles n'aient pas encore fait l'objet de méta-analyses, indiquent que cette possibilité de l'intervention précoce de la sémantique bien avant la fenêtre temporelle classique de la

N400 est bel et bien pertinente. Le décours temporel documenté ici correspond ainsi à une activation presque simultanée des informations à propos de la forme du mot (traitement des informations orthographiques et phonologiques) et de la sémantique de celui-ci.

2.10 Questions spécifiques et hypothèses de recherche

Il convient de rappeler que l'objectif général de la présente recherche est d'étudier la compensation en lecture, qui est opérée par une certaine proportion d'élèves qui ont une difficulté sévère d'identification des mots écrits sans que leur compréhension écrite ne soit drastiquement affectée. Le cadre théorique a traité plusieurs aspects pertinents à cet objectif.

Le cadre théorique présenté dans ce chapitre permet de constater que ce sont à la fois les habiletés de compréhension écrite et les habiletés d'identification des mots écrits qui devraient être considérées pour repérer les lecteurs compensateurs. Une première question spécifique peut donc être posée : Comment repérer les lecteurs compensateurs à l'aide de mesures de compréhension écrite et d'identification des mots écrits ?

À la lumière du cadre théorique, une deuxième question de recherche peut également être formulée : le traitement sémantique des lecteurs compensateurs est-il mobilisé de manière particulière en identification de mots écrits présentés de manière isolée. Vu la contribution potentielle de l'EEG à l'étude de la sémantique en identification des mots écrits, il convient de scinder cette hypothèse en deux sous-questions, l'une ayant trait aux mesures comportementales, et l'autre ayant trait aux mesures liées à l'EEG. Ces sous-questions appellent ensuite la formulation d'hypothèses. La première hypothèse, concernant les données comportementales, est que l'effet facilitateur d'un lien sémantique entre l'amorce et la cible devrait se manifester de manière plus forte chez les lecteurs compensateurs que chez les contrôles, et ce, autant lorsque l'amorce est présentée rapidement que lorsque l'amorce est présentée plus longuement. En d'autres termes, même sous une contrainte de temps, les lecteurs compensateurs vont avoir tendance à recourir à la sémantique lexicale. La seconde hypothèse, concernent les données d'EEG, et met de l'avant qu'il est possible que les lecteurs compensateurs activent plus tôt l'information sémantique pendant l'identification des mots écrits que ne le font les lecteurs contrôles, et ce, malgré un déficit du traitement de la forme visuelle et phonologique des mots.

Afin d'assurer un contraste suffisant entre le groupe de lecteurs compensateurs et les autres, il semble judicieux de sélectionner des élèves de la fin du primaire. En effet, en début d'apprentissage, l'identification des mots écrits concourt pour une très large part de la variance à la compréhension écrite. Ainsi, puisque ces processus spécifiques ne sont pas automatisés, même chez le lecteur qui ne présente pas de difficulté, la performance en compréhension écrite est fortement liée à la performance en identification des mots écrits, pour tous les élèves. Par conséquent, les tâches scolaires proposées en classe en début d'apprentissage sont des tâches où l'enjeu et le niveau de difficulté ne se trouvent pas dans les processus non spécifiques : les phrases sont simples, le vocabulaire est connu à l'oral, les thèmes sont communs, les structures sont connues, etc. Ceci permet aux élèves de centrer surtout leur attention sur la mise en œuvre des processus spécifiques d'identification des mots écrits, bien qu'on leur demande de comprendre ce qu'ils lisent. Sélectionner une population de lecteurs débutants pour cette étude ne permettrait donc pas d'obtenir un fort contraste entre les lecteurs compensateurs et les autres, puisque même les élèves qui ne présentent pas de difficulté au niveau des processus spécifiques d'identification des mots écrits ont une compréhension écrite qui est limitée par leur capacité au niveau des processus spécifiques.

Dans les systèmes d'écriture alphabétiques comme le français, il est documenté que, vers 9 ans, les mécanismes de base de l'identification des mots écrits sont minimalement maîtrisés afin d'être automatisés chez les élèves qui n'ont pas de difficulté au niveau de ces processus (p. ex. Duncan et Seymour, 2000; Seymour, 2008). Pour ces élèves, parce que leurs processus spécifiques sont automatisés, la compréhension écrite résulte davantage de la contribution des processus non spécifiques de compréhension langagière (Catts et Kahmi, 2005; Perfetti, 1985). Par conséquent, les tâches scolaires proposées en classe à partir de la 4^{e} année du primaire tendent à être centrées sur la « compétence à lire », soit la compréhension écrite (Programme de formation de l'école québécoise (PFEQ), MEQ, 2006), et incitent les élèves à mobiliser surtout les processus non spécifiques. Ainsi, bien que l'on puisse penser que les lecteurs compensateurs mettent en œuvre une forme de compensation dès le début de l'enseignement formel de la lecture en 1^{e} année du primaire, il n'existe pas un grand décalage entre ce qui est attendu des élèves de cet âge en classe et ce que les lecteurs compensateurs arrivent à faire, ce qui rend difficile l'identification de ces derniers. Par contre, ce décalage devient plus grand après la 4^{e} année du primaire; il est attendu des élèves plus âgés que leurs ressources cognitives lors de la lecture puissent être dédiées davantage aux processus non spécifiques de compréhension langagière. Les lecteurs de cet âge qui ont un déficit sur le plan des processus spécifiques doivent encore allouer beaucoup de ressources

cognitives à ces processus alors qu'ils sont automatisés chez leurs pairs. Ces élèves ont donc, par conséquent, moins de ressources cognitives disponibles à dédier aux processus non spécifiques. Plus les élèves qui ont une telle difficulté avancent dans leur cheminement scolaire, plus l'écart avec les élèves qui n'ont pas de difficulté s'agrandit. Qui plus est, les élèves au-delà de la 4e année qui présentent encore des difficultés d'identification des mots se retrouvent dans une situation particulière, où on leur demande beaucoup en classe de mobiliser leurs processus non spécifiques lors de la lecture de textes, alors que leurs processus spécifiques sont beaucoup moins fonctionnels que ceux de leurs pairs. Cette insistance sur les processus non spécifiques en classe pourrait les encourager encore davantage à compenser leur déficit. Il apparait ainsi que sélectionner des lecteurs au-delà de la 4e année soit judicieux. Ainsi, on assure le contraste entre les lecteurs compensateurs, qui ont par définition un déficit au plan des processus spécifiques, et leurs pairs. Afin d'obtenir le plus grand contraste possible au primaire, il a été choisi de sélectionner la fin de la 5e année du primaire, pour faire en sorte que les élèves soient en 6e année au moment de l'étude 2.

3.1.2.1 Recrutement

Préalablement au recrutement, l'approbation éthique a été obtenue auprès du Comité d'éthique de la recherche pour les projets étudiants impliquant des êtres humains (CERPE) de l'Université du Québec à Montréal (UQAM), ainsi qu'auprès d'un centre de services scolaires, dont les procédures en matière d'éthique exigeaient également une approbation interne supplémentaire. C'est au moyen de différents formulaires décrivant le projet et en présentant des exemples de matériel que ces approbations ont été obtenues. Ce sont donc des lecteurs de la fin 5e année du primaire qui ont été sollicités pour participer à l'étude 1. Le recrutement des participants a débuté au printemps 2019, alors que les participants potentiels en étaient à terminer leur 5e année. Puisque peu de données sont disponibles quant à la prévalence des lecteurs compensateurs, il a été choisi de considérer des écoles qui ont un indice de milieu socio-économique (IMSÉ) moyen, puisque la distribution des différents profils de lecteurs pourrait être différente dans les milieux relevant d'indice socioéconomique très faible ou très élevé. Les milieux d'indices 1 et 2, puis 9 et 10 n'ont donc pas été retenus dans le recrutement. Ce faisant, bien que leur territoire soit plus proche du lieu de collecte de l'étude 2 à l'UQAM, les trois centres de services scolaires[17] de l'île de Montréal n'ont pas été retenues, la grande majorité de leurs écoles ne répondant pas à ce critère.

[17] Il est à noter qu'au printemps 2020, l'appellation des commissions scolaires a été modifiée. Elles sont désormais désignées par le terme *centre de services scolaires*.

Des personnes-contacts de trois centres de services scolaires ont agi à titre d'intermédiaire entre la chercheuse et le centre de services scolaires pour compléter toutes les démarches éthiques et administratives préalables au recrutement de participants dans les écoles. Ces personnes-contacts ont également été celles qui ont communiqué avec les écoles dont l'IMSÉ est moyen pour sonder leur intérêt à participer. Une fois l'intérêt à participer communiqué par la direction, l'orthopédagogue ou les enseignantes de 5e année, un échange téléphonique ou par courriel avec la chercheuse a eu lieu. Cet échange a notamment porté sur le déroulement, l'horaire, les dispositions éthiques appropriées, ainsi les questionnements du milieu scolaire. Dans les trois centres de services scolaires, ce sont les premières écoles à avoir communiqué leur intérêt qui ont été retenues, pour un maximum de 2 écoles par commission scolaire. Les écoles retenues ont ensuite reçu une lettre d'information ainsi qu'un formulaire de consentement à envoyer aux parents des élèves par les voies de communication habituelles de l'école. Seuls les élèves dont le parent a signé ledit formulaire de consentement éclairé ont participé à l'étude.

3.1.2.2 Caractéristiques de l'échantillon

Deux écoles par centre de services scolaires ont été visitées, impliquant 15 classes de 5e année au total. Ce sont ainsi 165 élèves de ces classes qui ont participé à l'étude 1. Il convient de noter ici qu'aucun critère d'exclusion n'a été établi (autre que l'indice socioéconomique de l'école), ceci afin d'observer l'échantillon d'élèves le plus représentatif des milieux moyens sur le plan socioéconomique. Le tableau et le graphique ci-dessous présentent la répartition des élèves par centre de services scolaires.

Tableau 3.1 Répartition des élèves par centres de services scolaires, par écoles et par classes

Centre de services scolaires (CSS)	Nombre d'écoles	Nombre de classes	Nombre de participants	% de l'échantillon
CSS 1	2 écoles	3 classes	22 élèves	13,33%
CSS 2	2 écoles	6 classes	84 élèves	50,9%
CSS 3	2 écoles	6 classes	59 élèves	35,75%
TOTAL	6 écoles	15 classes	165 élèves	100%

L'âge moyen des élèves de l'échantillon est de 134,5 mois, ce qui correspond à 11,2 ans, avec un écart-type de 5,5 mois (il est à noter que 4 participants n'ont pas répondu à cette question dans le formulaire). Le pourcentage de filles est de 54,6%, et de 45,4% pour les garçons (2 participants n'ont pas indiqué de

réponse à cette question). La presque totalité des élèves de l'échantillon ont le français comme langue maternelle (92,1%), tel que le montre le tableau suivant. Il importe de mentionner à nouveau ici qu'aucun critère relatif à la langue maternelle n'a été formulé dans le recrutement. Vu la situation géographique des centres de services scolaires ciblés, ainsi que le critère en lien avec l'IMSÉ, il est relativement attendu que peu d'élèves allophones se retrouvent dans l'échantillon. Le fait que l'échantillon soit relativement homogène au plan de la langue maternelle des élèves semble un avantage. De la même manière que les milieux socioéconomiques moyens ont été retenus pour augmenter l'homogénéité de l'échantillon, il est heureux de constater cette homogénéité au plan de la langue maternelle. Évidemment, cet état de fait est considéré dans l'interprétation des résultats. Quoi qu'il en soit, pour les quelques élèves dont le français n'est pas la langue maternelle, le français est leur 2^{e} langue. Le tableau ci-contre rapporte la langue maternelle des élèves de l'échantillon (3 participants n'ont pas indiqué de réponse à cette question).

Tableau 3.2 Répartition des élèves de l'échantillon selon leur langue maternelle

Langue maternelle	Fréquence	Pourcentage
Français	152	92,10
Espagnol	2	1,20
Punjabi	1	0,60
Turc	1	0,60
Anglais	1	0,60
Roumain	1	0,60
Arabe	3	1,80
Créole	1	0,60
non disponible	3	1,80
Total	165	100,00

3.1.3 Sélection des instruments de mesure

Les instruments de mesure nécessaires à l'étude 1 doivent concerner les variables énoncées plus haut : 1) précision de la compréhension écrite de texte narratif; 2) précision de l'identification des mots écrits; 3) fluence (précision X temps) de l'identification des mots écrits.

3.1.3.1 Instrument de mesure de la compréhension écrite

Pour la sélection de l'instrument de mesure de la compréhension écrite, il convient de sélectionner un instrument qui comporte des tâches de sélection d'images, de choix multiples ainsi que celles ayant recours à des questions ouvertes, puisque ce sont celles qui induisent la meilleure variance des résultats (Collins *et al.*, 2017). Pour la présente étude, elles offrent effectivement plus de possibilités d'anticipation du contenu pour les lecteurs compensateurs. Pour réussir ces tâches, il est nécessaire non seulement que le lecteur identifie les mots écrits, mais en plus il doit interpréter, modifier et refléter dans sa réponse le modèle mental qu'il a élaboré. De plus, afin d'inciter les élèves à recourir effectivement à leurs processus cognitifs les plus complexes, le texte doit être d'une certaine longueur. Tel que mentionné précédemment, le texte à lire doit être un texte narratif. En effet, les élèves sont exposés aux textes narratifs depuis le début de leur scolarité, voire avant l'entrée à l'école. La structure de ce type de texte est donc la mieux connue, le vocabulaire ainsi que les connaissances antérieures nécessaires à mobiliser sont également généralement plus accessibles. Par conséquent, ce type de texte semble le plus à même de permettre la compensation d'un déficit d'identification des mots écrits en comparaison aux textes informatifs, directifs ou argumentatifs, sur lesquels les lecteurs possèdent généralement moins de connaissances antérieures, etc.

L'instrument doit aussi mesurer uniquement les processus mis en œuvre lors de la lecture du texte; non pas les processus qui concernent la lecture des questions ou ceux impliqués lors de la production de réponses écrites. Ainsi, il est souhaitable que les questions soient lues aux élèves, et si certaines réponses devaient être constituées de mots écrits ou de phrases écrites, les élèves devraient les fournir à l'oral. Ces réponses devraient être transcrites par l'évaluateur. Par conséquent, la passation envisagée est une passation individuelle.

Tableau 3.3 Résumé des caractéristiques primordiales de l'instrument de mesure de la compréhension écrite

Caractéristiques du texte
Longueur > paragraphe
Structure narrative
Thème familier
Texte adapté à des élèves de 5e année
Caractéristiques de la tâche demandée
Sélection d'image, choix multiples ou réponses à des questions ouvertes
Modalité orale uniquement pour les questions/réponses
Passation individuelle
Tâche qui vise une variété de processus cognitifs de lecture

Les tâches de Bianco *et al.* (2018) semblent appropriées pour la présente recherche. Ces tâches ont été élaborées pour évaluer la compréhension en lecture d'élèves de la 3e à la 5e année du primaire (CE2 à CM2) en France, par la lecture de textes et la réponse à des questions de compréhension. Une quinzaine de textes se retrouvant dans du matériel pédagogique de ces niveaux scolaires ont été sélectionnés et des questions de compréhension visant différents processus cognitifs[18] ont été élaborées. Les 8 textes les plus discriminants ont ensuite été administrés à 500 élèves de chaque niveau scolaire en vue de faire un étalonnage. Ainsi, le résultat final de ce travail est une collection de 3 textes par niveau scolaire, comportant chacun une dizaine de questions de compréhension. Le texte correspondant au niveau scolaire des élèves de l'échantillon, soit le dernier texte pour la 5e année du primaire, est intitulé « Le Lion », et a une longueur adéquate. Cela fait en sorte que le temps de lecture est relativement court.

[18] Les auteurs ont utilisé les concepts de « compréhension de surface (littérale) » et de « compréhension approfondie (inférentielle) », ainsi que les concepts de « cohérence locale » et de « cohérence globale » pour équilibrer leurs questions. Ramenés dans les termes des modèles cognitifs théoriques présentés dans cette thèse, il s'agit ainsi de questions qui visent tant les microprocessus (cohérence locale) que les macroprocessus (cohérence globale), et de questions qui visent tant la compréhension littérale que les processus d'inférence.

Considérant que la passation est individuelle, le temps de passation de l'épreuve doit demeurer réaliste. Aussi, la structure du texte correspond à un récit « standard » et les questions formulées sur le texte lu font appel à un inventaire pertinent de processus cognitifs de lecture décrits en détails plus bas et sont des questions ouvertes ou à choix multiples comme le proposent Collins *et al.* (2017). Le texte raconte l'histoire d'un aventurier qui se trouve en Afrique alors qu'il chasse pour trouver de la nourriture, et qui fait face au danger représenté par un lion qui convoite la même proie que lui. Les principaux thèmes du texte sont donc l'aventure, les animaux et la survie en milieu naturel; les deux premiers thèmes semblent, à notre connaissance, très communs pour des élèves du groupe d'âge ciblé, vu la récurrence de ce thème dans le matériel scolaire. Ceci permet ainsi de considérer que les connaissances des élèves à propos de ces thèmes puissent leur fournir ici aussi l'occasion de compenser (vocabulaire connu, scénario/schèmes nombreux en lien avec le thème, etc.) Le texte comporte 329 mots regroupés en trois paragraphes, et respecte la structure narrative avec, dans l'ordre, une situation initiale, un élément déclencheur, une péripétie et un dénouement. La lecture de ce court texte est suivie de 8 questions ouvertes ainsi que 2 questions à choix de réponse. Ces 10 questions ciblent principalement les processus non spécifiques de lecture et leur réponse implique soit une recherche dans le texte parce que la réponse est littérale (5 questions), soit une inférence de la part du lecteur (5 questions). Certaines questions nécessitent la mobilisation de microprocessus (7 questions), d'autres nécessitent la mobilisation de macroprocessus (2 questions) et une autre question se centre sur l'élaboration du modèle de situation (1 question)[19].

Cette épreuve a été réalisée par Bianco et ses collaborateurs en France pour évaluer des élèves français. Une analyse du texte et des questions ne souligne pas de différence culturelle pour la population québécoise dans le contenu évoqué. La forme linguistique ne pose pas non plus de problème majeur. Par contre, dans une des questions, la formulation a été adaptée au français standard québécois. La question comporte une utilisation de la préposition « du » si peu fréquente en français québécois qu'elle porte à

[19] Il convient de noter qu'une question classée dans un processus de haut niveau (p. ex. élaboration du modèle de situation) inclut nécessairement tous les processus qui y sont inférieurs. Une bonne réponse à cette question implique donc la mise en œuvre efficace du processus correspondant au classement de la question, ainsi que de tous les processus cognitifs inférieurs. À l'inverse, une mauvaise réponse à une question classée dans un processus pourrait être attribuable à ce processus ou à n'importe lequel de plus bas niveau. Le classement des questions fournit donc une indication sur le nombre de processus à mobiliser pour répondre adéquatement à ladite question.

confusion. Ainsi, dans la question « À la fin de l'histoire, que fait le lion du caama? », la préposition « du » est remplacé par « …, que fait le lion avec le caama»[20].

L'instrument de Bianco *et al.* (2018) comporte un protocole de passation collective afin que plusieurs élèves d'une même classe puissent être évalués simultanément. Dans le cas de l'étude 1, tel que mentionné précédemment, la passation doit être individuelle pour que l'élève puisse fournir ses réponses en modalité orale. Ce faisant, le protocole original a été adapté à ce contexte. Ces modifications concernent principalement les consignes formulées à l'élève, la possibilité de lire le texte silencieusement ou à voix haute – au choix de l'élève –, puis le fait que c'est l'évaluateur qui lit oralement les questions qui sont présentées simultanément à l'élève sous forme écrite, puis qui transcrit les réponses fournies oralement par l'élève. À l'exception de ces éléments, le protocole a été conservé intact.

3.1.3.2 Instrument de mesure de la précision et de la fluence de l'identification des mots écrits

Concernant l'instrument de mesure des processus spécifiques d'identification des mots écrits, il importe de mentionner d'abord que, puisque ces processus sont relativement automatisés chez des élèves à la fin du primaire qui n'ont pas de difficulté, il n'est pas rare de constater un effet plafond dans les outils qui considèrent les scores de précision uniquement. Par exemple, les normes françaises rapportées pour l'épreuve de lecture de mots et pseudomots dans la BALE (Batterie analytique du langage écrit, Zorman *et al.*, 2002), batterie dont il a été question précédemment, témoignent de cet effet plafond : pour les mots réguliers de haute fréquence, le score associé au 20^{e} rang centile (jusqu'au dernier rang centile) est de 20/20 pour des élèves de CM2 (5^{e} année du primaire). Peu importe l'épreuve choisie, puisque l'habileté à identifier des mots écrits est déjà automatisée pour les élèves de 10-11 ans (Seymour et Duncan, 2003), il y a de fortes probabilités que cet effet plafond existe sur certaines catégories d'items. Il est ainsi sans doute impossible de l'éviter complètement. Quoi qu'il en soit, le recours à un grand nombre d'items, de complexité suffisamment élevée, pourrait contribuer à diminuer le manque de sensibilité de l'instrument de mesure attribuable à cet effet plafond.

[20] Sans entrer dans le détail ici, il convient de mentionner qu'en français québécois, l'usage de « du » apparaît surtout dans les groupes nominaux (p. ex. « le chat du voisin »). Ce faisant, pour éviter une confusion, il a été choisi de modifier la formulation de la question pour empêcher d'interpréter les mots « du caama » comme faisant partie d'un groupe nominal.

Dans le cas présent, les variables concernant les processus d'identification des mots écrits relèvent tant de la précision que de la fluence. Ainsi, tant la qualité de la réponse que le temps consacré à la lecture des items présentés doivent être pris en compte. Cette considération du temps pourrait également permettre d'éviter l'effet plafond, si la variation entre les élèves s'avérait être assez grande. En effet, même si les élèves sans difficulté de l'échantillon arrivent à lire un nombre élevé de mots correctement, il est fort possible que l'on retrouve une variation dans le temps nécessaire pour compléter la tâche. Aussi, les variables concernant les processus d'identification des mots écrits relèvent tant des mots que des pseudomots. L'instrument sélectionné doit alors couvrir ces deux types d'items, et doit les apparier entre eux de telle sorte que l'on puisse comparer les résultats obtenus sur ces deux catégories d'items. Pour ce faire, les caractéristiques linguistiques des items doivent avoir été minutieusement contrôlées afin que seule la lexicalité distingue les mots des pseudomots (ou en tout cas le plus possible). Finalement, il est souhaité que l'instrument présente des items qui sollicitent tous les mécanismes d'identification des mots écrits. Tel que mentionné dans le cadre théorique, le mécanisme d'identification de mots mobilisé dépend des caractéristiques linguistiques des items à identifier (fréquence, régularité, etc.) L'assurance de la couverture du plus de mécanismes possibles réside donc dans la considération sérieuse des caractéristiques linguistiques. Les caractéristiques essentielles pour les instruments de mesure visant la précision et la fluence de l'identification des mots écrits sont décrites ci-dessous :

Tableau 3.4 Résumé des caractéristiques primordial à l'instrument de mesure de l'identification des mots écrits

Caractéristiques des instruments de mesure de la précision et de la fluence de l'identification des mots écrits
Grand nombre d'items
Comparaison possible de la performance sur les mots et les pseudomots (items appariés entre eux)
Contrôle minutieux du plus de caractéristiques linguistiques possible
Mobilisation de tous les mécanismes d'identification des mots écrits

Le sous-test MIM de la Batterie d'évaluation du langage écrit et ses troubles (BELEC; Mousty *et al.*, 1994) semble finalement un choix judicieux. D'abord, ce sous-test contient les deux principales catégories d'items souhaités (mots et pseudomots), dont certains items sont d'un niveau de complexité élevé. Elle comporte également un grand nombre d'items en comparaison à d'autres instruments, soit 72 items, regroupés en 12 blocs de 6 items. Cette tâche permet d'obtenir un score de réussite qui peut, si cela s'avère pertinent, être décliné en plusieurs scores selon les types d'items puisque tous les blocs d'items sont appariés entre eux. Qui plus est, le protocole prévoit la considération du temps pris pour lire les items

par blocs de 6 items. Il est possible, avec ce même outil, de calculer la fluence de la lecture des items en considérant un ratio du nombre d'items correctement lus sur le temps pris pour lire les 6 items d'un bloc. Ce sous-test permet donc d'obtenir deux scores, soit un score de précision et un score de fluence. La validité de contenu est excellente, mais les caractéristiques psychométriques ne sont pas toutes documentées. En effet, seules les données d'un petit échantillon d'élèves sont disponibles. Dans ce petit échantillon d'élèves de la 5e année du primaire, on dénote un effet plafond. La moyenne du pourcentage de réussite pour les mots fréquents est de 98,35%, de 95% pour les mots rares (qu'ils soient courts ou longs) et de 90,9% pour les pseudomots courts et 80% pour les pseudomots longs (qui représentent les items les plus complexes de la liste). Cet outil est tout de même retenu, en gardant en tête le problème potentiel de l'effet plafond.

Il existe également une autre façon de mesurer la fluence de l'identification des mots écrits. Il s'agit d'une tâche d'identification d'une liste d'items devant être lus oralement en un temps maximum. En effet, faire lire une liste d'items à l'intérieur d'un temps limite permet de calculer un ratio d'items lus correctement dans ce temps donné. Dans le cas où le temps pris pour lire la liste est moindre que le temps alloué, il est possible d'extrapoler le nombre d'items qui auraient été correctement lus si l'élève avait pu continuer à lire des items pendant le temps maximum alloué. Ceci permet d'obtenir un score du nombre d'items correctement lus en une minute pour tous les élèves. De la même manière que la mesure de fluence précédente, l'effet plafond a une chance d'être évité encore ici puisque de la variation dans le temps pris pour lire les items pourrait être présente malgré un manque de variation sur le score de précision. Ainsi, la tâche de lecture orale d'items en une minute maximum, aussi appelée *Test 1 minute,* est également retenue ici. C'est précisément le « Test 1 minute de lecture à voix haute de mots familiers et de mots inventés » de Sprenger-Charolles et Leloup (2013) et de Leloup (2011) qui est utilisé.

La lecture de mots familiers comprend 50 mots à lire oralement le plus rapidement possible, pour un maximum de temps d'une minute. Une communication personnelle avec les auteurs a permis de dégager les caractéristiques suivantes. La régularité de ces mots a été considérée (30 réguliers – « pétrole » – et 20 irréguliers – « seconde »), ainsi que leur longueur (25 courts – « farce » – et 25 longs – « domicile »). Les mots courts sont formés de mots qui comportent 4 et 5 lettres, alors que les mots longs sont constitués de 6 à 9 lettres. La fréquence écrite de ces mots est moyenne[21]. La présence ou non de graphies

[21] La fréquence considérée est celle exprimée en U (fréquence d'usage pour 1 million de mots de la base de données Manulex).

contextuelles (graphies dont la prononciation dépend du contexte) et la complexité syllabique et graphémique ont été minutieusement contrôlées. La deuxième partie de ce test comprend 40 pseudomots à lire le plus rapidement possible, pour un maximum de temps d'une minute. Les pseudomots sont appariés aux mots de la première partie. Ainsi, 20 pseudomots ont été appariés à certains mots réguliers (p. ex. « vagne » - « vigne »), les 20 autres aux mots irréguliers (p. ex. « aximon » - « examen »). Vingt pseudomots sont courts, alors que les 20 autres sont longs (« poute » - « ogréiple »). Puisque les items sont appariés, la présence ou non de graphies contextuelles ainsi que la complexité syllabique et graphémique sont également contrôlées.

Ce Test 1 minute permet de comparer la performance des élèves sur les mots et sur les pseudomots. Une telle comparaison est nécessaire ici puisque, tel qu'indiqué plus haut dans ce texte, la performance sur les mots (précision et fluence) représente une situation où la compensation par la sémantique est possible, alors que la performance sur les pseudomots empêche la compensation par la sémantique.

Ce test permet de mesurer la fluence des processus d'identification des mots écrits, puisque le score obtenu inclut tant la précision que le temps pris pour identifier les items. Si le temps pris pour lire la liste d'items est moindre qu'une minute, le temps réellement utilisé est noté. Pour obtenir le score final, on divise le nombre d'items réussis par le temps pris (en seconde), et on le ramène ensuite sur 60 secondes. Ceci nous permet d'indiquer combien d'items auraient été lus en une minute s'il y avait eu plus d'items sur la liste. Par conséquent, la performance à ces mesures est représentée par le nombre d'items lus en une minute. La manière de faire ce calcul dans les deux cas de figure sont présentés ci-dessous : soit l'élève a besoin de l'entièreté de la minute pour lire les items de la page, soit l'élève lit la page entière en moins d'une minute.

Tableau 3.5 Procédures de calcul du nombre de mots correctement lus en une minute selon les deux possibilités de considération du temps

Deux cas d'élèves	Nbr. items / temps pris (en sec.)	Ratio	Ratio X 60 secondes	Nombre de mots lus en 1 minute
Élève que l'on a arrêté après une minute (la liste n'est pas lue entièrement)	25 items / 60 secondes	-----	-----	25 mots
Élève qui a pris moins d'une minute	25 items / 30 secondes	0,833	0,833 X 60 secondes	50 mots

3.1.4 Déroulement

Cette collecte de données s'est déroulée d'avril à juin 2019. Les enseignantes titulaires des classes participantes ont été responsables de recueillir les formulaires de consentement signés et de prévoir la venue de l'équipe de recherche dans leur milieu (communiquer à la chercheure le nombre d'élèves à tester, prévoir les locaux et l'horaire, etc.)

Le jour désigné pour la collecte, les trois instruments de mesure ont été administrés en modalité individuelle, en une seule séance. Chaque élève dont les parents avaient fourni leur consentement écrit a été évalué individuellement, dans un local isolé de son école d'appartenance. La passation a été réalisée par la chercheuse ou l'une des trois assistantes de recherche impliquées (un formulaire d'entente à la confidentialité a été signé). À chaque séance de passation individuelle, l'évaluatrice mentionne explicitement à l'élève la portée des résultats (données anonymes non conservées à l'école, ne compte pas au bulletin, etc.) et s'assure que l'élève est toujours volontaire avant de débuter.

Au cours de la séance, l'élève complète d'abord l'évaluation de la compréhension écrite « Le lion ». L'évaluatrice fournit les consignes à l'élève (suivant le protocole) et l'élève lit le texte selon la modalité de son choix (haute voix ou silencieusement; l'évaluatrice consigne la modalité choisie). Une fois la lecture par l'élève terminée, l'évaluatrice lit à haute voix la première question du questionnaire qui est disposé devant l'élève, puis elle transcrit la réponse fournie par l'élève. Elle procède ainsi pour les autres questions. Une fois toutes les questions répondues, l'évaluatrice fournit les consignes et le matériel pour l'évaluation

de lecture de mots et pseudomots et consigne judicieusement les réponses fournies oralement par l'élève; un crochet correspond aux items réussis et les erreurs commises sont transcrites en alphabet phonétique.

3.2 Résultats de l'étude 1

Afin d'identifier les lecteurs compensateurs dans l'échantillon d'élèves de 5e année scolarisés en français de milieu socioéconomique moyen, les instruments de mesure « Le lion » (Bianco *et al.*, 2018), MIM (BELEC; Mousty *et al.*, 1994) et Test 1 minute (Sprenger-Charolles et Leloup, 2013) ont été administrés à l'ensemble des élèves de l'échantillon (n=165). Les résultats sont présentés ici. Les données descriptives obtenues à l'aide de chaque instrument sont d'abord présentées, puis des corrélations entre les différentes variables mesurées par ces instruments sont rapportées.

3.2.1 Données descriptives des trois instruments

Les résultats des participants à l'épreuve de compréhension écrite « Le lion » sont présentés à la figure suivante. Sur un score maximum possible de 10, la moyenne du groupe est de 5,78 (ET = 1,80).

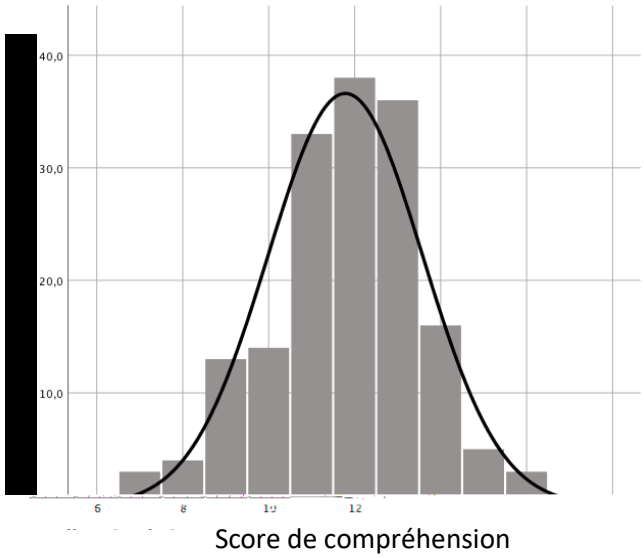


Figure 3.2 Distribution des scores obtenus à l'épreuve de compréhension écrite « Le Lion » (Bianco *et al.*, 2018)
Note. La ligne noire représente une distribution normale.

Tableau 3.7 Données descriptives de précision de l'identification des mots écrits selon MIM (BELEC)

Épreuve	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
MIM-précision (mots)	165	40	70	61,17	5,71	-1,04 (0,19)	1,44 (0,38)

La figure 3.4 illustre bien l'effet plafond obtenu pour cette l'épreuve. Ainsi, 71,5% de l'échantillon a un score de 59 et plus. Le score de 59/72 correspond à un taux de réussite de 82% des items de l'épreuve. Le calcul d'asymétrie de la courbe (*skewness*) et de coefficient d'aplatissement (*kurtosis*) soulignent également cette tendance des scores à être élevés, avec des valeurs de -1,04 et de 1,43, respectivement. Tel que mentionné précédemment, cet effet plafond est relativement attendu dans le cas d'une mesure en identification des mots écrits sur une population de ce niveau scolaire. La figure met également en lumière le fait que certains élèves se démarquent de leurs pairs à cause de leur score très faible (4% de l'échantillon a un score plus faible que 50/72).

Puisqu'il est prévu de comparer la performance sur les mots à celle sur les pseudomots, et puisque l'instrument est construit de manière à contrebalancer les items mots (48 mots) et pseudomots (24 pseudomots), il convient de rapporter les résultats par catégorie d'items; les figures ci-dessous présentent les données descriptives par catégorie. Pour les mots, la moyenne est de 44,15 réponses correctes (ET = 2,961; maximum possible 48) et pour les pseudomots, elle est de 17,01 (ET = 3,407; maximum possible 24).

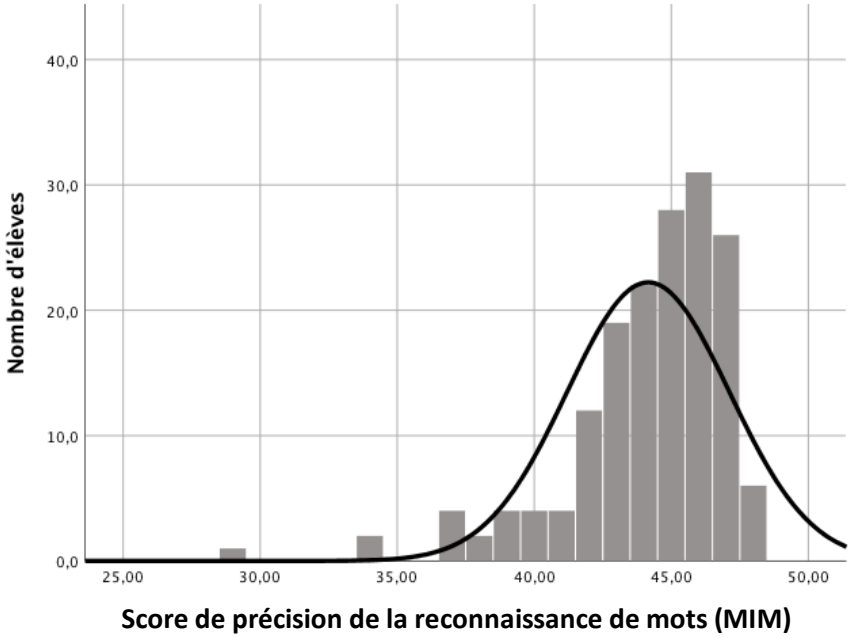


Figure 3.4 Distribution des scores de précision obtenus sur les mots du sous-test MIM (BELEC)

Note. La ligne noire représente une distribution normale.

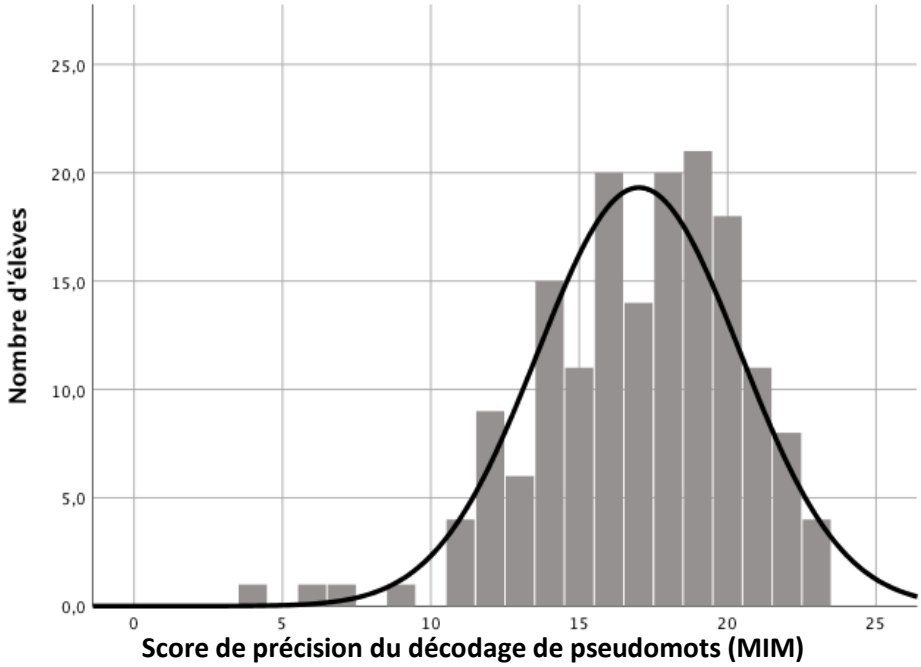


Figure 3.5 Distribution des scores de précision sur les pseudomots au sous-test MIM (BELEC)

Note. La ligne noire représente une distribution normale.

Tableau 3.8 Données descriptives de précision du sous-test MIM (BELEC) pour les mots et les pseudomots

Mesure	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
Précision sur les mots (MIM)	165	29	48	44,16	2,96	-1,78 (0,19)	4,81 (0,38)
Précision sur les pseudomots (MIM)	165	4	23	17,01	3,41	-0,74 (0,19)	0,98 (0,38)

Les scores obtenus sur les mots tendent davantage vers l'effet plafond que ceux des pseudomots. En effet, la moyenne des scores sur les mots correspond à un taux de réussite de 92% des items. Une moyenne si élevée ne permet pas beaucoup de variation, ce qui fait que 76% de l'échantillon se retrouve dans les scores 43/48 à 47/48. Le calcul d'asymétrie de la courbe (*skewness*) et de coefficient d'aplatissement (*kurtosis*) sont de -1,78 et de 4,81, respectivement. En ce qui concerne les pseudomots, la moyenne des scores correspond à un taux de réussite de 70,87% des items de la liste. Les scores y sont plus distribués que dans l'identification de mots, avec une valeur d'asymétrie de la courbe (*skewness*) de -0,74. Même si l'effet plafond n'est pas présent dans les scores de décodage de pseudomots, il n'en demeure pas moins que ces scores ne sont pas répartis de manière idéale jusqu'aux extrémités de la distribution tel qu'en témoigne le coefficient d'aplatissement (*kurtosis*) qui est de 0,98.

Le sous-test MIM permet également de calculer un ratio qui inclut le nombre d'items correctement lus et le temps pris pour ce faire. Elle permet ainsi de mesurer la fluence de l'identification de mots. Ce sont ainsi les mêmes scores qui ont été présentés dans les trois figures 3.4, 3.5 et 3.6 qui sont repris, en considérant le temps pris pour identifier ces mots, ce qui est désigné par le terme « ratio ». Le calcul est donc celui-ci :

$$ratio = \frac{nombre\ de\ mots\ bien\ lus}{temps\ pris\ pour\ lire\ tous\ les\ items}$$

Ainsi, la distribution des ratios obtenus est présentée à la figure suivante. La moyenne de la fluence sur les mots et les pseudomots considéré ensemble est de 0,64 mot lu par seconde (ET = 0,15).

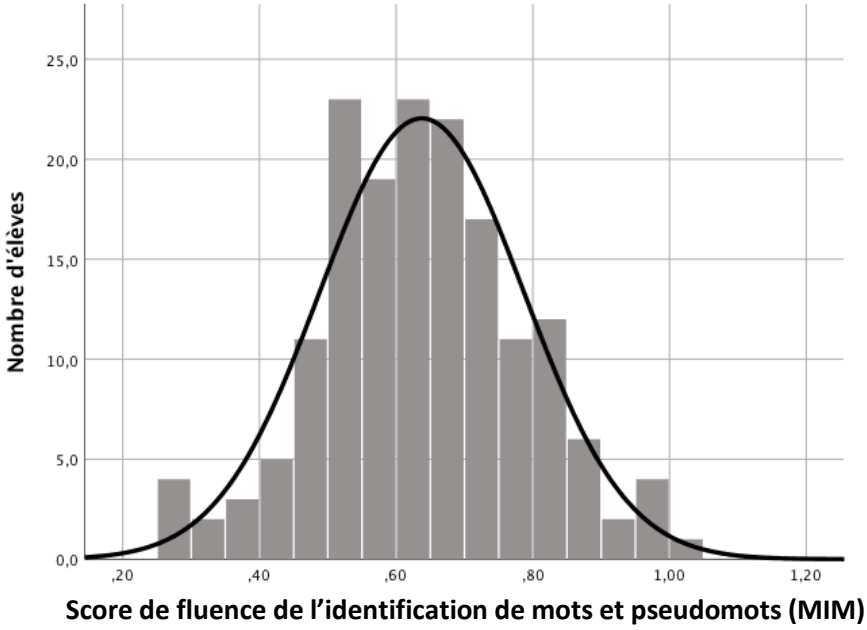


Figure 3.6 Distribution des scores de la fluence de l'identification de mots et pseudomots du sous-test MIM (BELEC)

Note. La ligne noire représente une distribution normale. La fluence de la MIM correspond au ratio du nombre de mots bien lus divisé par le temps pris pour le faire, en secondes.

Tableau 3.9 Données descriptives de fluence de l'identification des mots et des pseudomots au sous-test MIM (BELEC)

Épreuve	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
Fluence sur les mots et pseudomots (MIM)	165	0,25	1,01	0,64	0,15	-0,01 (0,19)	0,06 (0,38)

Le fait de considérer un ratio incluant le temps, plutôt que les scores uniquement, permet ici effectivement d'éliminer l'effet plafond. Les valeurs d'asymétrie de la courbe (*skewness*) et de coefficient d'aplatissement (*kurtosis*) sont excellentes, -0,01 et de 0,06, respectivement, ce qui indique que la distribution des ratios s'apparente à une distribution normale.

Les données sont également considérées en séparant les mots et les pseudomots. Les figures suivantes présentent la distribution de ces scores.

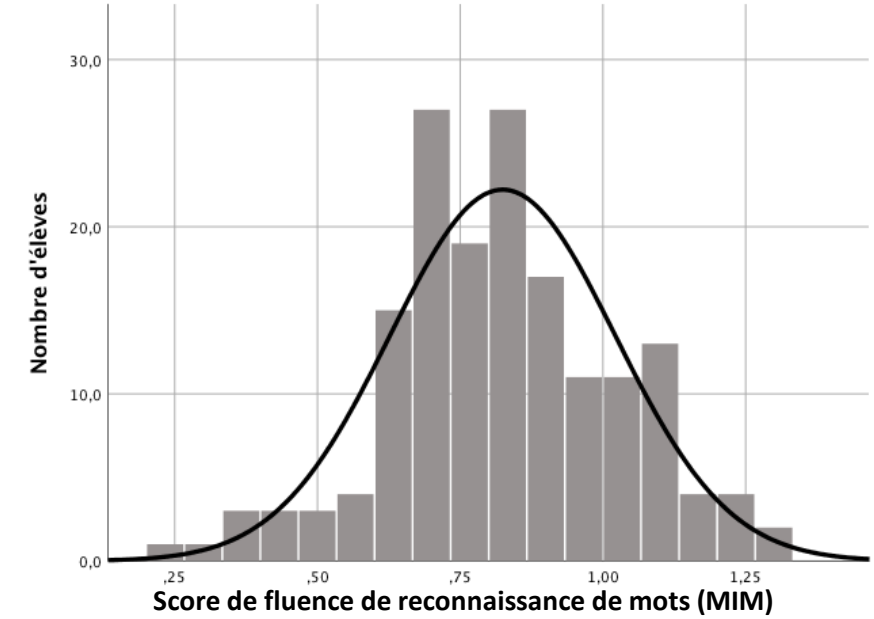


Figure 3.7 Distribution des scores de fluence de reconnaissance de mots, au sous-test MIM (BELEC)

Note. La ligne noire représente une distribution normale. La fluence de la MIM correspond au ratio du nombre de mots bien lus divisé par le temps pris pour le faire, en secondes.

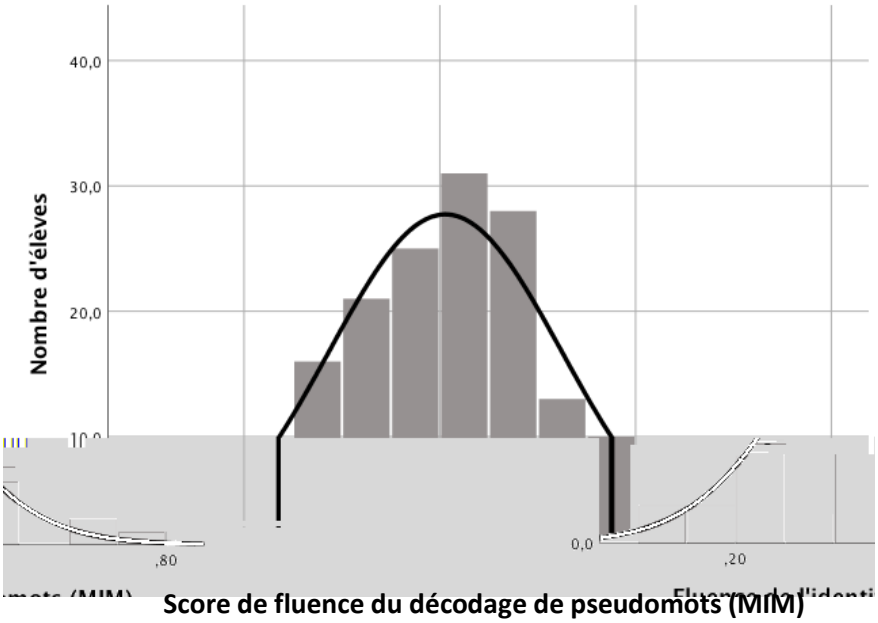


Figure 3.8 Distribution des scores de fluence du décodage de pseudomots au sous-test MIM (BELEC)

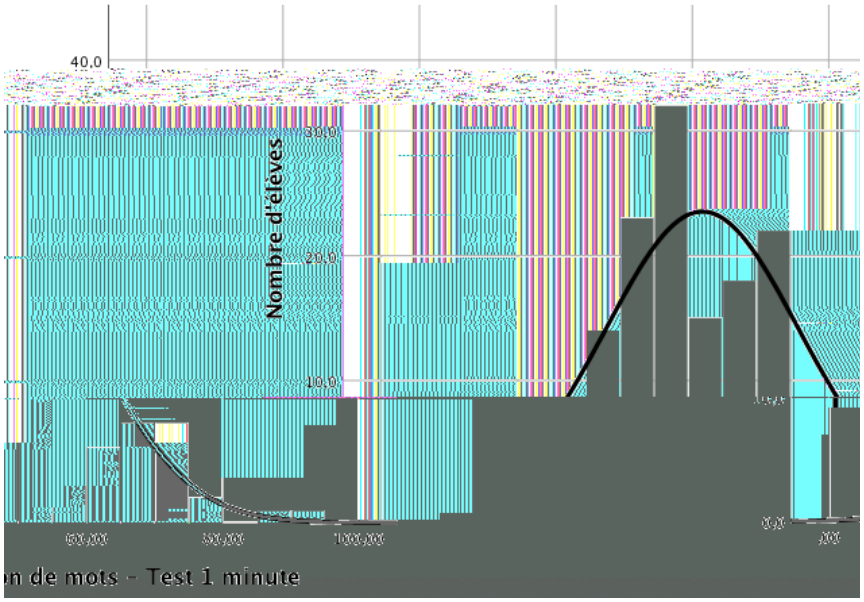
Note. La ligne noire représente une distribution normale. La fluence de la MIM correspond au ratio du nombre de mots bien lus divisé par le temps pris pour le faire, en secondes.

Tableau 3.10 Données descriptives de la fluence d'identification de mots et de pseudomots

Épreuve	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
Fluence sur les mots (MIM)	165	0,27	1,28	0,82	0,20	-0,01 (0,19)	0,19 (0,38)
Fluence sur les pseudomots (MIM)	165	0,11	0,78	0,41	0,12	0,14 (0,19)	0,50 (0,38)

La moyenne pour les mots est de 0,82 mots lus correctement par seconde (ET = 0,20), et de 0,40 pseudomots lus correctement par seconde (ET = 0,12). Dans les mots, la distribution des données s'apparente, en général, à une distribution normale. La moyenne est plus faible en décodage de pseudomots, ce qui est attendu selon la littérature scientifique qui rapporte ce phénomène, documenté maintes fois tant chez l'adulte (voir le résumé de Ferrand, 2007) que chez l'enfant (voir le résumé de Sprenger-Charolles, 2010). Les scores sont sensiblement centrés autour de la moyenne, s'apparentant ainsi à la courbe normale.

La dernière épreuve, constituée de deux sous-tests, permet également de mesurer la fluence. Tout comme l'instrument précédent, il est possible de considérer la performance sur l'ensemble des types d'items (mots et pseudomots) puisque les listes ont été appariées de cette manière. Il est également possible de considérer les résultats pour les mots et pseudomots séparément. Il convient de rappeler que l'instrument est conçu de sorte qu'une liste d'items doit être lue en 1 minute maximum. Si le temps pris pour lire tous les items de la liste est moindre, ce temps est noté (en secondes) et un ratio est calculé et extrapolé pour indiquer le nombre d'items qui auraient été réussis en une minute. Les résultats des participants sont présentés dans la figure 3.10 ci-dessous. la moyenne de mots et de pseudomots lus correctement en 1 minute pour les deux sous-tests considérés simultanément est de 47,05 (ET = 13,98).



Score de fluence de l'identification de mots et de pseudomots (Test 1 minute)

Figure 3.9 Distribution des scores de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Note. La ligne noire représente une distribution normale. La fluence correspond au nombre de pseudomots bien lus en 1 minute.

Tableau 3.11 Données descriptives de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Épreuve	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
Fluence sur les mots et les pseudommots (Test 1 minute)	165	14,00	93,44	47,05	13,98	0,32 (0,19)	0,34 (0,38)

Tout comme avec le sous-test MIM présenté précédemment, le fait de considérer un ratio incluant le temps permet encore ici d'éviter l'effet plafond. Les valeurs d'asymétrie de la courbe (*skewness*) et de coefficient d'aplatissement (*kurtosis*) sont relativement bonnes – 0,32 et de 0,34, respectivement, bien que peu de scores semblent se retrouver dans les extrémités.

En considérant les mots et les pseudomots séparément, tel que le montrent respectivement les figures 3.11 et 3.12 ci-dessous, la moyenne de mots lus correctement en 1 minute est de 60,84 pour les mots (ET = 19,28); elle est de 33,25 pour les pseudomots (ET = 10,20). Encore une fois, tel qu'attendu, la performance moyenne sur les pseudomots est moins élevée que celle sur les mots.

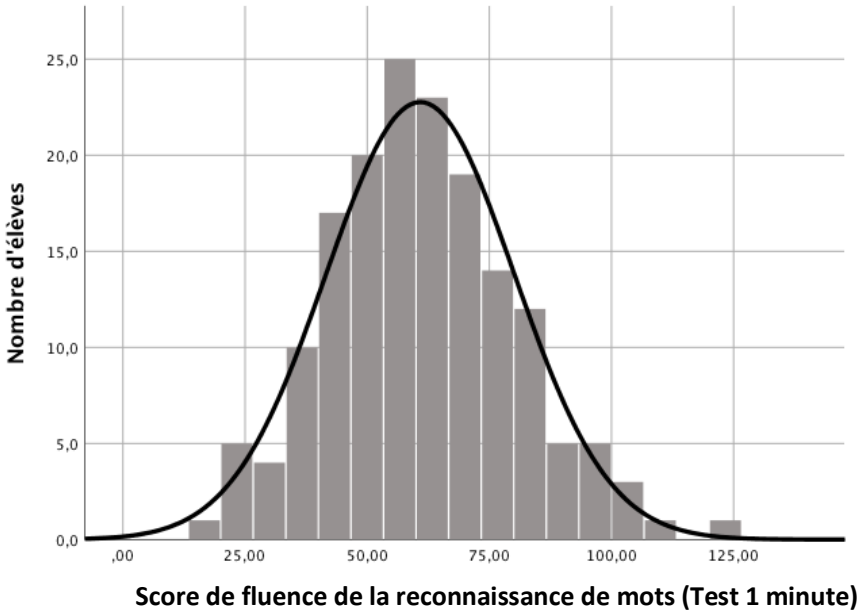


Figure 3.10 Distribution des scores de fluence de la reconnaissance de mots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Note. La ligne noire représente une distribution normale. La fluence correspond ici au nombre de mots bien lus en 1 minute.

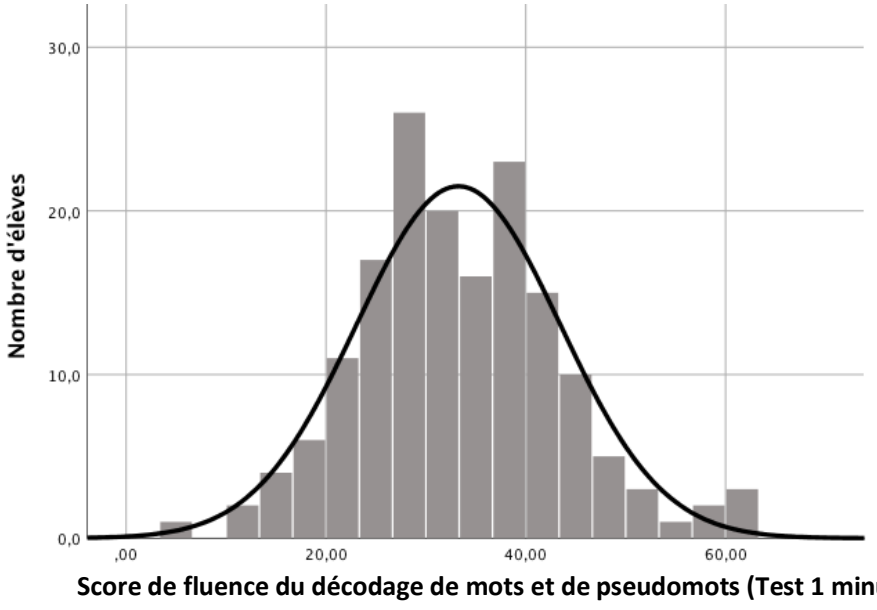


Figure 3.11 Distribution des scores de fluence du décodage de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Note. La ligne noire représente une distribution normale. La fluence correspond ici au nombre de mots et pseudomots bien lus en 1 minute.

Tableau 3.12 Données descriptives de fluence de l'identification de mots et de pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Épreuve	N	Minimum	Maximum	Moyenne	Écart-type	Skewness (erreur)	Kurtosis (erreur)
Fluence sur les mots (Test 1 minute)	165	14	125,22	60,84	19,28	0,30 (0,19)	0,26 (0,38)
Fluence sur les pseudomots (Test 1 minute)	165	4	62	33,25	10,20	0,32 (0,19)	0,42 (0,38)

Les scores au Test 1 minute qui concernent les mots sont davantage étalés, allant de 14 mots correctement lus à 125,22 mots bien lus. Les scores sur les pseudomots sont plus rapprochés, tel que l'indiquent le minimum et le maximum de la distribution, soit 4 pseudomots bien lus à 62 pseudomots bien lus. La variété des ratios obtenus par les élèves dans les mots et pseudomots permet aux scores d'être encore plus distribués de manière similaire à la courbe normale que c'était le cas dans le sous-test MIM.

3.2.2 Corrélations entre les variables

Les analyses descriptives présentées précédemment permettent d'observer la distribution des résultats pour chacune des variables. La section qui suit présente les corrélations entre ces variables. Ces corrélations permettent de mettre en lumière certains éléments qui seront pris en compte ultérieurement dans le choix des mesures pour identifier avec le plus de justesse possible les lecteurs compensateurs.

3.2.2.1 Corrélations des instruments qui mesurent l'identification des mots écrits

Deux épreuves mesurent la fluence de l'identification de mots; le sous-test MIM (mots bien lus/temps), et les sous-tests du Test « 1 minute » (nombre de mots ou de pseudomots bien lus en 1 minute). À l'instar des analyses descriptives, on peut ici aussi considérer les mots et les pseudomots ensemble, ou distinctement, ce qui résulte en 3 variables par instrument : fluence de l'identification de mots (mots seulement), fluence du décodage de pseudomots et fluence des 2 types d'items (mots et pseudomots). Les corrélations de ces 6 variables sont présentées dans le tableau 3.13, ci-dessous.

Tableau 3.13 Corrélations des différentes mesures de fluence en identification de mots et de pseudomots

	Moyenne (écart-type)	1.	2.	3.	4.	5.
1. Fluence – mots et pseudomots (MIM)	0,64 (0,15)					
2. Fluence – mots (MIM)	0,82 (0,20)					
3. Fluence – pseudomots (MIM)	0,41 (0,12)		,680**			
4. Fluence – mots et pseudomots (Test 1 min.)	0,78 (0,23)	,863**	,835**	,712**		
5. Fluence – mots (Test 1 min.)	60,84 (19,28)	,821**	,823**	,630**		
6. Fluence – pseudomots (Test 1 min.)	33,25 (10,20)	,815**	,735**	,761**		,779**

Note. ** indique que p < 0,01. Les corrélations qui impliquent des mesures redondantes du même concept ont été retirées du tableau.

Les deux instruments mesurant la fluence sont très fortement corrélés. Le score du MIM sur tous les items est fortement corrélé aux scores obtenus aux sous-tests « 1 minute » sur les mots et les pseudomots (r = 0,863, p < 0,01). Lorsque l'on considère les mots et les pseudomots séparément, les corrélations entre ces deux instruments sont également fortes pour les pseudomots (mots : r = 0,823, p < 0,01; pseudomots : r = 0,761, p < 0,01). Considérant ces corrélations, on constate que les deux mesures de fluence (mots et pseudomots ensemble), soit le MIM et le Test 1 minute, partagent 75% de la variance (r^2 = 0,75), ce qui est considérable. Ainsi, les deux instruments semblent bel et bien mesurer un construit similaire, soit la fluence de l'identification des mots écrits, et s'avèrent tous les deux adéquatement construits.

Il convient également d'analyser les corrélations entre les mesures de précision obtenues dans MIM et les mesures de fluence obtenues dans MIM et dans les sous-tests « 1 minute » (tableaux 3.14 et 3.15).

Tableau 3.14 Corrélations de la précision et de la fluence au sous-test MIM (BELEC)

	Moyenne (écart-type)	1.	2.	3.	4.	5.
1. Précision – mots et pseudomots (MIM)	61,17 (5,71)					
2. Précision – mots (MIM)	44,16 (2,96)					
3. Précision – pseudomots (MIM)	17,01 (3,41)		,608**			
4 Fluence – mots et pseudomots (MIM)	0,64 (0,15)	,498**	,523**	,381**		
5. Fluence – mots (MIM)	0,82 (0,20)	,401**	,508**	,230**		
6 Fluence – pseudomots (MIM)	0,41 (0,12)	,642**	,481**	,659**		,680**

Note. ** indique que p < 0,01. Les corrélations qui impliquent des mesures redondantes du même concept ont été retirées du tableau.

Tableau 3.15 Corrélations entre les mesures de précision du sous-test MIM et de fluence du Test 1 minute (Sprenger-Charolles et Leloup, 2013)

	Moyenne (écart-type)	1.	2.	3.	4.	5.
1. Précision – mots et pseudomots (MIM)	61,17 (5,71)					
2. Précision – mots (MIM)	44,16 (2,96)					
3. Précision – pseudomots (MIM)	17,01 (3,41)		,608**			
4 Fluence – mots et pseudomots (Test 1 min.)	0,78 (0,23)	,396**	,463**	,262**		
5. Fluence – mots (Test 1 minute)	60,84 (19,28)	,314**	,405**	,174*		
6 Fluence – pseudomots (Test 1 minute)	33,25 (10,20)	,493**	,502**	,389**		,779**

Note. ** indique que p < 0,01. * indique que p < 0,05. Les corrélations qui impliquent des mesures redondantes du même concept ont été retirées du tableau.

La mesure de précision de l'identification de mots du sous-test MIM est moyennement corrélée à la mesure de fluence sur les mots de ce même outil (r = 0,508, p < 0,01). Il en est de même pour la précision et la fluence dans le décodage de pseudomots (r = 0,659, p < 0,01). Considérant l'effet plafond rapporté précédemment en ce qui concerne le sous-test MIM sur les mots, la corrélation moins forte sur ce type d'items pourrait être attribuable uniquement au fait que l'effet plafond restreint la variation, empêchant ainsi de faire ressortir une corrélation. La corrélation entre les mesures de précision et de fluence du sous-test MIM sur tous les items est également moyenne (r = 0,498, p < 0,01), ce qui correspond à 25% de variance partagée (r^2 = 0,25). Les corrélations sont encore plus faibles lorsque l'on considère les mesures de précision de MIM et les mesures de fluence des sous-tests « 1 minute » (mots : r = 0,405, p < 0,01; pseudomots : r = 0,389, p < 0,01; tous les items : r = 0,396, p < 0,01). Ceci représente ainsi 16% et 15% de variance partagée entre la précision dans MIM et la fluence dans les sous-tests 1 minute sur les mots et les pseudomots respectivement (mots : r^2 = 0,16; pseudomots : r^2 = 0,15). Il apparaît ainsi, tel qu'attendu, que les mesures de précision et de fluence réfèrent à des construits qui se recoupent, mais qui ne sont pas totalement identiques.

3.2.2.2 Corrélations entre la compréhension écrite et l'identification des mots écrits

La performance en compréhension écrite, mesurée par les réponses aux questions de compréhension, sur le texte « Le lion » (Bianco *et al.*, 2018), peut être mise en relation avec les différentes mesures de l'identification des mots écrits.

Tableau 3.16 Corrélations entre la compréhension écrite, la précision et la fluence de l'identification des mots et des pseudomots au sous-test MIM (BELEC) et au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

	Mesures d'identification des mots et des pseudomots								
	Précision mots et psmots (MIM)	Précision mots (MIM)	Précision psmots (MIM)	Fluence mots et psmots (MIM)	Fluence mots (MIM)	Fluence psmots (MIM)	Fluence mots et psmots (Test 1 min.)	Fluence mots (Test 1 min.)	Fluence psmots (Test 1 min.)
Compréhension écrite	0,229**	0,275**	0,145	0,300**	0,349**	0,189*	0,323**	0,311**	0,296**

Note. ** indique que p < 0,01. * indique que p < 0,05. Psmots est l'abbréviation pour « pseudomots ».

On remarque ainsi que, malgré des corrélations significatives (sauf celle de précision de lecture de pseudomots-MIM), la mesure de compréhension écrite est faiblement corrélée à presque toutes les

variables associées à l'identification des mots écrits (avec des *r* variant de 0,349 à 0,275, $p < 0,01$), hormis les mesures de précision sur tous les items et les mesures de fluence sur les pseudomots (MIM) qui sont, quant à elles, très faiblement corrélées quoique significatives ($r = 0,229$, $p < 0,01$; $r = 0,189$, $p < 0,05$). La mesure de précision sur les pseudomots (MIM) n'est pas corrélée du tout. D'aussi faibles corrélations sont relativement attendues avec une population d'élèves de cet âge. En effet, tel que mentionné dans le cadre théorique, le modèle *Simple View of Reading* (Gough et Tunmer, 1986) indique que la compréhension écrite résulte du produit de l'identification des mots écrits et de la compréhension langagière (comprise ici comme étant la compréhension orale). Dans plusieurs études, l'identification des mots écrits est généralement hautement corrélée à la compréhension écrite en tout début d'apprentissage, mais cette corrélation diminue à mesure que l'identification des mots écrits devient automatisée (Vellutino, Tunmer, Jaccard et Chen, 2007). C'est vers 9 ans (fin 3^{e} année du primaire) que les processus spécifiques d'identification des mots écrits sont réputés être relativement automatisés (Seymour *et al.*, 2003; Sprenger-Charolles, 2012). Les élèves de l'échantillon, ayant en moyenne 11 ans, voient donc leurs processus spécifiques d'identification des mots écrits peu corrélés à la compréhension écrite, confirmant ainsi l'automatisation de ces processus spécifiques, en tout cas chez la majorité d'entre eux.

Malgré qu'elles soient faiblement corrélées à la compréhension écrite, il est intéressant de noter que ce sont les mesures de fluence qui ont les corrélations les plus fortes avec la compréhension écrite et non pas les mesures de précision. Particulièrement, la fluence de tous les items (MIM et Test 1 minute) et la fluence des mots seulement (MIM et Test 1 minute) sont les 4 mesures qui ont les corrélations les plus fortes.

3.3 Interprétation des résultats : vers l'identification de lecteurs compensateurs au sein de l'échantillon

L'étude 1 a pour objectif l'identification des lecteurs compensateurs au sein d'un échantillon d'élèves de même niveau scolaire. Afin d'y arriver, une mesure de compréhension écrite ainsi que plusieurs mesures d'identification des mots écrits ont été prises auprès des élèves de l'échantillon. De manière générale, toutes les mesures mènent à des distributions acceptables des scores, sauf la mesure de précision en identification de mots seulement (MIM) qui comporte un effet plafond. Ainsi, bien que certaines distributions s'apparentent plus à la courbe normale que d'autres, toutes les autres épreuves peuvent être utilisées; les données obtenues correspondent globalement à ce qui est attendu par la littérature scientifique de ce domaine (voir la section résultats).

Pour arriver à identifier les lecteurs compensateurs à l'aide de ces mesures, il faut, d'une part, sélectionner les variables qui seront considérées et, d'autre part, déterminer les seuils pour chacune de ces variables. En d'autres mots, on doit déterminer avec précision les critères pour les mesures de compréhension écrite et d'identification de mots écrits qui correspondent au profil de lecteur compensateur souhaité. Cette identification, la plus juste possible, est extrêmement importante puisqu'elle est garante de la qualité de l'étude 2, présentée ci-après.

3.3.1 Compréhension écrite

En ce qui concerne la compréhension écrite, une seule épreuve a été retenue dans le cadre de cette étude. C'est donc cette mesure qui sera utilisée pour identifier les lecteurs compensateurs. Cela ne pose pas problème, puisque les données obtenues sont adéquates (distribution qui s'apparente à une distribution normale, avec une absence d'effet plafond).

Tel que mentionné précédemment, les lecteurs compensateurs ne doivent pas présenter une faible performance en compréhension écrite; c'est effectivement pour cela qu'on les appelle ainsi. Il convient alors de déterminer le seuil à partir duquel la compréhension écrite sera considérée faible dans l'épreuve Le Lion (Bianco *et al.*, 2018). Les élèves dont la performance se situera au-dessus de ce seuil répondront ainsi au critère de compréhension écrite.

Se fier à la moyenne et l'écart-type obtenus sur un grand échantillon d'élèves de même année scolaire est une bonne façon de situer la performance d'un élève en particulier par rapport aux autres. Traditionnellement, pour qualifier de « moyenne » la performance d'un élève par rapport aux autres élèves de l'échantillon, on considère l'intervalle qui s'étend entre un écart-type sous la moyenne et un écart-type au-dessus de la moyenne. Dans le même sens, pour qualifier de « faible » la performance d'un élève, on considère le critère d'un écart-type sous la moyenne, et moins. Dans le cas de cette épreuve administrée à une population de 165 élèves de 5^{e} année du primaire, la moyenne est 5,78/10 et l'écart-type est de 1,8. Le score qui correspond à un écart-type sous la moyenne est donc 3,98/10, arrondi à 4/10 (puisque le score individuel de l'épreuve ne peut être qu'un nombre entier). Les élèves qui ont un score de 4/10 ou moins seraient donc les élèves qui sont faibles sur cette mesure de compréhension écrite et, par conséquent, ne pourraient pas faire partie du groupe de lecteurs compensateurs. À l'inverse, on pourrait considérer que tous les élèves qui ont un score de 5/10 et plus sont « dans la moyenne ou mieux » en ce qui concerne cette mesure de compréhension écrite et, ce faisant, répondraient aux critères de

sélection des lecteurs compensateurs. Cependant, pour sélectionner les lecteurs compensateurs retenus pour l'étude 2, il semble plus judicieux d'élever légèrement ce critère jusqu'à un score de 6/10. Il convient de rappeler ici que la distribution des données de cette épreuve s'avère peu discriminante en ce qui concerne les scores moyens de l'épreuve (5/10 à 7/10); ainsi, 63,8% de l'échantillon a un score de 5/10 à 7/10, ce qui est considérable. Ce manque de discrimination dans les scores moyens amène à augmenter le seuil à 6/10 pour faire partie des lecteurs compensateurs, ce qui réduit les risques que les élèves sélectionnés aient une faible compréhension écrite. Comparé à la moyenne, ce seuil de 6/10 correspond à la moyenne arrondie (moyenne : 5,78/10). Pour qualifier la performance des élèves qui ont un score de 6/10 et plus, on utilise généralement l'expression « dans la moyenne supérieure, ou mieux ». Un résumé de ces qualificatifs de performance est présenté sur la courbe normale ci-dessous.

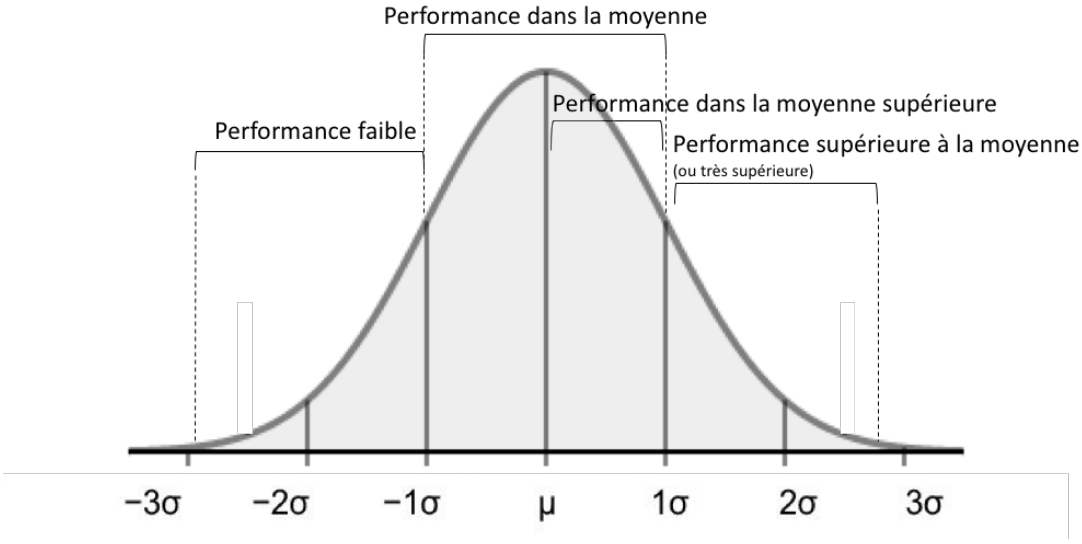


Figure 3.12 Terminologie utilisée pour qualifier la performance des élèves en comparaison à la moyenne du groupe, sur une courbe normale hypothétique

Dans l'échantillon, ce sont donc 98 élèves qui ont un score de compréhension écrite qui correspond à une performance dans la moyenne supérieure ou mieux. En conservant les cas limites des élèves qui ont obtenu 5/10, c'est une trentaine d'élèves qui auraient été conservés en plus. Quoi qu'il en soit, ces 98 élèves sont donc susceptibles de faire partie du groupe de lecteurs compensateurs, à condition évidemment que les critères de sélection concernant l'identification de mots soient également respectés.

3.3.2 Identification des mots écrits

Pour déterminer lesquels de ces 98 élèves, qui ont une bonne compréhension (score de 6/10 et plus), feront partie du groupe de lecteurs compensateurs, il convient maintenant de sélectionner quelles seront les variables de l'identification des mots écrits retenues ainsi que les seuils appropriés.

Tel que mentionné précédemment, les lecteurs compensateurs ont une identification des mots écrits déficitaire. Le seuil fixé d'une mesure d'identification de mots doit donc permettre d'identifier, avec le plus de justesse possible, les lecteurs faibles au regard de cette habileté. De la même manière que pour la compréhension écrite, la moyenne et l'écart-type seront considérés afin de déterminer le seuil correspondant à une performance faible (voir figure 3.13). Les élèves dont la performance se situe à un écart-type sous la moyenne, ou moins, sont donc les élèves « faibles » sur cette mesure.

Deux instruments de mesure ont été utilisés concernant l'identification des mots écrits : MIM (BELEC, Mousty *et al.,* 1994) et Test 1 minute (Sprenger-Charolles et Leloup, 2013). Le test MIM mène à 6 mesures distinctes de l'identification des mots écrits; le Test 1 minute en crée trois (voir la figure 3.14).

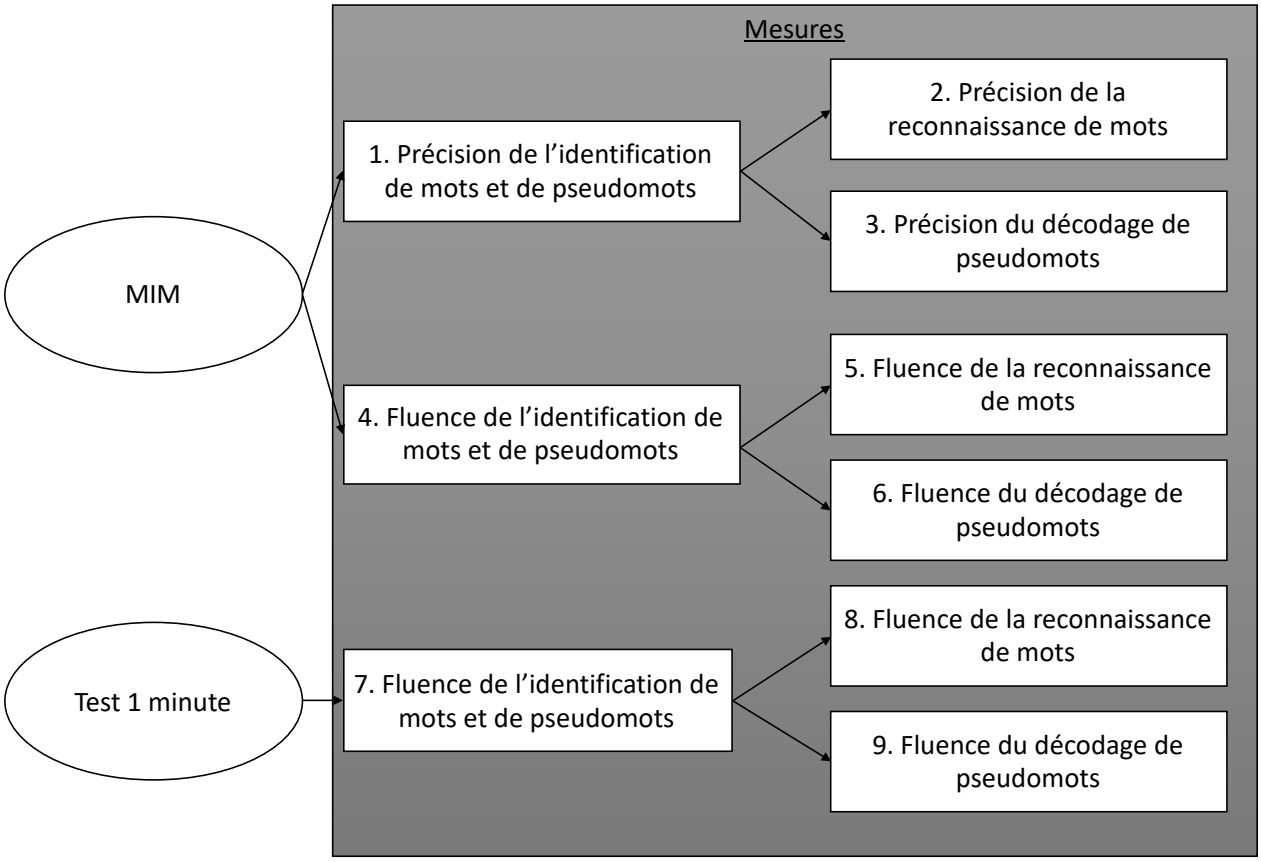


Figure 3.13 Les 9 variables de l'identification de mots et de pseudomots du sous-test MIM (BELEC) et du Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Pour la sélection des lecteurs compensateurs, il apparaît donc nécessaire d'observer si les élèves identifiés « faibles » par l'une de ces 9 mesures le sont également sur les autres mesures. Étant donné que les corrélations entre certaines de ces mesures sont relativement fortes (p.ex. les mesures de fluence obtenues par le MIM ou par le Test 1 minute), il semble a priori que ces mesures devraient permettre d'identifier comme « faibles » des élèves similaires. Les corrélations moyennes qui ont été constatées entre mots et pseudomots devraient, pour leur part, donner lieu à une sélection possiblement différente. Les corrélations plutôt faibles mesurées entre la précision et la fluence devraient quant à elles occasionner des sélections distinctes. En somme, si les élèves identifiés faibles (un écart-type ou plus sous la moyenne), ne sont pas les mêmes dans toutes les mesures de l'identification des mots écrits, il faudrait choisir judicieusement la ou les mesures à considérer pour le cas qui nous occupe, c'est-à-dire les lecteurs compensateurs. Les sections qui suivent font état de l'analyse des résultats obtenus au regard des différentes mesures d'identification des mots écrits et des décisions qui ont été prises quant aux critères permettant d'identifier les lecteurs compensateurs.

3.3.2.1 Précision ou fluence de l'identification des mots écrits

Parmi les 98 élèves qui ont une bonne compréhension écrite (score de 6/10 minimum), 33 élèves sont identifiés « faibles » sur l'une ou l'autre – ou plusieurs – mesures d'identification de mots. Ces élèves sont représentés dans le tableau ci-dessous par des numéros dans la colonne qui s'intitule « sujet ». Puisque le sous-test MIM est le seul qui contient à la fois une mesure de précision et une mesure fluence, ce sont ces résultats qui sont présentés dans le tableau pour fins de comparaison[22]. Les résultats sont présentés en score Z, ce qui fait en sorte qu'un score de -1 correspond au seuil de performance que l'on qualifie de « faible »; un score de -1,01 fait partie de l'intervalle des performances faibles.

[22] Effectivement, puisque le Test 1 minute ne comporte pas les mêmes items que le sous-test MIM, des différences entre le score de précision du MIM et de fluence du Test 1 minute pourraient être attribuables à la différence des items plutôt que la différence entre précision et fluence.

Tableau 3.17 Score Z des élèves « faibles » sur une des mesures du sous-test MIM (BELEC)

	MIM (mots et pseudomots)			MIM (mots et pseudomots)	
Participant	Précision	Fluence	Participant	Précision	Fluence
12	-1,78	-0,71	106	-1,25	-0,79
17	-0,73	-0,07	110	-1,25	0,25
19	-1,43	-0,86	112	0,32	-1,10
21	1,02	-1,30	117	1,02	-0,97
29	-1,25	-0,54	119	0,50	-1,83
37	-0,90	0,30	123	-3,36	-0,75
39	-1,25	1,13	128	-0,73	-1,09
42	-1,43	-0,41	131	-2,13	-1,51
53	-0,55	-0,80	133	0,15	-0,59
56	-0,55	-1,21	136	0,32	-1,19
57	-2,66	-2,31	137	-0,20	-1,13
70	-0,73	0,56	143	0,67	0,62
83	-0,38	-0,80	152	-0,38	-1,49
85	-1,78	-0,78	154	-0,55	-1,03
89	0,15	-0,66	163	-1,43	-1,62
90	-1,60	-1,27	165	-0,90	-0,36
105	-0,90	0,94			

Note. Les performances sont présentées en score Z. Les scores surlignés en gris foncé indiquent que la performance de l'élève est faible sur les deux mesures, alors que le score surligné en gris pâle signifie que la performance de l'élève est faible uniquement sur l'une des deux mesures. Les scores non surlignés correspondent à une performance qui n'est pas faible.

Ainsi, ce tableau indique que 11 élèves sur 33 (aucun surlignement) n'ont pas du tout été identifiés comme étant « faibles » par ces mesures du MIM (sujets 17, 37, 53, 70, 83, 89, 105, 117, 133, 143 et 165). 4 élèves (sujets 57, 90, 131 et 163) ont été identifiés tant par la mesure de précision que par la mesure de fluence. Neuf élèves (sujets 12, 19, 29, 39, 42, 85, 106, 110 et 123) ont été identifiés par la mesure de précision seulement, et 9 autres élèves (21, 56, 112, 119, 128, 136, 137, 152 et 154) ont été identifiés par la mesure de fluence uniquement. Il semble donc que, dans la majorité des cas d'élèves (18 élèves sur 33), une seule des deux mesures les identifie comme étant « faibles ». Il existe donc, dans ce sous-échantillon de 33 élèves, des élèves qui sont faibles en fluence mais pas en précision. Le comportement de ces élèves tend ainsi vers plusieurs bonnes réponses, fournies dans un temps relativement long en comparaison à leurs pairs. Il existe également dans ce sous-échantillon la situation inverse, où des élèves sont faibles en

précision mais pas en fluence. Leur comportement tend ainsi vers plusieurs mauvaises réponses, qui sont fournies dans un temps relativement raisonnable en comparaison à leurs pairs.

Selon la mesure choisie, les élèves identifiés comme étant faibles ne sont donc pas les mêmes. Il importe de choisir judicieusement laquelle de ces deux variables sera retenue pour identifier les lecteurs compensateurs de l'échantillon. Tel que mentionné dans le cadre théorique, le modèle de la compensation en lecture présenté par Walczyk (2000) propose qu'il existe deux manières de compenser : 1) contourner le processus déficitaire en recourant à d'autres processus pour obtenir l'information, ou 2) allouer beaucoup de ressources cognitives (mémoire de travail et attention) au processus déficitaire pour qu'il fonctionne le mieux possible et arrive à fournir une information. Ceci étant, il est possible que les élèves qui sont considérés faibles en fluence mais pas en précision soient des élèves qui utilisent efficacement ces deux moyens. Que ce soit pour contourner le processus déficitaire ou pour y allouer une partie importante des ressources cognitives disponibles, davantage de temps est nécessaire pour compenser. Les lecteurs qui compensent prennent ainsi plus de temps pour l'identification de mots que c'est le cas pour des pairs qui n'opèrent pas cette compensation. L'objectif de la mesure d'identification des mots écrits est de faire ressortir les difficultés et d'empêcher le plus possible le recours à la compensation. En ce sens, il semble tout à fait judicieux de sélectionner la mesure de fluence, étant donné que c'est celle qui considère le temps pris pour réaliser la tâche. Ce sont ainsi les mesures de fluence qui sont retenues. Le tableau suivant regroupe les élèves qui sont considérés faibles soit sur le plan de la précision, soit sur le plan de la fluence, soit les deux.

Tableau 3.18 Score Z des élèves « faibles » sur le plan de la précision de l'identification des mots écrits (à gauche) ou sur le plan de la fluence (à droite), au sous-test MIM (BELEC)

	MIM (mots et pseudomots)			MIM (mots et pseudomots)	
Participant	Précision	Fluence	Participant	Précision	Fluence
12	-1,78	-0,71	56	-0,55	-1,21
19	-1,43	-0,86	112	0,32	-1,10
29	-1,25	-0,54	21	1,02	-1,30
39	-1,25	1,13	119	0,50	-1,83
42	-1,43	-0,41	128	-0,73	-1,09
85	-1,78	-0,78	136	0,32	-1,19
106	-1,25	-0,79	137	-0,20	-1,13
110	-1,25	0,25	152	-0,38	-1,49
123	-3,36	-0,75	154	-0,55	-1,03
57	-2,66	-2,31			
90	-1,60	-1,27			
131	-2,13	-1,51			
163	-1,43	-1,62			

Note. Les performances sont présentées en score Z. Les scores surlignés en gris foncé indiquent que la performance de l'élève est faible sur les deux mesures, alors que le score surligné en gris pâle signifie que la performance de l'élève est faible uniquement sur l'une des deux mesures. Les scores non surlignés correspondent à une performance qui n'est pas faible.

3.3.2.2 Fluence de l'identification des mots écrits : quel type d'items?

Des 9 mesures qui ont été prises auprès des élèves de l'échantillon, 3 mesures concernent uniquement la précision de l'identification de mots. Puisque c'est la fluence qui est retenue ici, les résultats obtenus à l'aide de mesures de précision sont retirées des analyses qui suivent. Il ne reste plus qu'à départir les mesures de fluence (mesures 4 à 9 de la figure 3.14). Il convient d'aller observer d'abord si des différences de sélection de sujets existent lorsque l'on considère les mots ou les pseudomots.

Parmi les 33 élèves identifiés « faibles » sur l'une ou l'autre – ou plusieurs – des mesures d'identification de mots (voir tableau 3.19), 10 élèves ne sont pas du tout identifiés par les mesures de fluence (sujets 12, 17, 37, 39, 42, 53, 70, 105, 110 et 165, non surlignés dans le tableau suivant).

Tableau 3.19 Score Z des élèves « faibles » sur une mesure d'identification de mots et de pseudomots au sous-test MIM (BELEC) et au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

	Fluence au Test 1 min.		Fluence au sous-test MIM	
Participant	Mots	Psmots	Mots	Psmots
12	-0,54	-0,98	-0,74	-0,91
17	0,10	-0,81	-0,52	0,68
19	-1,39	-0,42	-1,38	0,23
21	-0,77	-1,50	-1,19	-0,99
29	-0,18	-0,52	-0,07	-1,31
37	-0,12	0,43	0,56	-0,38
39	1,61	1,83	1,42	0,07
42	-0,55	-0,58	-0,44	-0,77
53	-0,87	-0,81	-0,99	-0,61
56	-0,93	-0,81	-1,08	-1,27
57	-1,86	-2,18	-1,99	-2,12
70	0,81	-0,80	0,34	0,67
83	-0,59	-0,22	-0,52	-1,03
85	-0,25	-1,43	-0,49	-1,26
89	-0,40	-1,69	-0,40	-0,71
90	-1,76	-1,10	-0,99	-1,39
105	1,78	1,48	1,13	-0,05
106	-0,04	-0,52	-0,51	-1,22
110	0,49	-0,96	-0,04	0,39
112	-0,98	-1,69	-1,18	-0,76
117	-1,39	-1,20	-1,25	-0,40
119	-2,02	-1,50	-2,16	-0,82
123	0,19	-2,02	0,10	-2,45
128	-0,18	-0,42	0,01	-1,68
131	-0,04	-1,01	-0,80	-2,36
133	-0,11	-0,42	-0,01	-1,05
136	-1,34	-1,40	-1,03	-1,14
137	-0,47	-0,37	-1,09	-1,08
143	-1,00	0,49	1,00	0,11
152	-1,24	-0,81	-1,81	-0,61
154	-0,91	-0,42	-1,09	-0,77
163	-2,02	-1,20	-1,57	-1,24
165	-0,70	-0,52	-0,35	-0,28

Note. Les performances sont présentées en score Z. Les scores surlignés en gris foncé indiquent que la performance de l'élève est faible sur les deux mesures, alors que le score surligné en gris pâle signifie que la performance de l'élève est faible uniquement sur l'une des deux. Les scores non surlignés correspondent à une performance qui n'est pas faible. Psmot est l'abbréviation qui désigne « pseudomots ».

Le tableau 3.19 met en lumière le fait que, pour le Test 1 minute, 6 élèves (sujets 57, 90, 117, 119, 136 et 163) ont été identifiés par les deux mesures (surlignés en gris foncé), 6 élèves ont été identifiés par la mesure de décodage de pseudomots seulement (sujets 21, 85, 89, 112, 123 et 131, surlignés en gris pâle), et 3 autres élèves (sujets 19, 143 et 152, surlignés en gris pâle) ont été identifiés par la mesure d'identification de mots uniquement. Il existe donc effectivement, dans ce sous-échantillon de 33 élèves, des élèves qui sont identifiés « faibles » en fluence sur les pseudomots seulement mais pas sur les mots. Inversement, il existe aussi des élèves qui sont identifiés « faibles » sur les mots seulement mais pas sur les pseudomots, en quantité tout de même moins importante. Un portrait similaire se dégage des résultats de fluence obtenus pour MIM, où 5 élèves (sujets 56, 57, 136, 137 et 163, surlignés en gris foncé) ont été identifiés par les deux mesures (mots et pseudomots); 9 élèves ont été identifiés par la mesure de décodage de pseudomots seulement (sujets 29, 83, 85, 90, 106, 123, 128, 131 et 133, surlignés en gris pâle) et 7 autres élèves[23] (sujets 19, 21, 112, 117, 119, 152 et 154, surlignés en gris pâle) ont été identifiés par la mesure d'identification de mots uniquement. Se référant aux modèles développementaux de l'identification des mots écrits et aux données présentées dans le cadre théorique, il n'est pas surprenant que davantage d'élèves soient identifiés par la mesure de pseudomots, puisque le traitement alphabétique qui sous-tend le décodage revêt un caractère primordial pour l'apprentissage de l'identification des mots écrits.

Quoi qu'il en soit, encore ici, puisque les élèves identifiés « faibles » par les mesures sur les mots et les pseudomots ne sont pas les mêmes, il convient de choisir judicieusement laquelle de ces mesures sera utilisée. Comme il y a des disparités importantes dans la sélection des sujets, il ne semble pas approprié de sélectionner la mesure de fluence qui combine tous les items (correspond à la moyenne de la performance sur les mots et les pseudomots), puisque des subtilités qui peuvent s'avérer importantes d'un point de vue conceptuel entre les traitements cognitifs mobilisés sur les mots et les pseudomots seraient « diluées » dans ce ratio moyen. C'est pourquoi les mesures de fluence des mots et pseudomots par le MIM et le test 1 minute, de la figure 3.14, ne sont donc pas retenues. Par ailleurs, puisqu'il faut choisir entre la mesure qui concerne les mots ou celle qui concerne les pseudomots, il convient de discuter de ce choix sous l'éclairage de la compensation. Tel que soulevé dans le cadre théorique, une des possibilités de compensation est le recours à la sémantique (voir la section et les données empiriques sur

[23] Le fait que davantage d'élèves soient identifiés par la mesure sur les mots seulement dans le MIM pourrait être attribuable au type de mots présents dans la liste. Plusieurs sont si rares qu'ils pourraient s'apparenter à des pseudomots. L'interprétation des résultats doit donc être faite en conséquence.

l'effet de contexte). En situation de lecture de mots isolés, les études sur l'effet de contexte mettent en lumière le recours à la sémantique lexicale pour soutenir l'identification des mots écrits déficitaire. En ce sens, la reconnaissance de mots permet le recours à la sémantique, alors que le décodage de pseudomots l'empêche complètement. Dans la perspective où la mesure d'identification des mots écrits doit faire ressortir les difficultés, il semble pertinent de conserver uniquement la mesure de fluence en décodage de pseudomots.

Tant le sous-test de MIM que le Test 1 minute ont une mesure de fluence de pseudomots. Par contre, la consigne du Test 1 minute met davantage l'accent sur la vitesse que ne le fait la consigne du MIM. Il semble également que, puisqu'elle ne dure qu'une minute, cette mesure nécessite moins un effort cognitif soutenu que celle du MIM; il y a donc moins de chance que le résultat obtenu inclue des items ratés attribuable à un effort prolongé. C'est donc le Test 1 minute de décodage de pseudomots qui a été retenu pour la sélection des lecteurs compensateurs.

3.3.3 Sélection finale des lecteurs compensateurs

Pour conclure cette section, il convient de revenir sur l'ensemble des critères de sélection des lecteurs compensateurs de l'échantillon de 165 élèves de 5^{e} année du primaire. Les lecteurs compensateurs doivent être des élèves qui ont une relativement bonne compréhension écrite et qui ont par ailleurs une identification des mots écrits faible. Dans les sections précédentes, il a été choisi de conserver uniquement la mesure de compréhension écrite « Le lion » et la mesure de fluence du Test 1 minute sur le décodage de pseudomots. Pour être identifiés comme lecteurs compensateurs, les élèves doivent donc avoir un score de compréhension écrite dans la moyenne supérieure ou mieux (6/10 ou plus) et une performance sous la moyenne pour la fluence du décodage de pseudomots (23,05 pseudomots bien lus en 1 minute). La figure ci-dessous situe par un point chaque élève de l'échantillon de 165 élèves; chacune des coordonnées dans le graphique correspond au score Z en compréhension écrite (axe y) et en fluence de pseudomots (axe x).

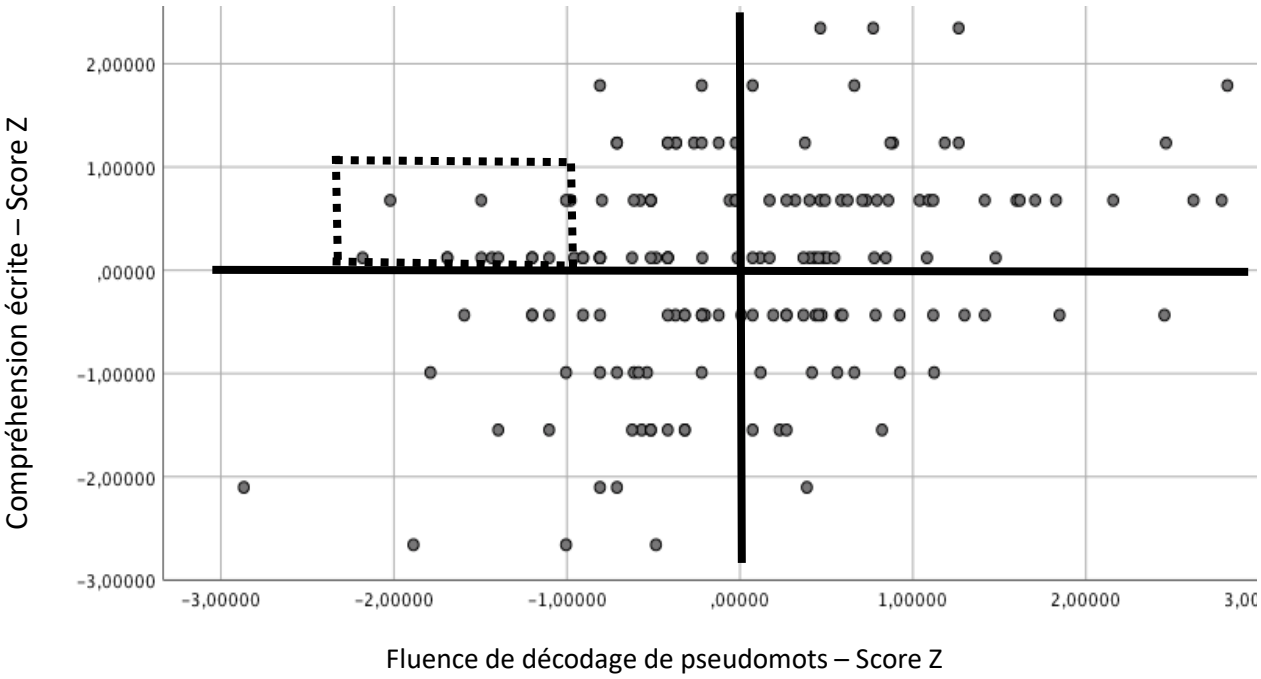


Figure 3.14 Répartition des élèves selon leur score Z de compréhension écrite (axe des y) et de fluence de décodage des pseudomots au Test 1 minute (Sprenger-Charolles et Leloup, 2013)

Note. Les lignes noires représentent la moyenne pour chacune des épreuves (5,78 pour la compréhension écrite, et 33,25 pour la fluence de décodage de pseudomots). L'encadré pointillé désigne les élèves qui rencontrent les critères de lecteurs compensateurs dans l'échantillon (score de 6 ou mieux en compréhension écrite, et score de 23,05 ou moins pour la fluence du décodage de pseudomots)

On remarque ainsi que les lecteurs compensateurs se situent dans le quadrant supérieur gauche de la figure, avec un score en compréhension écrite se situant dans la moyenne supérieure et un score de fluence de décodage de pseudomots sous la moyenne. La majorité des élèves ont un score de fluence de pseudomots qui suit généralement bien le score en compréhension écrite, puisque la grande majorité des points qui représentent les scores de chaque élève suit la droite attendue. Les lecteurs compensateurs quant à eux sont éloignés de cette droite attendue, ils témoignent donc d'un profil peu typique. Il convient également de noter qu'aucun lecteur compensateur n'a un score supérieur à 7/10 en compréhension écrite, tel qu'en témoigne le tableau ci-dessous. Ceci souligne le fait que la compensation que les lecteurs compensateurs mettent en œuvre a sans doute une limite; quand on lit un texte avec une difficulté d'identification des mots écrits, cela semble empêcher une compréhension maximale.

CHAPITRE 4

ÉTUDE 2 – MÉTHODOLOGIE ET RÉSULTATS

L'étude 1 a permis d'identifier des lecteurs de la fin du primaire dont le comportement en lecture s'apparente à de la compensation. Ces élèves, appelés lecteurs compensateurs, représentent des participants potentiels à l'étude 2. L'étude 2 vise à répondre à la question suivante : le traitement sémantique des lecteurs compensateurs est-il mobilisé de manière particulière en identification de mots écrits présentés de manière isolée? Elle vise ainsi à vérifier les hypothèses suivantes, la première ayant trait à des données comportementales, la seconde ayant trait à des données d'électroencéphalographie : 1) l'effet facilitateur d'un lien sémantique entre l'amorce et la cible devrait se manifester de manière plus forte chez les lecteurs compensateurs que chez les contrôles, et ce, autant lorsque l'amorce est présentée rapidement que lorsque l'amorce est présentée plus longuement; 2) les lecteurs compensateurs activeraient plus tôt l'information sémantique pendant l'identification des mots écrits que ne le font les lecteurs contrôles, et ce, malgré un déficit du traitement de la forme visuelle et phonologique des mots. Dans un premier temps, un commentaire général sur l'approche méthodologique retenue pour cette étude est formulé. Ensuite, la méthodologie est décrite, puis les résultats sont présentés. Une discussion des résultats obtenus conclut finalement le chapitre.

4.1 Commentaire général sur l'approche méthodologique

La présente étude, tout comme la précédente, s'inscrit dans le courant positiviste. En effet, comme l'identification de mots écrits se déroule extrêmement rapidement, souvent en moins d'une demi-seconde, même chez le lecteur en difficulté (Rayner, 1988; Ferrand, 2007), et que plusieurs mécanismes peuvent être mis en œuvre de façon automatique ou inconsciente au cours de cette identification (p. ex. Laberge et Samuel, 1974; Perfetti, 1985; Walzyck *et al.*, 2004), il est nécessaire de recourir à des données qui ne dépendent pas de la capacité de verbalisation et d'introspection du sujet, comme c'est le cas dans le paradigme interprétatif (Creswell, 2003; Cromley et Azevedo, 2007; Fox *et al.*, 2011; Fraenkel et Wallen, 2000). Dans le courant positiviste, pour l'étude de l'identification des mots écrits, on soumet généralement les participants à des tâches de lecture d'items isolés, où différentes variables pouvant influencer les traitements mis en œuvre font l'objet d'une manipulation expérimentale et d'un contrôle minutieux (p. ex. Alegria et Mousty, 1996). Les mesures collectées concernent principalement la précision des réponses, mais aussi le temps de réponse. Le choix des variables à manipuler, ainsi que la façon d'interpréter les données ainsi obtenues sont tributaires de modèles ou de théories existantes sur le sujet; il s'agit donc

d'une approche hypothético-déductive. Dans ce paradigme, on considère qu'une réalité complexe (p. ex. la lecture) doit être décomposée en phénomènes isolés plus simples pour pouvoir être adéquatement étudiée (Creswell, 2003; Kazdin, 2003). De plus, on considère qu'il existe une certaine part de réalité objective, que l'on peut observer et étudier sans nécessairement recourir à l'interprétation que le participant en fait. C'est ainsi que l'on accorde une signification à des résultats mesurables et objectifs comme la précision des réponses et le temps de réponse, pour lesquels les participants ne fournissent pas eux-mêmes d'interprétation. Ce choix paradigmatique permet ainsi d'étudier des phénomènes qui revêtent un caractère inconscient ou automatique. Cependant, il est essentiel de garder en tête que la décomposition d'une tâche aussi complexe que la lecture en de plus petites unités, plus simples, comme l'identification de mots écrits présentés de manière isolée, ne permet souvent pas de rendre compte de toute la complexité d'une situation ou d'un phénomène. Finalement, afin d'être en mesure de tirer des conclusions justes à propos des résultats obtenus sur des tâches décontextualisées impliquant certains des processus de la compétence à lire, il est primordial de contrôler très minutieusement les variables autres que celles à l'étude, qui pourraient influencer l'identification de mots écrits, ceci afin d'assurer que les différences observées sont attribuables aux variables d'intérêt de l'étude plutôt qu'à ces autres variables.

Qui plus est, afin d'être en mesure de bien distinguer les sous-composantes de l'identification des mots écrits, il convient de recourir à l'électroencéphalographie (EEG) en plus des données comportementales liées à la précision et aux temps de réponse. La précision des réponses ainsi que les temps de réponse dans une tâche de décision lexicale mesurent le résultat de la mise en œuvre de toutes les étapes de l'identification de mots écrits, en plus des processus liés à la prise de décision et ceux liés au mouvement moteur pour appuyer sur un bouton. Les composantes EEG, quant à elles, permettent une mesure en continu, dès l'apparition du mot à identifier à l'écran, et offrent la possibilité de cibler le décours temporel d'une variable linguistique en particulier. Par exemple, il est possible de se centrer sur le traitement sémantique dans l'identification des mots écrits, et d'en repérer les moments d'influence bien avant que le participant ne fournisse une réponse motrice. La considération conjointe des données comportementales (précision et temps de réponse) et des données de l'activité électrique du cerveau permet un regard particulier sur certains phénomènes; des questionnements peuvent être soulevés avec cette technique, questionnements qu'il serait difficile d'envisager avec l'une ou l'autre de ces techniques seules. La considération de la précision des réponses, des temps de réponse et de l'activité électrique en

continu permet ainsi un nouvel éclairage sur la question du moment auquel la sémantique peut être mobilisée dans l'identification des mots écrits.

Quoique quasiment absentes en didactique cognitive, de telles recherches recourant à l'EEG pour l'étude de l'identification des mots écrits existent dans le paradigme positiviste, notamment en psychologie cognitive ou en linguistique, tel que rapporté dans le cadre théorique.

4.2 Méthodologie

Afin de vérifier les hypothèses mentionnées ci-dessus, un devis expérimental de groupe avec comparaisons intra-sujet et inter-sujets est mis en place. Les éléments méthodologiques associés à ce devis sont présentés dans les sections qui suivent.

4.2.1 Devis de recherche et description des variables

Cette section présente le devis expérimental de groupe mis en place afin de tester les hypothèses concernant le rôle de la sémantique lexicale chez les lecteurs compensateurs. Une tâche expérimentale est élaborée, dans laquelle une variable indépendante est manipulée. Dans cette tâche, la manipulation de cette variable indépendante, plus particulièrement les comparaisons entre les conditions qui sont associées à cette variable, permettront de tester les hypothèses. Puisque les hypothèses de recherche concernent le traitement des aspects sémantiques impliqués dans l'identification des mots écrits, c'est la sémantique lexicale qui constitue cette variable indépendante. Tel que mentionné dans le cadre théorique, les tâches où le participant doit identifier un mot et prendre une « décision » par rapport à celui-ci sont celles qui sont les plus utilisées pour étudier l'identification des mots écrits (Ferrand, 2007). La tâche expérimentale implique donc que les participants déterminent si l'item présenté de manière isolée sur l'écran d'ordinateur est un vrai mot du français ou ne l'est pas, ce que l'on appelle « décision lexicale » [24]. Particulièrement pour cibler les aspects sémantiques de l'identification des mots écrits, ce sont les tâches d'amorçage sémantique qui sont les plus pertinentes (voir le cadre théorique pour les détails). C'est donc

[24] Il convient de noter ici qu'il aurait été possible de sélectionner une tâche de décision sémantique plutôt qu'une tâche de décision lexicale. Une tâche de décision sémantique implique que le participant indique si le mot présenté fait partie d'une catégorie donnée (le mot représente-t-il un « vivant » ou un « non vivant »). Puisque nous supposons que les lecteurs compensateurs recourent naturellement à la sémantique lexicale pour compenser leur difficulté, il nous a semblé plus pertinent de ne pas attirer particulièrement l'attention des lecteurs sur la sémantique.

cette tâche qui est sélectionnée ici. Ainsi, on prévoit des paires de mots (mot-amorce et mot-cible) dont les concepts sont reliés sémantiquement, et d'autres qui ne le sont pas.

Cette variable indépendante, soit la sémantique lexicale, consiste en une variable intra-sujet puisque celle-ci sera manipulée de façon à créer différentes conditions pour un même sujet; les performances dans chacune de ces conditions seront ensuite comparées entre elles, pour un même participant. Une deuxième variable indépendante est présente dans ce devis. En effet, la performance des lecteurs compensateurs sera comparée à celle d'élèves contrôles, afin d'être en mesure de bien interpréter leur performance. La deuxième variable indépendante est donc le type de lecteur (soit lecteur compensateur, soit lecteur contrôle). Cette deuxième variable indépendante est une variable inter-sujets, puisque les comparaisons entre les différentes conditions impliquent différents participants.

Finalement, il convient de noter que la tâche sera réalisée une fois avec un délai entre la présentation de l'amorce et de la cible (SOA) qui est court, puis une fois avec un SOA qui est long. C'est cette variation du délai de présentation qui permettra de tester la sous-hypothèse qui concerne l'effet facilitateur d'un lien sémantique qui devrait se manifester de manière plus forte chez les lecteurs compensateurs, autant lorsque l'amorce est présentée rapidement que lorsque l'amorce est présentée plus longuement. Puisque les items pour le SOA court et ceux pour le SOA long ne sont pas appariés, il ne s'agit pas d'une variable indépendante de l'étude.

En résumé, le devis de recherche général de cette étude est représenté par la figure suivante :

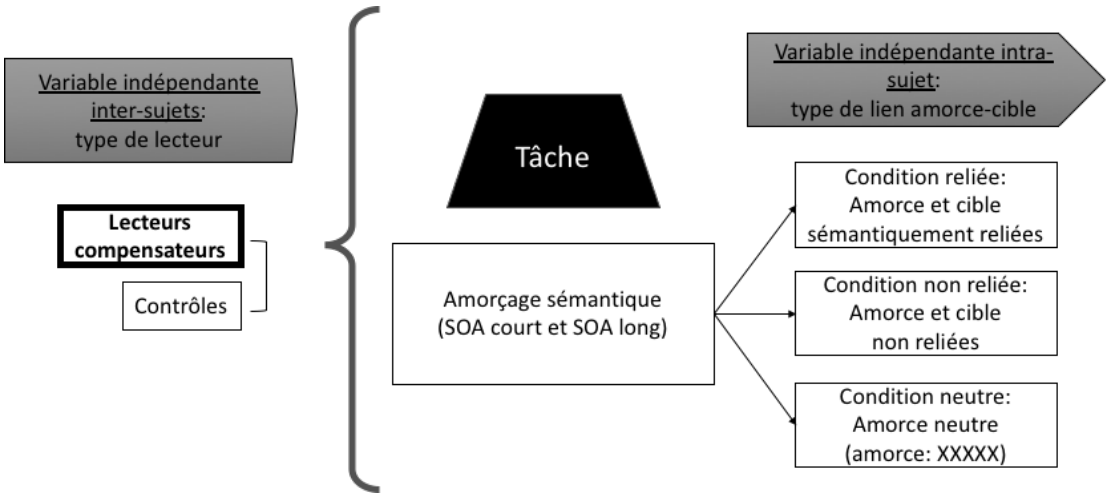


Figure 4.1 Résumé du devis de recherche de l'étude 2 mettant en évidence les variables indépendantes

Les variables présentées dans ce schéma sont décrites en détails dans les prochaines sections, ainsi que les variables de contrôle et les variables dépendantes.

4.2.1.1 Description des variables indépendantes

Les variables indépendantes de l'étude ont été brièvement discutées à la section précédente, mais sont reprises ici pour davantage de détails.

La première variable indépendante correspond à la variable qui est manipulée par l'expérimentateur au sein des tâches d'amorçage sémantique avec décision lexicale afin de créer différentes conditions pour un même sujet. Dans le cas présent, la manipulation du type de lien sémantique entre l'amorce et la cible permet de mesurer l'apport de la sémantique lexicale à l'identification des mots écrits. En effet, l'amorce fournit un « contexte » à l'identification de la cible. Dans la première condition, le mot écrit présenté en amorce est sémantiquement relié au mot écrit présenté comme cible (p. ex. rame-BARQUE); la présence de l'amorce établit ainsi un contexte sémantique qui précède l'identification du mot-cible. Dans la deuxième condition qui est utilisée dans les comparaisons, le mot écrit présenté en amorce n'est pas sémantiquement relié au mot-cible (p. ex. athlète-ROSEAU); le contexte mis en place n'est donc pas un contexte sémantique. La troisième condition présente une amorce neutre constituée d'une série de « X » (p. ex. XXXXX-CRATÈRE); le contexte mis en place par l'amorce est donc un contexte neutre, qui ne contient aucun mot. La comparaison de la performance des lecteurs entre ces différentes conditions permettra de

vérifier des hypothèses de recherche. Dans la première condition, l'amorce et la cible sont considérées comme étant reliées sémantiquement, soit parce qu'elles ont une forte association verbale, soit parce que les concepts désignés par l'amorce et la cible partagent un nombre important de traits sémantiques (voir le chapitre 2 pour les explications conceptuelles). Les items de la condition non reliée sont constitués quant à eux de deux mots écrits qui ne sont pas reliés (ni par catégorie ni par fonction) et n'ont pas de trait sémantique commun. Les items de la condition non reliée ne partagent pas non plus de suite de phonèmes communs ni de suite de lettres similaires, pour éviter un amorçage phonologique ou orthographique.

La seconde variable indépendante correspond au type de lecteur concerné par l'étude. Les lecteurs visés par le sujet de cette thèse sont les lecteurs compensateurs. La manière de repérer des lecteurs compensateurs dans un échantillon d'élèves de la 5e année du primaire a fait l'objet de l'étude 1. La définition de ce type de lecteur ainsi que les critères permettant d'assigner des élèves au type « lecteur compensateur » dépendent ainsi des résultats obtenus à l'étude 1, qui sont brièvement rappelés ici. De manière théorique, les lecteurs compensateurs sont définis comme étant des lecteurs qui présentent une difficulté importante au niveau des processus spécifiques d'identification des mots écrits, sans toutefois que leur compréhension écrite ne soit dramatiquement affectée. De manière empirique, tel qu'établi par l'étude 1, les lecteurs compensateurs sont donc des élèves qui présentent une très faible performance au « Test 1 minute de lecture à voix haute de mots inventés » (Test 1 minute de pseudomots; Sprenger-Charolles et Leloup, 2013), mais dont le score obtenu en compréhension écrite (Épreuve « Le lion », Bianco *et al.*, 2018) est minimalement dans la moyenne supérieure. Les lecteurs compensateurs sont donc effectivement des élèves qui compensent une difficulté marquée en identification des mots écrits pour tout de même parvenir à comprendre un texte qu'ils ont lu.

Concernant le groupe contrôle, la principale caractéristique des élèves de ce groupe est qu'ils ne doivent pas compenser en lecture, contrairement aux lecteurs compensateurs. Ces lecteurs contrôles doivent également représenter une forme d'« équivalence » aux lecteurs compensateurs. Cette équivalence implique que le moins d'éléments possible les distinguent, à l'exception de ce que l'on veut observer, c'est-à-dire la compensation en lecture. Il a donc été choisi de sélectionner des élèves provenant du même échantillon que les lecteurs compensateurs (échantillon de participants à l'étude 1); les élèves proviennent des mêmes milieux, des mêmes écoles et sont exposés aux mêmes pratiques pédagogiques. Ils sont donc

de même âge chronologique que les lecteurs compensateurs[25]. Ils présentent également un score similaire en compréhension écrite; ils ne présentent toutefois pas de déficit au niveau de l'identification des mots écrits; c'est-à-dire un score au moins dans la moyenne supérieure au « Test 1 minute de pseudomots » (Sprenger-Charolles et Leloup, 2013). Puisque ces élèves ne présentent pas de difficulté en identification de mots écrits, on peut ainsi déduire qu'ils n'ont pas besoin de compenser une lacune sur ce plan.

4.2.1.2 Description des variables dépendantes

Les variables dépendantes de cette étude sont de nature comportementale ainsi que psychophysiologique. Les variables dépendantes de nature comportementale sont la précision des réponses ainsi que le temps de réponse, et ce, pour les deux tâches. Concernant la précision des réponses, l'item est considéré comme réussi s'il remplit la condition suivante : la cible présentée à l'écran est un mot et elle est identifiée comme tel par le participant par la pression d'un bouton avant la limite de temps imparti pour ce faire. La somme des items réussis est calculée, puis est ensuite convertie en proportion, c'est-à-dire un pourcentage d'items réussis. La variable « précision des réponses » est donc mesurée par le « taux de réussite » des items de la tâche. La deuxième variable dépendante comportementale concerne les temps de réponse, c'est-à-dire le temps qui s'écoule entre la présentation du mot-cible à l'écran et la pression d'un bouton par le participant pour indiquer sa réponse. Ici, seules les bonnes réponses sont considérées, afin d'assurer que le temps mesuré a effectivement servi à faire les traitements cognitifs prévus plutôt que n'importe quel autre qui aurait mené à une réponse erronée. Ainsi, pour étudier le phénomène qui nous intéresse, les temps pris pour fournir une bonne réponse à chaque item-cible d'une condition sont moyennés, pour chaque groupe de participants. Les comparaisons entre toutes ces moyennes de temps (en millisecondes) permettent d'observer s'il y a des différences de traitement entre les différentes conditions présentées, et entre les types de lecteurs.

Finalement, la dernière variable dépendante de cette recherche est de nature psychophysiologique. En effet, l'activité électrique émise par le cerveau est mesurée au moyen de l'électroencéphalographie (EEG). Cette mesure de l'activité électrique qui émane d'un groupe de neurones œuvrant de manière synchronisée fournit des données continues pendant l'identification d'un mot écrit par les participants. En

[25] Il aurait été très intéressant d'avoir des élèves de même niveau de lecture en plus d'élèves de même âge chronologique. Cela s'est effectivement avéré primordial dans les études portant notamment sur les élèves présentant un trouble spécifique d'apprentissage de la lecture-écriture (Vellutino, 1979). Pour une raison de faisabilité, un seul groupe contrôle a été choisi pour cette recherche doctorale, mais les travaux pourraient se poursuivre en ce sens ultérieurement.

ce qui concerne cette thèse, les données obtenues par l'EEG fournissent des indications quant au potentiel électrique qui est provoqué par la présentation de l'item-cible; généralement désigné par le terme « mesure des potentiels évoqués par un événement », ou « *event-related potentials »* (ERP) en anglais. Les potentiels électriques évoqués par un item-cible sont mesurés chaque milliseconde, sur chacune des 64 électrodes apposées sur la tête du participant, à partir de l'apparition de l'item-cible à l'écran jusqu'à la réponse fournie. Ces potentiels évoqués sont ensuite moyennés à chaque milliseconde pour chacune des conditions, pour chacun des groupes de participants[26].

4.2.1.3 Description des variables contrôles

Plusieurs variables contrôles sont nécessaires dans ce devis de recherche. Ces variables contrôles permettent de renforcer la qualité de l'étude et d'assurer des interprétations adéquates des données obtenues.

Afin d'assurer que les différences mesurées dans les différentes conditions d'items sont effectivement attribuables à la manipulation des variables indépendantes et non pas à d'autres variables, certaines caractéristiques linguistiques des mots doivent être conservées constantes à travers les conditions « relié », « non relié » et « neutre ». Ainsi, les différences observées sont plus susceptibles d'être attribuables au type de lien sémantique qu'à d'autres caractéristiques linguistiques. Les détails de ces variables contrôles sont décrits à la section 4.2.1.3. De manière succincte, on peut indiquer que tous les mots choisis sont des noms communs concrets, peu polysémiques, qui ne contiennent pas d'affixe. Certaines autres caractéristiques linguistiques doivent être contrôlées, mais ne peuvent être conservées constantes à travers les conditions, puisque cela s'avère impossible. Dans ces cas, ces variables sont « balancées » à travers les différentes conditions de sorte que, considérés ensemble, les mots écrits de la condition « relié » sont « équivalents » à ceux de la condition « non relié » et « neutre ». C'est le cas de la fréquence orale et écrite, de la complexité syllabique et graphémique, de la présence de graphie contextuelle, etc.

[26] Un certain nombre de détails supplémentaires sont nécessaires afin d'être en mesure de faire des comparaisons entre les conditions et entre les groupes. Ces détails primordiaux sont décrits dans la section sur la préparation des données EEG (section 4.2.5).

Puisque la tâche implique une décision lexicale, il faut des cibles qui sont des mots et d'autres qui sont des pseudomots. Pour ne pas induire une stratégie chez les participants en faveur de l'un ou l'autre de ce type d'items, les mots et les pseudomots doivent être présentés en nombre égal dans la tâche, et alternés de manière aléatoire. De cette façon, l'incertitude maximale est créée pour le participant. La lexicalité des cibles est ainsi contrôlée.

Finalement, certaines caractéristiques des participants sélectionnés doivent également être contrôlées afin que les participants soient les plus similaires possible. Ces caractéristiques représentent des variables contrôles entre les participants et sont de diverses natures. Elles sont décrites en détails à la section 4.2.1.3., mais à titre d'exemple, on peut mentionner le fait de ne pas retenir les élèves qui ont un retard intellectuel ou un trouble grave non médicamenté. D'autres instruments de mesure servent également à considérer la performance des élèves sur des composantes fortement associées à la lecture-écriture dans la littérature scientifique. Ces mesures concernent le vocabulaire réceptif, la fluence verbale, la mémoire de travail phonologique, la conscience phonémique ainsi que la dénomination rapide. Le fait de considérer la performance des élèves sur ces composantes permettra d'assurer que des différences entre les lecteurs compensateurs et les lecteurs contrôles ne soient pas simplement le résultat d'une différence de performance sur l'une ou l'autre de ces composantes. Les mesures de ces composantes sont décrites en détails à la section 4.2.1.3.

4.2.2 Participants

L'échantillon vise les lecteurs compensateurs d'une population du primaire, ainsi que des élèves de même niveau scolaire qui agissent à titre de contrôles.

De la même manière que dans l'étude 1, les lecteurs compensateurs sont définis dans cette thèse comme étant ceux qui présentent une difficulté sur le plan des processus spécifiques d'identification des mots écrits, sans que leur habileté à comprendre un texte écrit n'en soit affectée considérablement. Afin d'agir à titre de contrôles, les élèves doivent présenter la même habileté de compréhension écrite que les lecteurs compensateurs, sans toutefois avoir de difficulté sur le plan de l'identification des mots écrits.

Les processus de lecture étant des caractéristiques constitutives des individus, il n'est pas possible d'assigner aléatoirement les élèves à l'un ou l'autre des groupes. Il s'agit donc plutôt d'observer ces caractéristiques chez les élèves et de les assigner à l'un ou l'autre des groupes le cas échéant.

4.2.2.1 Recrutement des participants

L'approbation éthique pour cette étude a été obtenue en même temps, et exactement de la même manière que pour l'étude 1. Des modifications à cette approbation ont toutefois dû être réalisées auprès de l'Université du Québec à Montréal à la suite de modifications d'échéanciers pour la collecte de données.

Le recrutement a débuté à l'automne 2019, alors que les participants potentiels en étaient à débuter leur 6e année du primaire. Pour cette étude, le recrutement ne peut avoir lieu auprès de tous les enfants d'une classe, puisque les participants doivent être sélectionnés selon leur appartenance potentielle à l'un ou l'autre des types de lecteurs visés (lecteurs compensateurs ou contrôles); il est donc nécessaire d'avoir des informations quant à la performance des élèves sur certaines mesures de lecture avant de procéder au recrutement.

Ces critères pour les lecteurs compensateurs ont été établis rigoureusement dans le cadre de l'étude 1 : un score de 6/10 ou plus en compréhension écrite (« Le lion », Bianco *et al.*, 2018), ainsi qu'un score de 23,05 pseudomots bien lus en 1 minute pour l'identification des mots écrits (Test 1 minute de pseudomots, Sprenger-Charolles et Leloup, 2013). Tel que décrit à la section 4.2.1.1, les élèves contrôles présentent la même performance en compréhension écrite (6/10 ou plus). Puisque les résultats de l'étude 1 indiquent que les élèves du groupe de lecteurs compensateurs ont des scores de 6/10 ou de 7/10 en compréhension écrite, ce sont des élèves qui obtiennent ces scores qui sont potentiellement sélectionnés comme contrôles. Ces derniers ne présentent pas de difficulté en identification des mots écrits et ont conséquemment un score au moins dans la moyenne supérieure au Test 1 minute de pseudomots (Sprenger-Charolles et Leloup, 2013), sans être au-delà de 2 écart-types au-dessus de la moyenne.

Ce sont donc les élèves ayant participé à l'étude 1 qui ont constitué la base de recrutement pour l'étude 2. Tel que mentionné dans les résultats de l'étude 1, 12 élèves de l'échantillon initial répondent aux critères du groupe des lecteurs compensateurs. Pour les contrôles, ce sont 34 élèves qui ont des scores qui correspondent aux critères.

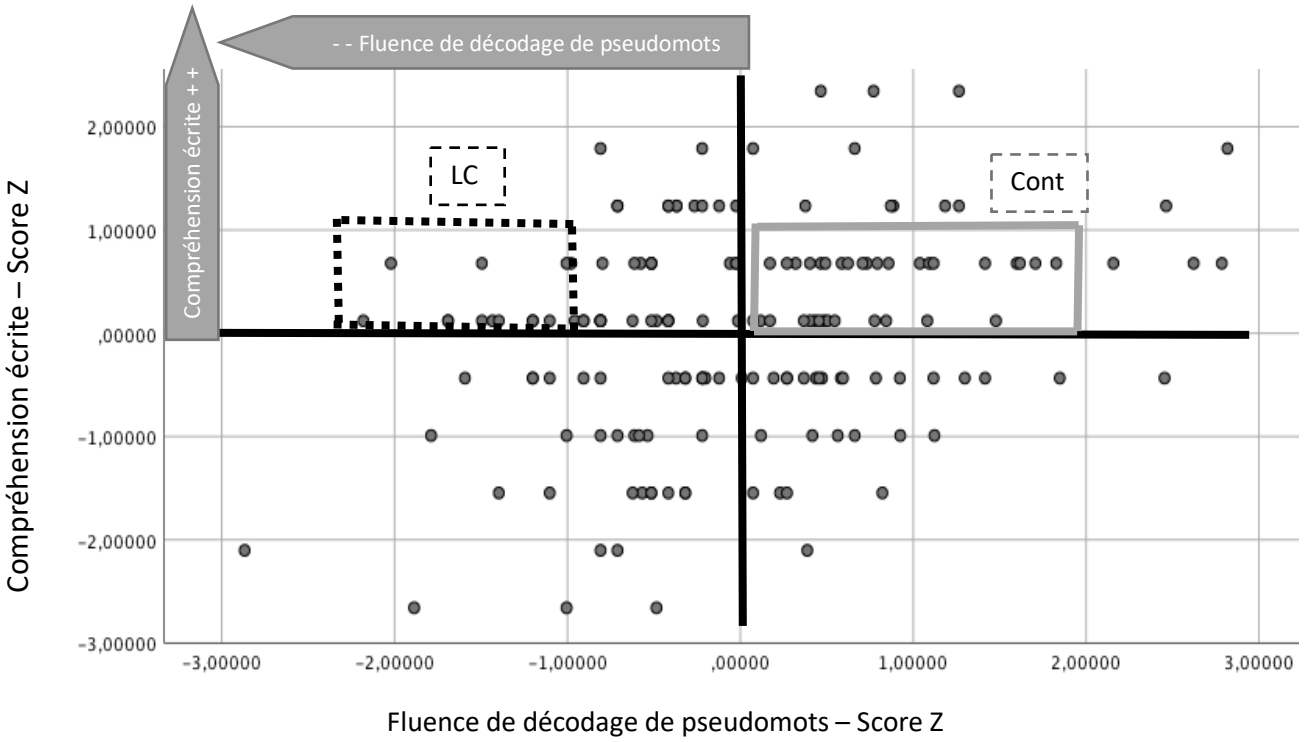


Figure 4.2 Répartition des élèves de l'échantillon de l'étude 1, selon leur score de compréhension écrite et en fluence de décodage de pseudomots, et identification des participants potentiels à l'étude 2

Note. Mise en évidence des élèves qui rencontrent les critères de lecteurs compensateurs (LC) dans l'encadré pointillé et de contrôles (Cont) dans l'encadré grisé.

L'étude 2 se déroulant sur une année scolaire différente de l'étude 1 (les élèves potentiellement participants sont en 6e année), un centre de services scolaire n'a pas été en mesure de poursuivre la collaboration. Par conséquent, trois lecteurs compensateurs de l'échantillon initial n'ont pu faire partie de l'étude 2. De plus, à cause de certains critères d'exclusion, qui seront présentés à la prochaine section, encore plus de lecteurs compensateurs de l'échantillon initial de 12 élèves n'ont pu être retenus. Il va sans dire qu'il a ainsi fallu trouver davantage de lecteurs compensateurs que ceux identifiés par l'étude 1, en étendant le recrutement à d'autres écoles des mêmes centres de services scolaires, ayant les mêmes indices de milieu socio-économique.

Pour ce faire, un message a été envoyé aux orthopédagogues de ces écoles, leur demandant si elles pensaient avoir des élèves dans leur école qui sont des lecteurs compensateurs. Tel que soulevé dans l'étude 1, il est possible que certains lecteurs compensateurs ne soient pas connus des milieux puisque

leur mécanisme de compensation camoufle efficacement leurs difficultés de lecture et n'attirent donc pas l'attention du personnel scolaire. Nous avons donc proposé aux orthopédagogues de consulter également les enseignantes… Malgré cela, il demeure fort possible que les lecteurs compensateurs supplémentaires ainsi recrutés représentent des cas un peu extrêmes puisque ceux-ci sont connus des intervenants de l'école. Quoi qu'il en soit, dans le cas où les orthopédagogues pensaient avoir de tels élèves, celles-ci devaient signer un formulaire d'engagement à la confidentialité et au respect de la propriété intellectuelle, puis ils/elles devaient obtenir le consentement éclairé du parent et de l'élève. Ensuite, les orthopédagogues faisaient passer l'épreuve de compréhension écrite (Bianco *et al.*, 2018) et le Test 1 minute de mots et de pseudomots (Sprenger-Charolles et Leloup, 2013) utilisés dans l'étude 1, en suivant le protocole très strict de passation qui leur était fourni. La cotation finale et l'interprétation des réponses étaient réalisées par la chercheuse pour conserver une uniformité. Les élèves qui présentaient une performance dans l'étendue ciblée pour faire partie des lecteurs compensateurs de l'étude étaient invités à participer à l'étude 2.

C'est sous forme d'invitation personnelle que le recrutement auprès des élèves s'est déroulé, à l'aide des outils préparés à cette fin, soit une vidéo de démonstration du déroulement d'une visite au Neurolab, une lettre d'information détaillée sur l'activité ainsi qu'un formulaire de consentement. Ce sont les personnes-contacts dans les écoles (la plupart du temps des orthopédagogues) qui ont présenté le projet aux élèves invités à participer à l'étude 2. Les élèves intéressés à y participer ont ramené les outils de présentation du projet à la maison pour les montrer à leurs parents et obtenir leur consentement éclairé le cas échéant. Sur la trentaine d'élèves invités à participer, trois ont refusé.

4.2.2.2 Caractéristiques des participants

Ce sont 29 élèves qui ont finalement participé à l'étude 2, au NeuroLab de l'Université du Québec à Montréal. Ces 29 élèves proviennent des deux centres de services scolaires qui ont été en mesure de poursuivre leur collaboration l'année suivant l'étude 1. Le tableau et le graphique ci-dessous présentent la répartition des élèves par centre de services scolaires.

Tableau 4.1 Répartition des élèves par centre de services scolaires et par type de lecteurs

Centre de services scolaires (CSS)	Type de lecteurs	Nombre de participants	% de l'échantillon
CSS 2		16 élèves	55%
	Lecteurs compensateurs	8 élèves	
	Contrôles	8 élèves	
CSS 3		13 élèves	45%
	Lecteurs compensateurs	6 élèves	
	Contrôles	7 élèves	
		29 élèves	100%

Pour l'ensemble des participants à l'étude 2, l'âge moyen des élèves de l'échantillon est de 143 mois, ce qui correspond à 11,9 ans, avec un écart-type de 5,5 mois (il est à noter qu'un participant n'a pas répondu à cette question dans le formulaire). Les filles sont au nombre de 17 et les garçons au nombre de 12, ce qui fait que le pourcentage de filles est de 58,62% et est de 41,38% pour les garçons.

Aucun des élèves ne présente un trouble grave de comportement ou de santé mentale attesté, ni de trouble déficitaire de l'attention (TDAH) non médicamenté. Aucun élève ne présente un retard intellectuel ou un trouble grave de la vision. La totalité des élèves de l'échantillon ont le français comme langue maternelle et sont droitiers.

Chez les 15 élèves contrôles, aucun n'est en situation d'échec en lecture-écriture, ni ne l'a été au cours des deux dernières années scolaires (voir le tableau 4.2 pour un résumé de ces caractéristiques). Chez les 14 lecteurs compensateurs, 4 élèves présentent un ou des échec(s) au bulletin en lecture-écriture au cours des deux dernières années, un élève a connu un échec en 1^{e} année seulement puis un autre en 3^{e} année seulement. Seuls deux de ces élèves étaient en situation d'échec en lecture-écriture au moment de leur évaluation au NeuroLab. Parmi les élèves contrôles, un élève a, et a toujours eu, un plan d'intervention dû à une déficience auditive qui nécessite le port d'un appareil. Huit lecteurs compensateurs ont actuellement un plan d'intervention, et 4 d'entre eux avaient également un plan d'intervention dans les années antérieures (il est à noter que l'information est manquante pour trois élèves quant au plan d'intervention

des années précédentes). Parmi les élèves contrôles, 4 élèves ont reçu des interventions spécialisées. Trois élèves ont reçu des interventions orthopédagogiques au préscolaire (conscience phonémique et son des lettres). Parmi ceux-ci un élève a également reçu des interventions orthopédagogiques en première année (conscience phonémique) et un autre en a reçu en 5e année (fluidité et processus non spécifiques de compréhension langagière). Un élève a également reçu des interventions orthophoniques avant son entrée à l'école vu sa déficience auditive. Dix des 14 lecteurs compensateurs ont, quant à eux, reçu des interventions orthopédagogiques. Pour 4 élèves, les premières interventions orthopédagogiques reçues remontent à la 1e ou 2e année. Pour 4 autres, ces interventions ont débuté plus tardivement, soit en 3e, 4e ou 5e année (les informations sur les interventions orthopédagogiques antérieures ne sont pas disponibles pour 2 élèves). Finalement, trois des 14 élèves lecteurs compensateurs n'étaient pas identifiés d'une quelconque façon comme étant un élève en difficulté ou à risque de difficulté de lecture dans son école. Bien entendu, ces trois élèves proviennent de l'échantillon de l'étude 1. Dans le groupe d'élèves contrôles, 2 élèves sont médicamentés pour le trouble déficitaire de l'attention (TDAH) et un élève porte un appareil pour pallier sa déficience auditive. Dans le groupe de lecteurs compensateurs, un élève a été identifié comme ayant un trouble du spectre de l'autisme (TSA) léger, un autre a un trouble de comportement très léger, 3 élèves présentent un trouble spécifique d'apprentissage de la lecture-écriture documenté par un professionnel, puis un élève présente ce même trouble avec un TDAH (médicamenté) concomitant. Il semble pertinent de noter ici que l'identification formelle des troubles par un diagnostic n'est pas obligatoire au Québec pour recevoir de l'aide. Ainsi, il arrive fréquemment que les troubles ne soient pas identifiés chez un élève mais que celui-ci reçoive tout de même de l'aide. C'est pourquoi la présence du diagnostic et la présence d'un plan d'intervention doivent être toutes deux considérées. Il est certain qu'il aurait été souhaitable que les élèves présentant un trouble quelconque soient exclus de l'étude 2. Cependant, vu le petit nombre d'élèves répondant aux critères de lecteurs compensateurs dans l'échantillon initial, il a été choisi de conserver l'échantillon « naturel » de l'étude 1 en priorité, puis de compléter le plus efficacement possible par d'autres lecteurs compensateurs ensuite. Le tableau suivant souligne le fait que deux lecteurs compensateurs étaient en échec en lecture-écriture au moment de la collecte de données. Il semble important de noter que cela ne remet pas en doute leur capacité à compenser, puisque le texte utilisé pour la sélection des participants lecteurs compensateurs était un texte narratif sur un thème bien connu, alors que les textes proposés pour l'évaluation de la lecture en classe de 6e année du primaire incluent généralement des textes informatifs ou directifs dans lesquels il semble beaucoup moins facile de compenser. Par ailleurs, l'échec en écriture a également été considéré. Par conséquent, certains élèves pourraient être en échec en écriture et non pas en lecture. Il a été choisi de

considérer l'échec en écriture également puisqu'il est beaucoup plus difficile de compenser en écriture étant donné l'éventail plus restreint de moyens de compensation possibles (p. ex. ne pas utiliser un mot que l'on ne sait pas écrire). Ce faisant, il serait envisageable que les lecteurs compensateurs ne présentent pas de difficulté manifeste en lecture alors qu'ils en présentent en écriture. Une difficulté ou un échec en écriture seulement pourrait ainsi être un témoin de difficultés en lecture.

Tableau 4.2 Caractéristiques supplémentaires des participants

	Échec en lecture-écriture	**Présence d'un plan d'intervention (PI)**	**Interventions spécialisées reçues**	**Présence d'un trouble attesté par un professionnel**
Lecteurs contrôles (n =15)	Aucun échec	1 élève a un PI en 6e année	6 élèves ont reçu des interventions orthopédagogiques au préscolaire (1 aussi en 1e année, 1 aussi en 5e année) 1 élève a reçu des interventions d'une orthophoniste	2 élèves TDAH (médicamenté) 1 élève avec une légère déficience auditive
Lecteurs compensateurs (n=14)	6 élèves ont eu au moins un échec en lecture-écriture entre la 1e et la 3e année 2 élèves en échec en 6e année	8 élèves ont un PI en 6e année	10 élèves ont reçu des interventions orthopédagogiques (4 en 1e et 2e année, 4 en 3-4-5e année, 2 indéfini)	1 élève TSA 1 élève avec un trouble léger du comportement 4 élèves ayant un TSAL-E (dont 1 avec un TDAH concomitant)

4.2.3 Instruments de mesure

L'instrument de mesure utilisé pour la tâche expérimentale d'amorçage sémantique avec décision lexicale a été entièrement élaboré dans le cadre de cette recherche. Dans un premier temps., la fabrication de cette tâche est décrite, puis les instruments de mesure sélectionnés pour les composantes associées à la lecture-écriture sont présentés.

4.2.3.1 Élaboration des tâches expérimentales

Le choix de recourir à une tâche d'amorçage sémantique permet de manipuler une variable indépendante, qui est le type de lien sémantique entre l'amorce et la cible. Ainsi, chaque tâche comprend au moins 3 conditions, correspondant aux trois types de liens sémantiques pertinents : relié, non relié et neutre (voir le tableau ci-dessous pour le rappel des différentes conditions[27]).

Qui plus est, puisqu'il s'agit d'une tâche de décision lexicale, il faut également assurer la présence de pseudomots écrits, en quantité égale à celle de mots écrits. Ainsi, la composition des trois premières conditions précédentes est répétée en plaçant un pseudomot comme cible plutôt qu'un mot (conditions 4 et 5 dans le tableau ci-dessous). Puisque la cible est un pseudomot, elle ne peut donc pas, par définition, être sémantiquement reliée ou non reliée, c'est pourquoi le parallèle entre les conditions ne peut être complet; la condition 4 regroupe autant d'items que les conditions 1 et 2 cumulées.

Tableau 4.3 Description des différentes conditions et des items de la tâche expérimentale d'amorçage sémantique

Condition		**Description**	**Exemple**
Condition 1	Relié	Mot-amorce et mot-cible reliés sémantiquement	rame - BARQUE
Condition 2	Non relié	Mot-amorce et mot-cible non reliés sémantiquement	athlète - ROSEAU
Condition 3	Neutre	Amorce qui consiste en une série de « X »	XXXXX - CRATÈRE
Condition 4	Pseudomot	Mot-amorce suivi d'un mot-cible qui est un pseudomot	griffe-ÉDRISE
Condition 5	Pseudomot avec amorce neutre	Amorce qui consiste en une série de « X » suivie d'un mot-cible qui est un pseudomot	XXXXX - COUVENNE

[27] Les amorces neutres peuvent parfois être constituées de pseudomots. Puisque la population de la présente étude représente des élèves pour qui la lecture de pseudomots est déficitaire, les X ont été choisis comme amorce.

Pour chaque SOA (court et long), une liste de paires d'items doit donc être élaborée. Dans un premier temps, pour chacune des listes, c'est la condition 1 qui a été élaborée d'abord, étant donné qu'il s'agit de la condition la plus contraignante : les amorces et les cibles doivent être sémantiquement reliées, soit par une forte association, soit par un important partage de traits sémantiques (sans être phonologiquement ou orthographiquement similaires). Il a été choisi de sélectionner autant de mots qui ont une forte association que de mots dont les concepts partagent un grand nombre de traits sémantiques.

Des tables d'association verbale du français ont été consultées. Faute de mieux, ce sont les « normes d'association verbales pour 366 noms d'objets concrets » de Ferrand et Alario (1998) qui ont été retenues, même si les associations ont été mesurées chez l'adulte. Ces normes ont été complétées par celles de De La Haye (2003), élaborées avec des élèves de 9 à 11 ans. Au contraire des premières normes, celles-ci incluent toutes les catégories grammaticales et portent sur seulement 200 mots. Toutes les paires de mots de ces tables ont été considérées si leur force d'association était de 15% et plus. Dans le cas qui nous occupe, le premier mot correspond à l'amorce et le deuxième mot à la cible. Puisque la cible ne doit pas être reconnue instantanément, il convient de sélectionner des paires dont le 2^{e} mot est le moins fréquent possible (tant à l'oral qu'à l'écrit). La base de données MANULEX-infra (Peereman *et al.*, 2007) a donc été consultée et toutes les paires dont la cible a une fréquence écrite U de 100 ou plus ont été retirées et les fréquences les plus basses ont été préférées. Toutes les cibles ont ainsi une fréquence écrite plus petite que 50, mises à part quelques exceptions qui ont dû être conservées tout de même et qui ont été balancées le plus possible à travers les conditions. Afin de conserver le plus d'homogénéité possible dans les items présentés, il a été choisi de ne conserver que les paires de mots écrits formés par deux noms communs présentés au singulier, qui ne contiennent pas d'affixe et dont le mot-cible comporte entre 6 et 12 lettres. Ces noms communs représentent des réalités concrètes. Ces critères ont permis de relever 34 paires de mots fortement associés. La moitié de ces paires (17) sert à la première liste d'items (pour le SOA court) et la seconde moitié, à la deuxième liste d'items (pour le SOA long). Aucune de ces paires ne présente une possibilité d'amorçage phonologique ou orthographique.

Pour compléter les items de la condition 1, 34 paires de mots écrits dont les concepts partagent un nombre important de traits sémantiques devaient être élaborées. Pour ce faire, vu l'absence de telles informations en français, il a été choisi de recourir à la table de similarité des concepts de McRae (2005). Cette table présente le degré de similarité (nombre de traits sémantiques partagés) entre 541 concepts, exprimé sous la forme d'un nombre entre 0 et 1; 1 représentant une similarité complète. Les paires constituées de ces

deux concepts ont été conservées si la similarité de leur concept était de 0,3 et plus. Puisque la table était en anglais, les concepts ont dû être traduits. La traduction de certains noms de l'anglais mène parfois en français à une expression (« shotgun » – fusil à pompe), alors ces items ont été retirés. Ensuite, les mêmes critères que ceux utilisés pour les paires de mots fortement associés ont été appliqués, c'est-à-dire que le mot-cible doit comporter entre 6 et 12 lettres, avoir une fréquence écrite U de 50 ou moins (à moins d'exception), et que les amorces et les cibles soient des noms communs sans affixe. La concrétude des concepts a également été vérifiée à l'aide des « Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mot » de Bonin et coll. (2003), et tous les noms communs non concrets ont été retirés. Du nombre de paires restantes (environ 41), quelques paires ont été retirées parce qu'elles présentaient une possibilité d'amorçage phonologique ou orthographique. Trente-quatre (34) paires ont ensuite été sélectionnées parmi les paires restantes parce qu'elles avaient des amorces qui présentaient des caractéristiques linguistiques similaires aux amorces des items fortement associés (notamment en ce qui concerne la fréquence écrite et la fréquence orale chez des enfants québécois).

Au total, ce sont donc 68 paires de mots écrits qui ont été sélectionnées pour la condition 1 (mots sémantiquement reliés), 34 par liste. Ceci fait en sorte que les autres conditions doivent également comprendre 34 paires d'items (voir la figure ci-dessous pour une représentation visuelle de la répartition des items par condition pour une liste).

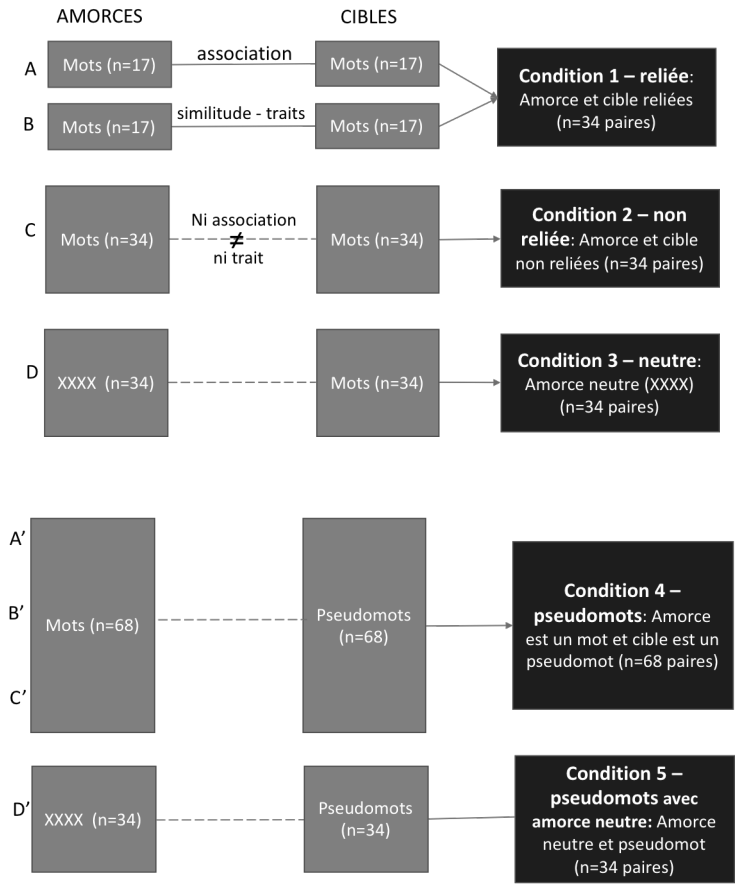


Figure 4.3 Constitution d'une liste d'items pour la tâche expérimentale

Concernant l'élaboration des listes d'amorces et de cibles non reliées, ce sont 102 amorces qui ont été sélectionnées pour les conditions 2 et 4 de chaque liste. Ces amorces doivent être appariées aux amorces de la condition 1. Pour ce faire, la fréquence écrite U > 60 et la classe grammaticale *nom singulier* ont été entrées dans Manulex-infra (Peereman *et al.,* 2007), puis le nombre de lettres et de syllabes orales correspondant au mot à apparier ont également été considérés pour produire une liste de mots potentiels. Les mots ainsi obtenus ont ensuite été considérés seulement s'ils représentaient des concepts concrets et s'ils ne comportaient pas d'affixes. Les trois amorces qui représentent les meilleurs appariements en termes de nombre de graphèmes, de structure syllabique, de présence ou non de graphème contextuel ou de semi-voyelle ont été sélectionnés. Si plusieurs candidats sont toujours disponibles, l'amorce représentant une entité bien connue des élèves a été choisie.

Concernant les cibles, ce sont les 34 mots-cibles de la condition 1 qui ont servi de base pour élaborer les 68 mots-cibles des conditions 2 et 3. En ce qui concerne les mots-cibles, de la même manière que pour les amorces, certains critères de ces mots à apparier ont été entrés dans la base de données Manulex-infra (Peereman *et al.*, 2007) pour trouver des mots potentiellement utilisés: fréquence écrite U < 50, nom singulier et maximum de 12 lettres. Le nombre de lettres et le nombre de syllabes correspondant au mot écrit à apparier ont également été entrés pour restreindre les mots proposés. Les mots ainsi obtenus ont ensuite été considérés seulement s'ils représentaient des concepts concrets et s'ils ne présentaient pas d'affixe. Pour chaque mot-cible de la condition 1, ce sont deux mots-cibles qu'il a fallu choisir pour les conditions 2 et 3. Les deux cibles qui ont une fréquence écrite la moins élevée possible et représentant les meilleurs appariements en termes de nombre de graphèmes, de structure syllabique, de présence ou non de graphème contextuel ou de semi-voyelle ont été sélectionnés.

Tous les pseudomots ont été créés à partir des mots des conditions qui leur sont associés. Ainsi, les pseudomots de la condition 4 proviennent des mots-cibles des conditions 1 et 2, et les pseudomots de la condition 6 proviennent des mots-cibles de la condition 3. Les mots-cibles concernés ont été entrés dans le logiciel de création de pseudomots Wuggy (Keuleers et Brysbaert, 2010). Le logiciel a ainsi proposé une dizaine de pseudomots correspondant aux règles graphotactiques du français (le module sélectionné est le module « Français orthographique »), et partageant avec le mot à apparier le nombre de lettres, la longueur des syllabes[28] et au moins un segment identique sur 3 (p. ex. « jambon » devient « **j**aup**on** »). Les pseudomots comportant une finale qui pourrait être interprétée comme une marque du pluriel ont été retirés (p. ex. « campiés »), ainsi que ceux qui comportaient des éléments plus ou moins réalistes en français (tout ce qui concerne les accents est difficilement traité par Wuggy; p. ex. « artuere ») ou trop proche de vrais mots (p. ex. « mainseche »).

Une fois les mots-amorces, les mots-cibles et les pseudomots sélectionnés, les paires amorce-cible ont pu être construites en réunissant les amorces et les cibles élaborées pour chaque condition. Dans le cas où une paire présentait un lien sémantique quelconque (même ténu) dans les conditions 2 et 3, la paire a été déconstruite pour être reconstruite avec un autre item de la même condition, de la même liste (p.ex. la paire vache-ASSIETTE). Il en a été de même pour les possibilités d'amorçage phonologique ou orthographique (p. ex. bulle-BUFFLE). Puisque ce sont les moyennes des catégories d'items qui sont

[28] L'unité sélectionnée par le logiciel ici est la syllabe écrite étant donné que le module sélectionné est le module orthographique.

considérées, les items sont interchangeables dans ces conditions, au contraire de la condition 1 où les items doivent être sémantiquement reliés.

Plusieurs caractéristiques linguistiques des mots (pour les mots seulement) ont ensuite été considérées pour balancer les items entre les conditions. Les moyennes de ces caractéristiques ont été calculées pour chaque condition, puis certaines paires ont été changées de liste ou de condition pour que celles-ci soient les plus équivalentes possible. Ces caractéristiques linguistiques sont les suivantes : fréquence écrite du nom commun et consistance des correspondances graphèmes-phonèmes du mot (Manulex-infra, Peereman *et al.*, 2007); fréquence en production orale chez des enfants québécois (Préfontaine, 1998); fréquence du mot oral dans des films, nombre de voisins orthographiques et nombre de voisins phonologiques (Lexique, New *et al.,* 2001); nombre de sens associés au mot (Trésor de la langue française informatisé (TLFi)), nombre de phonèmes et nombre de graphèmes.

Il s'est avéré impossible de trouver une manière d'organiser les items pour assurer une équivalence parfaite des listes, mais un travail intensif a été réalisé pour s'approcher le plus possible de cette équivalence. Quoi qu'il en soit, le tableau suivant présente toutes ces moyennes par condition.

Tableau 4.4 Caractéristiques linguistiques des items par condition

	Fréquence écrite (U)	Fréquence orale (enfants)	Fréquence orale films	Consistance des CGP	Nombre de sens	Nombre de phonèmes	Nombre de graphèmes	Nombre de voisin ortho.	Nombre de voisin phono.	Nombre de lettres	Nombre de syllabes
Liste 1											
Amorces AB – Condition 1	81,72	22,66	49,29	80,75	1,65	4,00	4,53	6,29	12,06	5,41	1,44
Amorces C – Condition 2	90,63	21,93	34,83	83,38	1,56	3,82	4,47	5,26	14,12	5,35	1,44
Amorces AB' – Condition 4	63,00	18,52	36,90	81,34	1,55	4,06	4,50	3,88	10,12	5,44	1,44
Amorces C' – Condition 4	79,23	15,23	43,65	83,27	1,33	3,97	4,50	4,32	11,29	5,38	1,59
Cibles AB – Condition 1	18,89	27,30	13,28	78,43	1,09	5,70	6,24	1,18	3,24	7,38	2,18
Cibles C – Condition 2	18,49	31,92	9,38	77,66	1,15	5,39	6,00	1,41	3,85	7,32	2,12
Cibles D – Condition 3	16,21	14,55	12,93	81,19	1,18	5,47	6,03	1,38	3,47	7,32	2,26
Liste 2											
Amorces AB – Condition 1	73,56	24,61	30,71	83,26	1,38	3,97	4,32	4,12	10,32	5,50	1,62
Amorces C – Condition 2	53,64	16,46	48,59	83,68	1,53	4,03	4,44	3,88	11,44	5,56	1,62
Amorces AB' – Condition 4	65,77	18,22	30,56	81,17	1,47	4,00	4,50	4,35	11,18	5,59	1,62
Amorces C' – Condition 4	63,88	26,40	39,85	84,07	1,44	4,09	4,47	4,50	13,32	5,53	1,56
Cibles AB – Condition 1	24,83	18,62	11,72	81,67	1,21	5,13	5,81	1,15	4,41	7,29	2,12
Cibles C – Condition 2	23,90	24,91	15,77	81,73	1,18	5,21	5,85	1,71	3,32	7,29	2,09
Cibles D – Condition 3	19,52	17,97	12,05	83,76	1,18	5,18	5,94	1,44	5,03	7,32	2,09

La séquence de présentation des éléments pour chacune des listes (SOA court ou long) est présentée à la figure 4.4. Pour chaque paire d'items, c'est d'abord un point de fixation qui est présenté à l'écran (une séquence de deux « + » qui sert à fixer l'attention du participant au centre de l'écran, c'est-à-dire l'endroit où l'item à lire va apparaître). Ensuite, l'amorce apparaît à cet endroit pendant un temps donné (plus ou moins long selon le SOA), puis est suivie d'une série de dièses qui servent de masque. Ce masque « efface » en quelque sorte la trace visuelle, ou effet résiduel, qui pourrait être laissée par l'amorce en mémoire de travail. Ce masque empêche donc le sujet de poursuivre le traitement de l'item alors qu'il n'est plus sous ses yeux. La cible est ensuite présentée et le participant doit alors fournir sa réponse en appuyant sur un bouton du clavier de l'ordinateur.

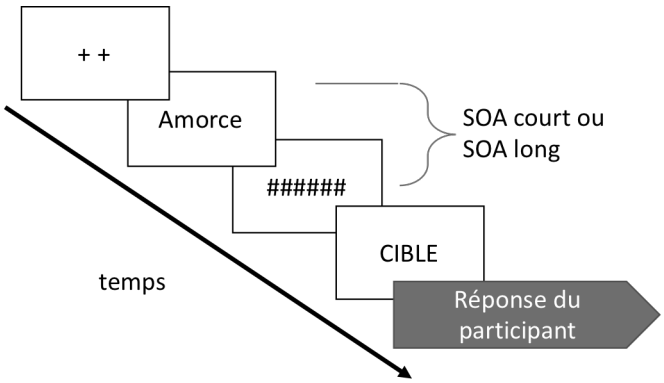


Figure 4.4 Structure générale de la séquence des stimuli visuels dans la tâche d'amorçage sémantique

Deux commentaires valent la peine d'être formulés à cette étape-ci. D'abord, afin que les participants fournissent leur réponse lors de la tâche, l'option la plus répandue dans les études publiées est celle qui consiste à associer un bouton à la réponse « oui, il s'agit d'un vrai mot, un mot de la langue française » puis un autre bouton à la réponse « non, ce n'est pas un mot de la langue française, c'est un mot inventé ». Il existe également une alternative à cette façon de faire, il s'agit du paradigme Go/NoGo. Dans ce paradigme, les participants sont invités à appuyer sur un bouton seulement si l'item appartient à une catégorie donnée. Dans le cas qui nous occupe, on pourrait par exemple donner l'instruction suivante : « Tu appuies sur le bouton seulement si l'item présenté à l'écran est un vrai mot, un mot existant en français; si ce qui apparaît à l'écran n'est pas un vrai mot, alors tu ne fais rien, tu te retiens d'appuyer. » Cette manière de fournir leur réponse est possiblement plus simple à comprendre pour les enfants d'âge scolaire primaire puisque la consigne implique seulement un bouton. De plus, cette manière de faire permet également de réduire l'influence que peut avoir l'opération cognitive liée à la décision entre un type d'item et un autre, puisqu'il s'agit d'appuyer sur un bouton lorsque l'on « reconnait » le mot comme faisant partie de la langue écrite (plutôt que de devoir décider à laquelle des deux catégories l'item présenté fait partie). Il est également documenté que ces paradigmes de réponse entraînent des résultats similaires (Lee *et al.*, 2019; Moret Tatay et Perea, 2011; Perea et Gomez, 2002;). Il est ainsi anticipé que la tâche puisse être plus accessible pour des élèves en difficulté de la fin du primaire, et ce, sans compromettre la qualité des résultats. Le paradigme Go\NoGo est donc retenu.

Afin de tester les listes d'items, le paradigme de réponse, ainsi que les temps d'exposition de chaque élément, tel que présenté à la figure 4.5, il a été choisi de réaliser une mise à l'essai des deux tâches. Les

tâches ont été programmées à l'aide du logiciel gratuit de présentation des stimuli DMDX (Forster et Forster, 2003). Dans un premier temps, 7 élèves de 6e année (4 présentant des difficultés similaires à celles des lecteurs compensateurs; 3 élèves sans difficulté) ont participé à une séance individuelle d'essai des tâches sur un ordinateur portable. Pendant ces séances, les élèves font une partie de la tâche puis s'arrêtent. Des ajustements ont lieu avant de reprendre la mise à l'essai et des questions sont également posées pour avoir les commentaires des élèves ou obtenir des renseignements précis (p. ex. concernant la visibilité de l'amorce). Ensuite, 35 élèves provenant de deux classes d'un Centre de services scolaire participant ont été interpellés pour la suite de la mise à l'essai. Dix ordinateurs portables ont été utilisés pour recueillir les données de 10 élèves à la fois. Des entretiens individuels très brefs ont été réalisés avec ces élèves pour recueillir leurs commentaires et pour répondre à deux courtes questions concernant le niveau de difficulté de la tâche et la visibilité des amorces.

La version finale de la structure générale de la tâche est présentée dans la figure ci-dessous.

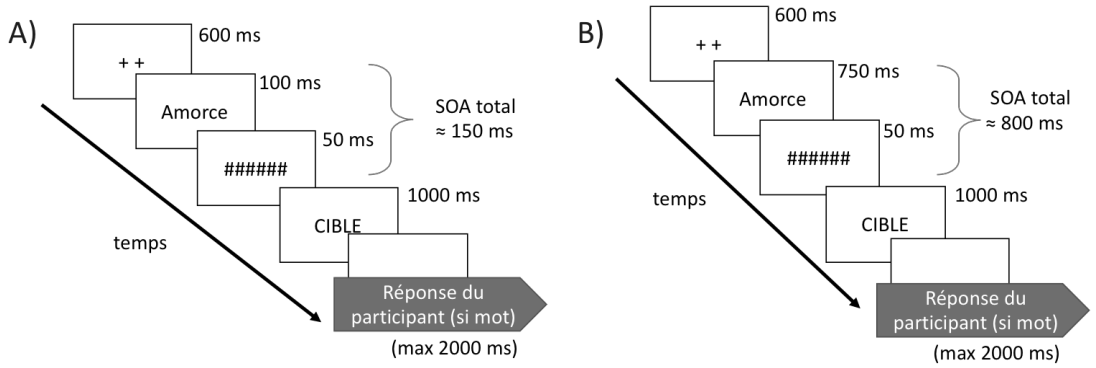


Figure 4.5 Structure de la tâche d'amorçage sémantique. SOA court à gauche; SOA long à droite.

Les points de fixation apparaissent pendant 600 ms, puis l'amorce pendant 100 ms pour la tâche au délai court ou 750 ms pour la tâche au délai long. Le masque apparaît ensuite pour 50 ms. Le délai (SOA) court est donc de 150 ms, le délai (SOA) long est donc de 800 ms. Le temps accordé au délai (SOA) court est en réalité un compromis. En effet, il aurait été souhaitable que le SOA court soit si rapide qu'il soit sous le seuil de la conscience pour l'ensemble des élèves. Cependant, la mise à l'essai montre qu'il faut un minimum de temps aux élèves en difficulté pour que l'amorce soit traitée; c'est pourquoi le SOA a été augmenté à 150 ms. À cette exposition, les élèves sont conscients que quelque chose apparaît sur l'écran, mais cela laisse une chance aux élèves plus faibles en identification de mots écrits de traiter l'amorce. La cible s'affiche ensuite pendant 1000 ms, puis il apparaît un écran blanc pendant 1000 ms supplémentaires.

Les participants ont donc un maximum de 2 secondes pour appuyer sur le bouton désigné si l'item est un mot. Après ces 2 secondes, la prochaine paire est présentée, débutant par les points de fixation.

4.2.3.2 Sélection des instruments de mesure pour les composantes associées à la lecture

Un ensemble d'autres variables font également l'objet de mesures. Ces mesures concernent des habiletés ou des connaissances qui sont associées à la lecture, et sont prévues être utilisées comme variables contrôles dans les statistiques. Dans un premier temps, comme la sémantique lexicale est le concept central de l'hypothèse, il est primordial d'aller observer le niveau de vocabulaire des participants, ainsi que leur fluence verbale. De plus, puisque les tâches expérimentales consistent en la lecture de mots et de pseudomots, il convient également d'aller mesurer la conscience phonologique, la dénomination rapide, ainsi que la mémoire de travail phonologique, qui sont toutes hautement associées à ces capacités de lecture.

Parmi les épreuves disponibles pour mesurer ces composantes, il a été choisi de ne conserver que celles qui étaient des instruments de très haute qualité, standardisés et normalisés avec une population d'enfants. Le choix final des instruments est également réalisé afin que le temps de passation demeure réaliste (préférence d'épreuves à règle d'entrée et règles d'arrêt).

Pour le vocabulaire, c'est l'outil EVIP (Dunn *et al.*, 1993), qui est une adaptation en langue française du Peabody Picture Vocabulary Test – Revised, qui a été sélectionné. Cet outil est une mesure d'étendue du vocabulaire réceptif. Les participants sont exposés à quatre images présentées simultanément sur une page. Ils doivent ensuite choisir, parmi ces images, laquelle illustre le mieux le mot prononcé à haute voix par l'expérimentatrice. Des normes sont disponibles pour des enfants de l'âge de ceux des participants, en considérant l'âge chronologique de ces derniers par tranche de deux mois, chez des enfants du Canada francophone. La passation individuelle dure entre 8 et 15 minutes et il existe une règle de point de départ ainsi qu'une règle d'arrêt.

L'outil qui a été sélectionné pour évaluer la fluence verbale est la batterie d'épreuves « langage oral, langage écrit, mémoire, attention » - 2^{e} édition (L2MA-2; Chevrier-Muller *et al.*, 2010). Cet outil mesure la fluidité de la production lexicale des participants, c'est-à-dire l'aisance avec laquelle ces derniers peuvent retrouver et évoquer des mots selon une consigne donnée, dans un temps donné. Dans le cas qui nous occupe, deux tâches sont proposées aux participants : la première concerne les métiers (nommer le plus

de noms de métier en une minute), la deuxième concerne les moyens de transport (nommer le plus de moyens de transport en une minute). Les normes concernent les enfants français de 7 à 11 ans. La passation individuelle dure entre 3 et 4 minutes.

Puisque la plupart des élèves de la fin du primaire ont atteint un excellent niveau d'expertise en conscience phonologique, il n'est pas rare que les épreuves qui la mesurent présentent un effet plafond. C'est effectivement le cas de l'épreuve qui a été choisie ici. Il s'agit de l'« évaluation clinique des notions langagières fondamentales – Version pour francophones du Canada » (CELF; Semel *et al.*, 2009). Tel que mentionné dans le cadre théorique, la phonologie joue un rôle primordial dans l'apprentissage de l'identification des mots écrits . Elle joue également un rôle important dans l'explication des difficultés de lecture. Plus précisément, le décodage de mots ou de pseudomots implique la capacité à fusionner des phonèmes, qui est une catégorie d'habiletés de conscience phonémique. Il semble ainsi primordial de mesurer la conscience phonologique ici, et particulièrement la conscience phonémique. L'épreuve sélectionnée est donc une épreuve qui inclut 8 items faisant appel à chacune de ces opérations mentales, dans l'ordre : élision syllabique, discrimination de rimes, identification du phonème initial, identification du phonème médian, élision du phonème initial, substitution de phonèmes, fusion de phonèmes et segmentation de phonèmes. Les normes sont présentées pour des élèves francophones du Canada de 9 à 16 ans. La passation individuelle dure entre 12 et 20 minutes et il existe une règle d'entrée selon l'âge chronologique ainsi qu'une règle d'arrêt.

Une autre composante qui joue un rôle essentiel dans l'identification des mots écrits, ainsi qu'un rôle potentiel dans les difficultés de lecture, notamment pour le trouble spécifique d'apprentissage de la lecture-écriture, est la dénomination rapide (p. ex. Wolf, 2001). Selon Wolf (2001), la dénomination rapide expliquerait la composante « rapidité » dans l'identification de mots écrits. L'épreuve classique pour des élèves de la fin du primaire consiste en une épreuve de dénomination rapide de lettres. C'est donc cette épreuve qui a d'abord été choisie. En plus de celle-ci, une épreuve de dénomination rapide de chiffres a également été ajoutée, au cas où les lecteurs compensateurs seraient pénalisés par leur moins grande exposition aux lettres que leurs pairs de même âge chronologique. La dénomination rapide de lettres est celle de la batterie « Batterie informatisée d'examen du Langage Oral, Langage écrit, Mémoire, Attention, Compétences transversales (Exalang 8-11; Thibault *et al.*, 2012). Dans cette épreuve, les élèves doivent dire à haute voix le nom des lettres majuscules (A, H, L, O et Z) présentées de manière randomisée sur une feuille, dans l'ordre de présentation, le plus rapidement possible. Pour la dénomination rapide de chiffres,

c'est l'épreuve de la « Batterie analytique du langage écrit » (BALE; Peereman *et al.*, 2007) qui a été sélectionnée. Cette épreuve contient 10 séries de 5 chiffres à lire à haute voix, dans l'ordre de présentation, le plus rapidement possible. Dans ces deux épreuves, tant le score de réussite que le temps doivent être notés. La passation individuelle de ces deux épreuves prend environ 5 minutes et les normes ont été élaborées avec des enfants français.

Le dernier instrument de mesure évalue la mémoire de travail phonologique, fortement sollicitée lors de la lecture, notamment pour les pseudomots (traitement alphabétique). L'épreuve sélectionnée fait partie de l' « Échelle d'intelligence de Wechsler pour enfants – 5e édition » (WISC-V; Wechsler, 2015) et est donc bien documentée et robuste, normée sur des francophones du Canada. L'épreuve s'appelle « Séquences de chiffres » et comporte trois parties. La première consiste en un rappel des chiffres mentionnés par l'évaluateur dans l'ordre direct, c'est-à-dire à l'endroit. La deuxième partie consiste en un rappel des chiffres dans l'ordre inverse, puis la dernière partie consiste en un rappel des chiffres dans l'ordre croissant (peu importe l'ordre dans lequel l'évaluateur les a prononcés). Un maximum de 18 séquences de chiffres à rappeler sont présentées dans chaque partie de l'épreuve. La passation individuelle de cette épreuve prend entre 3 et 5 minutes, et comporte une règle d'arrêt. Des normes sont disponibles pour des enfants de l'âge de ceux des participants, en considérant l'âge chronologique de ces derniers par tranche de quatre mois.

4.2.4 Déroulement

Les participants se sont déplacés à l'Université du Québec à Montréal (UQAM), dans les locaux de NeuroLab, accompagnés de leur parent ou d'un intervenant scolaire. Pour la plupart d'entre eux, c'est un intervenant scolaire qui les a accompagnés, en groupes de 2 ou 3 élèves à la fois. En échange de la libération des intervenants lors de cette journée, des journées de formation sur différents sujets touchant lecture-écriture ont été offertes aux centres de services scolaires. Les participants étaient accueillis à leur arrivée par la chercheuse principale ou un des 6 assistants de recherche[29] impliqués dans la collecte de données. À leur arrivée, la visite des lieux débute leur journée : 1) local d'attente et vestiaire; 2) local de passation des épreuves portant sur les composantes associées à la lecture-écriture; 3) local de collecte de données EEG. À ce dernier lieu, les participants sont donc introduits aux équipements qui seront utilisés et familiarisés avec le fonctionnement des électrodes, ceci afin de minimiser le sentiment d'insécurité qui

[29] Les assistants de recherche ont tous préalablement signé le formulaire d'engagement à la confidentialité et au respect de la propriété intellectuelle tel qu'il est de mise pour tous les assistants travaillant à ce projet.

pourrait être induit par ce type de collecte. Simultanément, les assistants de recherche et la chercheuse répondent aux questions sur l'EEG, puis le déroulement de leur présence à l'UQAM est expliqué, et on réitère évidemment le caractère volontaire de la participation à cette recherche. Dès le début de la journée, les mesures de têtes sont prises, ainsi que l'essai des bonnets d'EEG, question de pouvoir préparer d'avance le matériel d'EEG et de diminuer le temps d'attente pour les participants. Pour la majorité des participants, on débute par les tâches expérimentales dans le NeuroLab (collecte de données EEG), on accorde une pause d'une trentaine de minutes minimum (parfois toute l'heure du dîner), puis on invite ensuite le participant dans l'autre local pour passer les épreuves portant sur les composantes associées à la lecture-écriture. Dans certaines circonstances, surtout pour répondre aux contraintes de temps des intervenants scolaires, l'ordre a été inversé entre les tâches expérimentales et les épreuves portant sur les composantes pour quelques élèves. Dans ces cas, une pause d'au moins 1 heure a été assurée avant la réalisation des tâches expérimentales, pour éviter que les épreuves sur les composantes « préactivent » des connaissances ou stratégies utiles aux tâches expérimentales et en biaisent les données.

Pour la partie de la collecte de données qui se déroule dans le NeuroLab, le participant est invité à s'asseoir sur une chaise devant un écran d'ordinateur. Le bonnet qui contient déjà les 64 électrodes actives est installé sur sa tête et ajusté (des mesures sont prises pour assurer qu'il est bien installé et le confort est également considéré). Afin de créer le contact nécessaire entre le bout de chaque électrode et le cuir chevelu juste en-dessous des cheveux, du gel doit être inséré sous chacune des 64 électrodes. Ceci est expliqué au participant et deux assistants se mettent à la tâche pendant que le participant écoute de la musique, joue au IPad, lit un livre ou discute avec les assistants de recherche. Cette étape, qui est la plus longue de toute l'installation, est d'une durée d'environ 15 à 20 minutes. Une fois l'appareillage EEG installé et fonctionnel, la chaise du participant est avancée et sa hauteur réglée pour que le participant ait les pieds au sol et qu'il soit le plus confortable possible. À ce moment, un des assistants de recherche démarre l'enregistrement pendant que l'autre demande au participant de faire des mouvements précis (p. ex. cligner des yeux, serrer les mâchoires, etc.) Les consignes standards générales pour l'EEG sont ensuite fournies au participant (p. ex. bouger le moins possible, tenter de cligner des yeux seulement après la disparition du mot, etc.) Dans le cas où un élève semble avoir de la difficulté à ne pas bouger, l'assistant lui offre un « lézard lourd », outil sensoriel composé de petites billes lourdes qui donnent un effet apaisant et qui contribue souvent à diminuer l'agitation motrice. La première tâche expérimentale est expliquée au participant à l'aide des consignes rédigées préalablement, puis l'assistant demeure à côté de lui pendant qu'il fait les 20 items d'essai. L'assistant s'assure que le participant est confortable et qu'il comprend bien

ce qu'il faut faire. Au besoin, les items d'essai sont répétés. L'assistant se déplace ensuite derrière une vitre sans tain et le participant débute ensuite la première tâche. Tel que mentionné à la section précédente, le logiciel de présentation des stimuli est DMDX (version 6.1.1.1; Forster et Forster, 2003). C'est ce logiciel, installé sur l'ordinateur de la salle de collecte, qui permet de récolter la réponse du participant, ainsi que le temps pris pour ce faire. DMDX envoie des signaux qui marquent le signal EEG qui lui est enregistré sur un ordinateur distinct (à chaque début de stimulus et à chaque réponse du participant). Pour la première tâche, les éléments visuels (mots, pseudomots, points de fixation, masques) sont présentés dans la police de caractère Arial au centre de l'écran. Puisque certains élèves ont des difficultés majeures en identification de mots écrits, il a semblé judicieux de ne pas utiliser la méthode standard de présentation des stimuli chez l'adulte, soit l'amorce en minuscule et la cible en majuscule. En revanche, afin d'assurer la clarté, la cible est présentée aux participants dans une couleur distincte du reste, c'est-à-dire en bleu foncé. Lors de la passation, les items sont présentés en ordre aléatoire et sont présentés les uns après les autres de manière continue. Des pauses sont insérées après chaque groupe de 20 items. Lors de ces pauses, le participant peut en profiter pour bouger un peu, repositionner ses mains, se reposer, etc. C'est à ce moment que les assistants entrent dans la pièce s'ils ont besoin de réajuster quoi que ce soit. Quand il est à nouveau prêt, le participant appuie sur une touche et la tâche continue. Lorsque le participant a terminé la première tâche, les instructions lui sont fournies pour la deuxième tâche, et le déroulement est exactement le même, ainsi que la structure des stimuli. Il y a encore la présentation de 10 items d'essai, qui est réalisée sous la supervision de l'assistant de recherche. La passation de ces deux tâches, l'une après l'autre, prend environ 40 à 50 minutes. Les données d'EEG sont enregistrées sur les 64 électrodes actives du bonnet, placées selon le système « 10-20 ». Pendant l'enregistrement, toutes les électrodes seront référencées à l'électrode Fz et amplifiées par un ampli *BrainAmp* avec un filtre passe-bande de 0,01 à 100 Hz. Le signal EEG est échantillonné de manière continue à 1000 Hz. Les signaux reçus du logiciel DMDX sont marqués en temps réel dans l'enregistrement de l'EEG pour faciliter le repérage ultérieur des stimuli et des réponses du participant. L'impédance des électrodes est maintenue sous 20 kΩ. C'est le logiciel Brain Vision (version 2.2) qui est utilisé pour l'acquisition de données EEG.

4.2.5 Préparation des données d'électroencéphalographie

Avant de procéder à l'analyse des données de potentiels évoqués, certaines étapes de traitement préalables sont nécessaires. Les données d'EEG sont ainsi filtrées pour ne conserver que les fréquences entre 0,2 et 30 Hz (filtre coupe-bande de 60 Hz). Les canaux correspondant aux électrodes qui n'ont pas

enregistré correctement sont retirés (inspection visuelle) ainsi que ceux dont le signal varie de plus de 50 µV entre deux points de données successifs ou de 140 µV dans un intervalle de 100 ms. Les données sont ensuite re-référencées à la moyenne du signal de toutes les électrodes. Les données brutes continues de l'EEG sont segmentées en épisodes d'une durée de 900 ms, débutant 100 ms avant la présentation d'un item à identifier jusqu'à 800 ms après la présentation de ce dernier. Un algorithme qui permet de corriger l'activité électrique mesurée par les électrodes lors des clignements des yeux est ensuite appliqué de manière semi-automatique sur chaque segment, sur chaque électrode. Finalement, les 100 ms précédant la présentation du stimulus sont utilisées pour procéder à une correction du signal à l'aide du niveau de base. Les segments associés aux derniers artéfacts sont ensuite retirés manuellement après une inspection visuelle. Pour être conservées, les données des participants devaient avoir au moins 19 segments valides par type d'items, ce qui a mené au retrait de toutes les données recueillies auprès d'un participant. Les données d'un autre participant ont également dû être retirées de certaines analyses à cause de la mauvaise qualité du signal des électrodes dans la région frontale (certaines électrodes F, AF et Fp).

Les fenêtres temporelles ainsi que les localisations à utiliser pour identifier les composantes N400 et P200 dans les deux tâches ont été déterminées en faisant une inspection visuelle des données et en se basant sur la littérature scientifique à propos de participants similaires. Ainsi, dans ce qui suit, la P200 correspond au point le plus positif du signal entre 120 et 220 ms pour la tâche au SOA court et entre 180 et 250 ms pour la tâche au SOA long; la N400 correspond au point le plus négatif entre 280 et 380 ms pour la tâche au SOA court et entre 280 à 500 ms pour la tâche au SOA long. Pour la tâche au SOA court, tant la P200 que la N400 ont été mesurés dans la région frontale (voir la figure 4.6), avec deux regroupements d'électrodes selon la latéralité de l'activité électrique : AF7, F5, F7, AF8, F6, F8, FT7, FC5, FC6 et FT8. Pour la tâche au SOA long, les électrodes ont été regroupées en 4 groupes distincts (voir la figure 4.6) : F5, F7, FC5, FT7 et AF7; F6, F8, FC6, FT8 et AF8; C1, C3, CP1 et CP3; C2, C4, CP2 et CP4.

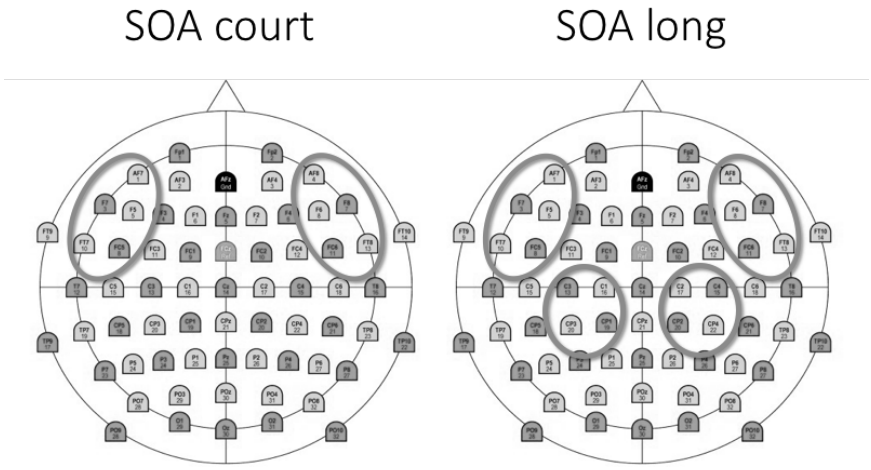


Figure 4.6 Regroupements d'électrodes pour les données d'ERP d'amorçage sémantique. Délai (SOA) court à gauche, délai (SOA) long à droite

4.3 Résultats de l'étude 2

Les résultats de la tâche expérimentale d'amorçage sémantique ainsi que les résultats des épreuves sur les composantes associées à la lecture sont rapportés ici. Les données de cette partie de l'étude peuvent être considérées comme des résultats exploratoires vu le petit nombre de participants. En effet, pour certains tests statistiques, la puissance n'est pas celle que l'on aurait souhaité. Ceci fait ainsi partie intégrante des limites de l'étude qui seront discutées plus en détails en conclusion. Par ailleurs, alors que les composantes associées à la lecture-écriture étaient supposées être considérées comme des variables contrôles dans la tâche d'amorçage sémantique (c'est-à-dire être entrées comme covariables dans les statistiques), le manque de puissance engendré par le faible échantillon a empêché de recourir à cette méthode. Les résultats obtenus sur ces composantes sont donc présentés à part, et permettent tout de même de documenter le phénomène de la compensation en lecture.

4.3.1 Résultats des composantes associées à la lecture-écriture

Chacune des mesures des composantes associées à la lecture-écriture a fait l'objet d'une ANOVA ayant le type de lecteur (lecteur compensateur ou lecteur contrôle) comme facteur intersujet. Ces analyses ont été conduites avec les données de 29 participants, 14 lecteurs compensateurs et 15 lecteurs contrôles. Les moyennes et les écart-types sont rapportés ci-dessous.

Tableau 4.5 Moyennes et écarts-types des mesures portant sur les composantes associées à la lecture pour chaque type de lecteurs

	Lecteurs compensateurs	**Contrôles**
	Moyenne (écart-type)	
Mémoire de travail phonologique (WISC)	20,64 (4,67)	21,47 (2,50)
Conscience phonologique (CELF) ★	59,36 (8,65)	69,47 (12,40)
Dénomination rapide de chiffres (BALE) ★	27,51 (4,54)	21,79 (3,20)
Dénomination rapide de lettres (Exalang 8-11)	12,16 (1,82)	11,31 (1,95)
Vocabulaire réceptif (EVIP)	129,93 (10,03)	133,47 (13,45)
Fluence (L2Ma)	20,93 (4,34)	23,2 (5,75)

Note. Les étoiles désignent les mesures qui sont statistiquement significatives ($p < 0,05$)

Le tableau 4.5 met en évidence le fait que, au premier coup d'œil, certaines moyennes semblent fortement similaires entre les lecteurs compensateurs et les contrôles. C'est le cas notamment de la mesure de mémoire de travail phonologique (WISC-V), dont la moyenne des scores pour les lecteurs compensateurs est de 20,64 et est de 21,47 pour les contrôles. Bien que l'étendue des scores soit plus grande et que la distribution soit plus dispersée pour les lecteurs compensateurs (scores de 12 à 30) que pour les contrôles (16 à 26) – ce qui se traduit également dans les écarts-types (respectivement 4,67 et 2,50) – la différence entre les moyennes de ces deux groupes n'est pas significative ($F(1,27) = 0,36$, $p < 0,56$, $\eta^2p = 0,01$). Les moyennes obtenues sur la mesure de vocabulaire (EVIP) sont également très similaires entre les deux types de lecteurs. Les lecteurs compensateurs ont une moyenne de 129,93 et celle des contrôles se retrouve à 133,47. Dans cette épreuve, ce sont les lecteurs contrôles qui présentent des données plus dispersées, tel que le montre la figure 4.8. L'écart-type pour les lecteurs compensateurs est de 10,03 et il est de 13,45 pour les contrôles. Il n'y a pas non plus de différence significative entre les moyennes des lecteurs compensateurs et des contrôles sur cette mesure ($F(1,27) = 0,64$, $p < 0,43$, $\eta^2p = 0,02$). Les mesures de fluence verbale (L2MA-2) et de dénomination rapide de lettres (Exalang 8-11) n'entraînent pas non plus de différence significative entre les moyennes des deux types de lecteurs ($F(1,27) = 1,43$, $p < 0,24$, $\eta^2p = 0,05$ et $F(1,27) = 1,50$, $p < 0,23$, $\eta^2p = 0,05$). Encore ici, les contrôles ont des scores légèrement plus dispersés que ceux des lecteurs compensateurs.

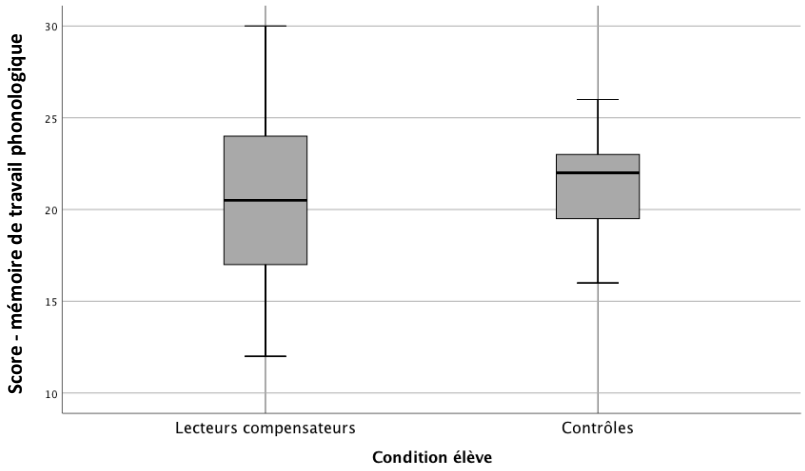


Figure 4.7 Scores obtenus en mémoire de travail phonologique (WISC) selon le type de lecteurs

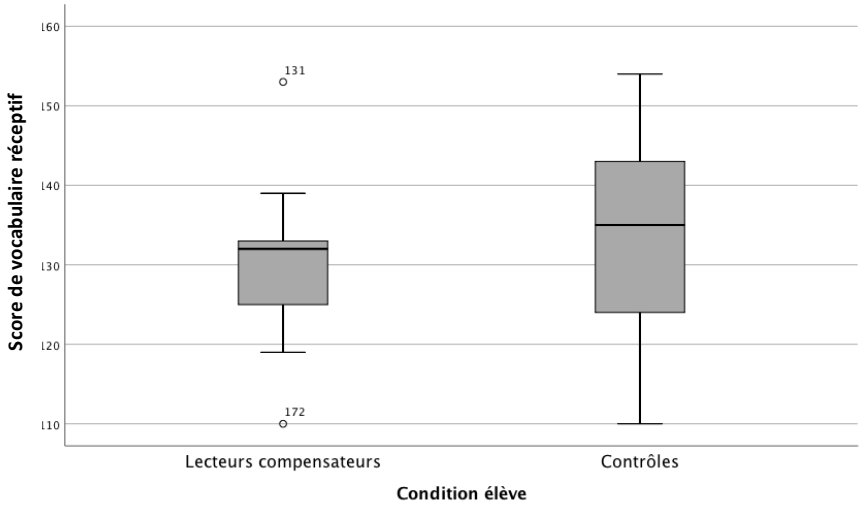


Figure 4.8 Scores obtenus en étendue de vocabulaire réceptif (EVIP) selon le type de lecteurs

La mesure de conscience phonologique (CELF), quant à elle, mène à des résultats significativement différents entre les lecteurs compensateurs et les contrôles ($F(1,27) = 6,39$, $p < 0,02$, $\eta^2 p = 0,19$), les lecteurs compensateurs ayant une moyenne de 59,36 alors que les contrôles ont une moyenne de 69,47. La figure suivante montre que l'étendue des données est similaire entre les deux groupes.

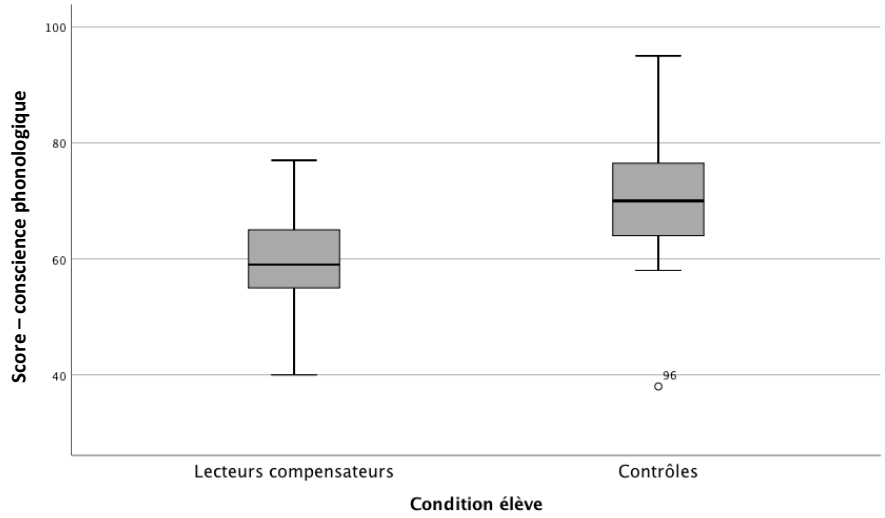


Figure 4.9 Scores obtenus en conscience phonologique (CELF) par type de lecteurs

La mesure de dénomination rapide de chiffres (BALE) mène elle aussi à des résultats significativement différents entre les lecteurs compensateurs et les contrôles ($F(1,27) = 15,51$, $p < 0,01$, $\eta^2 p = 0,37$), les lecteurs compensateurs ayant une moyenne de 27,51 alors que celle des contrôles est de 21,79. La figure suivante montre la dispersion des données.

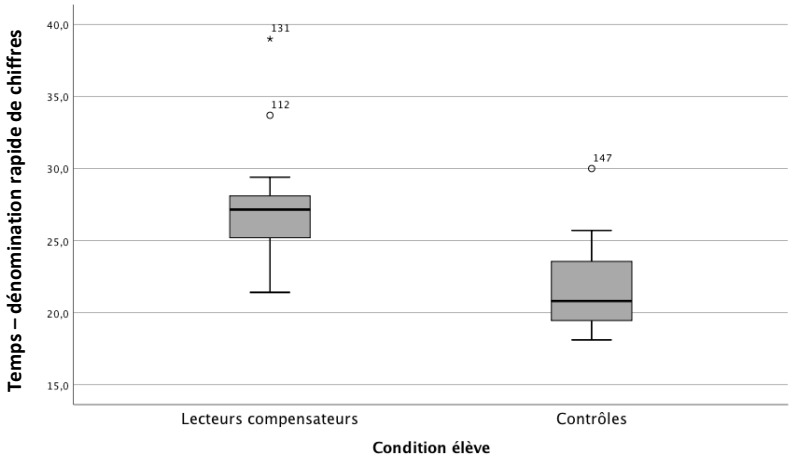


Figure 4.10 Temps obtenus en dénomination rapide de chiffres (BALE) selon le type de lecteurs

4.3.2 Tâche expérimentale

Avant de présenter les données comportementales, il convient de noter que les latences rapportées ne concernent que les bonnes réponses fournies sur les mots. En effet, puisqu'il s'agit d'une tâche

« Go/NoGo », seuls les mots sont indiqués par la pression d'un bouton, et, ce faisant, les temps de réponse ne peuvent être calculés que sur ces items. De plus, afin d'assurer que les élèves ont bel et bien fait la tâche demandée, les participants dont le taux de réussite sur les pseudomots est inférieur à 54% n'ont pas été retenus dans l'échantillon (n=3).

4.3.2.1 Données comportementales (précision et latence)

Pour les données comportementales, ce sont des analyses de variance (ANOVA 2 (lecteurs compensateurs/lecteurs contrôles) X 3 (items reliés, non reliés et neutre)) à mesures répétées qui ont été réalisées sur les taux de réussite ainsi que sur les latences. Dans ces analyses, le type de lecteur (lecteur compensateur ou lecteur contrôle) est un facteur intersujet, alors que le type de lien entre l'amorce et la cible (relié, non relié ou neutre) est un facteur intrasujet.

Pour la tâche au délai (SOA) court, les résultats globaux (n=26), qui sont rapportés au tableau 4.6, indiquent des différences entre les types d'items. En effet, le type de lien entre l'amorce et la cible mène à des taux de réussite significativement différents dans l'ensemble ($F(2,48) = 28,13$, $p < 0,01$, $\eta^2p = 0,54$). Les analyses de contraste à propos de ces différences de taux de réussite indiquent que les items reliés sont mieux réussis que les items neutres (taux de réussite de respectivement 87,62% et 81,56%). Cette différence de 6,06% est significative ($F(1,24) = 10,01$, $p < 0,004$, $\eta^2p = 0,29$). Les items reliés sont également mieux réussis que les items non reliés (74,58%). Cette différence est également significative ($F(1,24) = 82,92$, $p < 0,01$, $\eta^2p = 0,776$). La différence de performance entre les items non reliés et les items neutres, qui est de 6,98%, est aussi significativement différente ($F(1,24) = 15,77$, $p < 0,001$, $\eta^2p = 0,397$). Les résultats obtenus à propos des latences dans la tâche au délai court présentent un patron de résultats similaire. Dans l'ensemble, le type de lien entre l'amorce et la cible mène à des temps de réponse significativement différents ($F(2,48) = 28,06$, $p < 0, 01$, $\eta^2p = 0,54$). Les analyses de contraste à propos de ces différences indiquent que les items reliés (temps de réponse de 840 ms) sont identifiés plus rapidement que les items neutres (889 ms). Cette différence de 49 ms est significative ($F(1,24) = 32,41$, $p < 0, 01$, $\eta^2p = 0,58$). Les items reliés sont également identifiés plus rapidement que les items non reliés (905 ms). Cette différence de 66 ms est également significative ($F(1,24) = 46,40$, $p < 0, 01$, $\eta^2p = 0,66$). La différence de latence entre les items non reliés et les items neutres, qui est de 17 ms est également significativement différente ($F(1,24) = 4,73$, $p < 0,04$, $\eta^2p = 0,17$).

Pour le délai (SOA) long (n=25), le type de lien entre l'amorce et la cible ne mène pas à des taux de réussite significativement différents dans l'ensemble. Ceci pourrait être attribuable notamment à l'effet plafond que l'on retrouve chez les élèves contrôles (ceci sera présenté plus en détails ci-dessous). La tendance dans les moyennes semble indiquer que les items reliés sont mieux réussis que les items non reliés qui sont, à leur tour, mieux réussis que les items neutres. Cette tendance (reliés < non reliés < neutres) est légèrement différente de celle observée dans le SOA court (reliés < neutres < non reliés), mais n'est pas statistiquement significative. Concernant les latences, les résultats indiquent des différences significatives entre les types d'items ($F(2,46) = 34,15$, $p < 0,01$, $\eta^2 p = 0,60$), selon la séquence reliés < non reliés < neutres, séquence qui est légèrement différente que celle du SOA court (reliés < neutres < non reliés). Les analyses de contraste à propos de ces différences indiquent que les items reliés (temps de réponse de 813 ms) sont identifiés plus rapidement que les items non reliés (879 ms). Cette différence de 66 ms est significative ($F(1,23) = 34,04$, $p < 0, 01$, $\eta^2 p = 0,60$). Les items reliés sont également identifiés plus rapidement que les items neutres (921 ms). Cette différence de 108 ms est également significative ($F(1,23) = 86,90$, $p < 0, 01$, $\eta^2 p = 0,79$). Les latences des items non reliés et des items neutres, dont la différence est de 42 ms, sont aussi significativement différentes ($F(1,23) = 6,40$, $p < 0,02$, $\eta^2 p = 0,22$).

Les résultats par type de lecteurs mènent à certaines différences significatives. Au délai (SOA) court, une différence significative est relevée, malgré le petit nombre de participants, entre les performances des lecteurs compensateurs et celle des lecteurs contrôles, et ce, tant concernant la précision des réponses ($F(2,48) = 3,34$, $p < 0,04$, $\eta^2 p = 0,12$), que la latence de celles-ci ($F(2,48) = 6,21$, $p < 0,01$, $\eta^2 p = 0,21$). Les analyses de contraste sur les taux de réussite font ressortir que c'est particulièrement les performances entre les items non reliés et les items neutres qui sont significativement différentes entre les lecteurs compensateurs et les contrôles ($F(1,24) = 6,35$, $p < 0,02$, $\eta^2 p = 0,21$). Tel qu'illustré à la figure 4.11, les moyennes sont similaires chez les lecteurs compensateurs pour les items non reliés et pour les items neutres, alors qu'elles semblent plus distinctes chez les lecteurs contrôles. Tous les contrastes impliquant les items reliés ne sont pas significativement différents entre les lecteurs compensateurs et les contrôles. Le patron général des résultats de la figure 4.11 met en lumière le fait que, chez les lecteurs compensateurs, les items reliés sont davantage réussis que les items non reliés et les items neutres (qui eux, sont sensiblement réussis aussi bien les uns que les autres). Chez les lecteurs contrôles, les items reliés et les items neutres sont également bien réussis, alors que les items non reliés sont moins bien réussis. Pour les latences, il convient d'abord de noter que celles-ci sont toutes plus longues chez les lecteurs compensateurs que les latences des lecteurs contrôles. Les analyses de contraste indiquent que ce sont

les performances entre les items reliés et les items neutres qui sont significativement différentes entre les lecteurs compensateurs et les contrôles ($F(1,24) = 15{,}56$, $p < 0{,}01$, $\eta^2 p = 0{,}39$). Les lecteurs contrôles ont des latences plus courtes pour les items reliés que pour les items neutres, alors que ce n'est pas le cas des lecteurs compensateurs qui présentent des latences similaires pour ces catégories d'items. Les lecteurs compensateurs ont des latences légèrement plus longues pour les items non reliés (voir figure 4.11). Il convient également de noter que les écarts-types sont toujours plus grands chez les lecteurs compensateurs que chez les lecteurs contrôles.

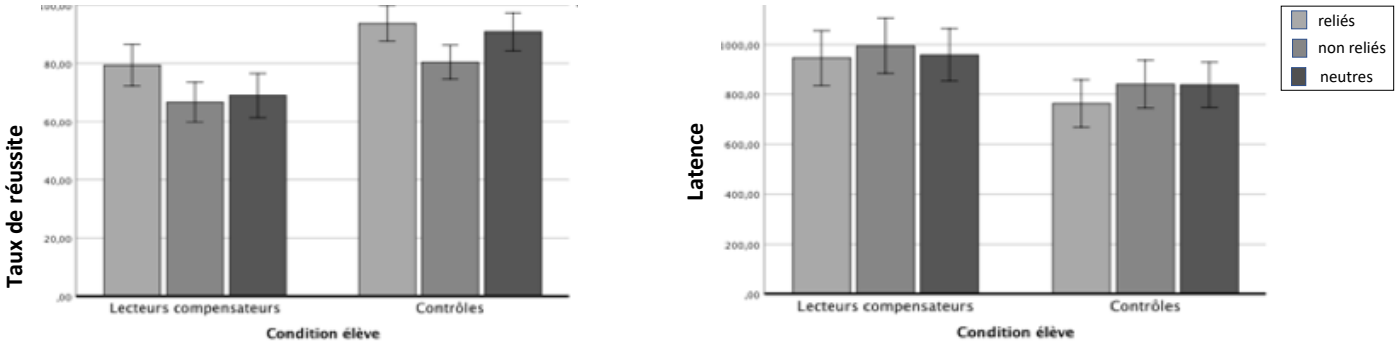


Figure 4.11 Moyennes des taux de réussite et des latences pour le délai (SOA) court par type d'items et par type de lecteurs.

Note. La précision est à gauche; les latences sont à droite.

Avec le SOA long, aucune différence significative n'est relevée selon le type de lecteurs, ni sur la précision des réponses ni sur la latence de celles-ci. Les moyennes des lecteurs compensateurs et des lecteurs contrôles présentées au tableau 4.6 indiquent que, bien que ces différences ne soient pas significatives, la tendance dans les moyennes de taux de réussite n'est pas la même. Les lecteurs compensateurs réussissent mieux les items reliés (78,83%) que les items neutres (74,02%) et que les items non reliés (72,09%). Les lecteurs contrôles, quant à eux, réussissent aussi bien les items reliés (91,60%) que les items non reliés (91,60%), qui sont tous légèrement mieux réussis que les items neutres (89,92%). En réalité, les taux de réussite des lecteurs contrôles pendant cette tâche sont très élevés en général, ce qui témoigne d'un effet plafond pour cette catégorie de participants, tel qu'illustré à la figure 4.12. Cet effet plafond empêche donc la variabilité entre les différents types d'items pour la précision. Pour les latences, encore ici, celles des lecteurs compensateurs sont généralement plus grandes que celles des lecteurs contrôles. La tendance dans les moyennes est similaire entre les deux types de lecteurs, et ne ressort donc pas statistiquement significative : les items reliés (LC :896 ms; Cont :748 ms) sont identifiés plus rapidement

que les items non reliés (LC :997 ms; Cont :786 ms), qui sont à leur tour identifiés plus rapidement que les items neutres (LC :1032 ms; Cont :833 ms).

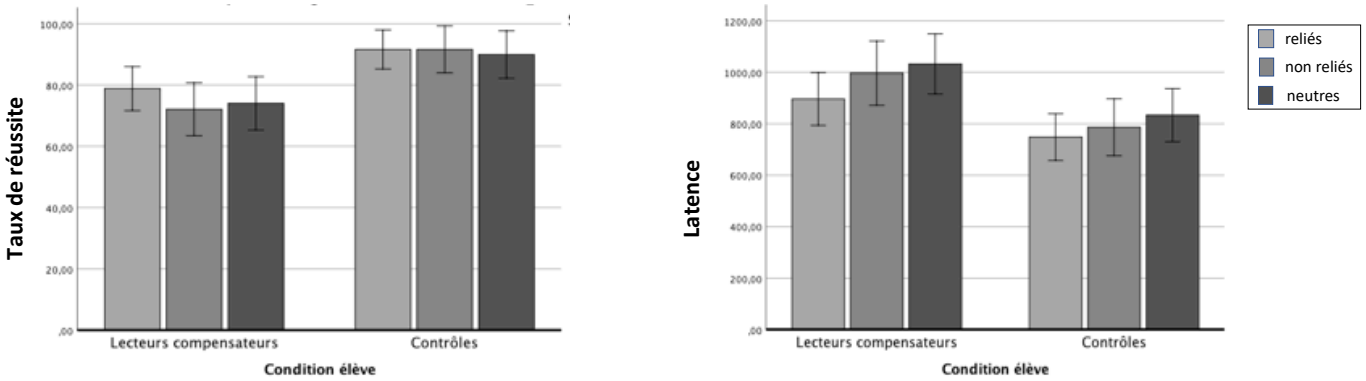


Figure 4.12 Moyennes des taux de réussite et des latences pour le délai (SOA) long par type d'items et par type de lecteurs.

Note. La précision est à gauche; les latences sont à droite.

Tableau 4.6 Moyennes et écarts-types à la tâche d'amorçage sémantique

		Reliés	Non reliés	Neutres
		SOA court – Précision (%)		
Lecteurs compensateurs		79,37 (17,03)	66,66 (15,00)	68,98 (17,96)
Contrôles		93,73 (4,43)	80,39 (6,73)	90,78 (5,08)
	TOTAL	87,65 (13,39)	74,58 (12,78)	81,56 (16,25)
		SOA court – Latence (ms)		
Lecteurs compensateurs		945 (259)	995 (245)	958 (235)
Contrôles		763 (79)	840 (110)	837 (100)
	TOTAL	840 (197)	905 (192)	889 (177)
		SOA long – Précision (%)		
Lecteurs compensateurs		78,83 (16,09)	72,09(20,03)	74,02 (20,06)
Contrôles		91,60 (5,98)	91,60 (5,63)	89,92(4,43)
	TOTAL	85,98 (13,00)	83,01 (16,79)	82,92 (15,89)
		SOA long – Latence (ms)		
Lecteurs compensateurs		896 (232)	997 (288)	1032 (269)
Contrôles		748 (81)	786 (84)	833 (79)
	TOTAL	813 (178)	879 (223)	921 (209)

4.3.2.2 Données obtenues par les potentiels évoqués (amplitude et latence)

Réviser « amplitudes les moins négatives » - « Plus ou moins grande amplitude négative »

Pour les données d'ERP, ce sont des analyses de variance (ANOVA) à mesures répétées qui ont été réalisées sur l'amplitude des potentiels évoqués, ainsi que sur leur latence. Des ANOVAs distinctes ont été réalisées sur chaque composante concernée par les hypothèses de recherche, soit la P200 et la N400. Dans ces analyses, le type de lecteur (lecteur compensateur ou lecteur contrôle) est un facteur intersujet, alors que le type de lien entre l'amorce et la cible (relié, non relié ou neutre) ainsi que la latéralité (ou la localisation lorsqu'il y a 4 regroupements d'électrodes) de l'activité électrique sont des facteurs intrasujet. Dans les analyses statistiques présentées ci-après, toutes les valeurs P ont été ajustées avec la correction pour la non-sphéricité de Greenhouse-Geisser chaque fois que nécessaire. Seules les données sur les mots sont conservées pour les mêmes raisons que celles évoquées précédemment pour les données comportementales. Sont rapportées dans les premières sections les données concernant la tâche avec un

SOA court – d'abord avec les données de précision et de latence de la composante N400, ensuite avec les données de précision et de latence de la composante P200. Les données concernant la tâche avec le SOA long sont ensuite présentées suivant la même organisation que le SOA court, c'est-à-dire la présentation des données de précision et de latence de la composante N400, suivi de la présentation des données de précision et de latence de la composante P200.

4.3.2.2.1 Données d'amplitude de la N400 (SOA court)

Dans la tâche au SOA court, les données concernant l'amplitude de la N400 témoignent d'un effet principal de la latéralité ($F(1,18) = 9,38$, $p < 0,01$, $\eta^2p = 0,34$), davantage de négativité à droite. On retrouve également un effet principal de la sémantique ($F(2,36) = 6,95$, $p < 0,01$, $\eta^2p = 0,28$), avec une amplitude globalement plus négative pour les items neutres (environ 2 µV de moins) que pour les items reliés et non reliés.

Il existe un effet d'interaction entre la latéralité et le type de lecteur ($F(1,18) = 7,54$, $p < 0,01$, $\eta^2p = 0,30$), les lecteurs compensateurs ayant une activité plus négative dans l'hémisphère droit (-9,75 µV) que dans l'hémisphère gauche (-3,32 µV), alors que les lecteurs contrôles ont une négativité similaire (-5,64 µV à gauche et -6 µV à droite). Une interaction entre la latéralité et le lien sémantique est également mesurée ($F(2,36) = 4,30$, $p < 0,02$, $\eta^2p = 0,19$). Dans l'hémisphère gauche, les items sémantiquement reliés sont les moins négatifs (-3,52 µV), suivis des items non reliés (-4,01 µV), et des items neutres (-6,95 µV). Dans l'hémisphère droit, ce sont les items non reliés qui sont les moins négatifs (-6,46 µV), suivis des items reliés (-7,27 µV), puis finalement des items neutres (-8,19 µV). Il y a un effet d'interaction entre la latéralité, le lien sémantique et le type de lecteur ($F(2, 36) = 4,44$, $p < 0,02$, $\eta^2p = 0,20$). La figure 4.13 présente visuellement les amplitudes associées à cette interaction. Dans l'hémisphère gauche, les lecteurs compensateurs présentent une amplitude moyenne moins négative pour les items reliés que pour les items non reliés (différence de 2,64 µV). Ils présentent également une différence similaire entre les items non reliés et les items neutres (différence de 3,20 µV), ces derniers provoquant les amplitudes les plus négatives. Les lecteurs contrôles, quant à eux, présentent une amplitude négative très similaire pour les items reliés et les items non reliés (-5,15 µV et -4,48, µV respectivement), et ont une amplitude plus négative pour les items neutres (-7,28 µV, différence de 2,13 µV en comparaison aux items reliés). À droite, les contrôles présentent encore une fois des amplitudes similaires pour les items reliés et non reliés (-5,46 µV et -5,76 µV, respectivement), et pour les items neutres (légèrement plus négative, différence de 1,28 µV avec les items reliés). Pour les lecteurs compensateurs, l'amplitude est similaire pour les items reliés et

les items neutres (-10,63 µV et -10,87µV, respectivement), et est moins négative pour les items non reliés (différence d'environ 3,13 µV).

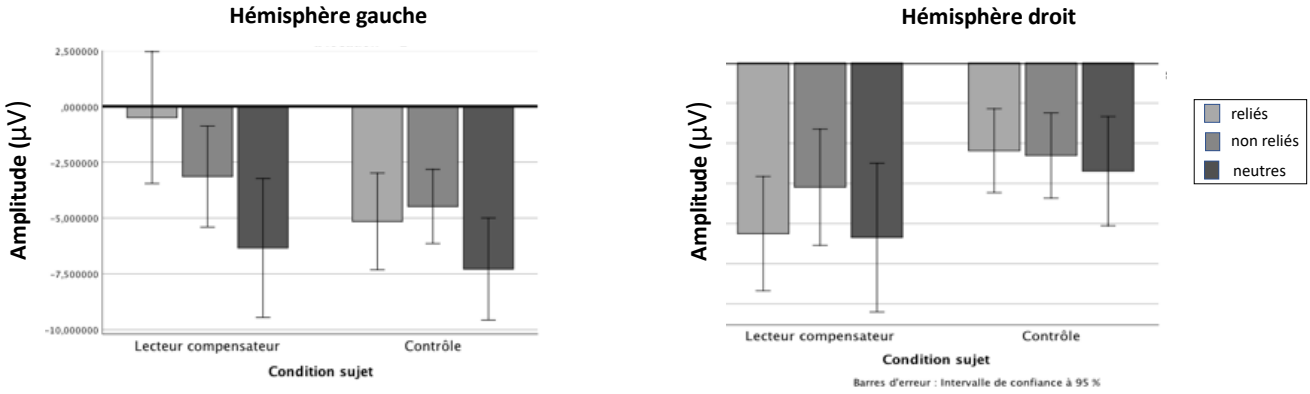


Figure 4.13 Résultats de l'amplitude des potentiels évoqués sur la N400 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite.

4.3.2.2.2 Données de latence de la N400 (SOA court)

En ce qui concerne la latence de la N400, les données obtenues ne relèvent aucun effet principal, ni de la sémantique, ni de la latéralité. Un effet d'interaction entre la latéralité et le lien sémantique est par contre présent ($F(2,36) = 4,77$, $p < 0,01$, $\eta^2p = 0,21$). Dans l'hémisphère gauche, les items neutres sont associés à une latence légèrement plus longue que les items non reliés, tel qu'indiqué par l'analyse de contraste, alors que c'est l'inverse dans l'hémisphère droit ($F(1,18) = 8,13$, $p < 0,01$, $\eta^2p = 0,31$). L'interaction entre la latéralité, le lien sémantique et le type de lecteur n'est pas significative, indiquant que les deux types de lecteurs ont une latence similaire pour la N400, tel qu'en témoigne la figure ci-dessous. Dans l'hémisphère gauche des lecteurs compensateurs, les latences sont de 337 ms pour les items reliés, de 311 ms pour les items non reliés et de 336 ms pour les items neutres. Pour les lecteurs contrôles, ces latences sont respectivement de 330 ms, de 326 ms et de 339 ms. Il n'y a donc pas de gain du lien sémantique sur la latence de la N400, et à cet égard, les lecteurs compensateurs ne sont pas distincts des lecteurs contrôles.

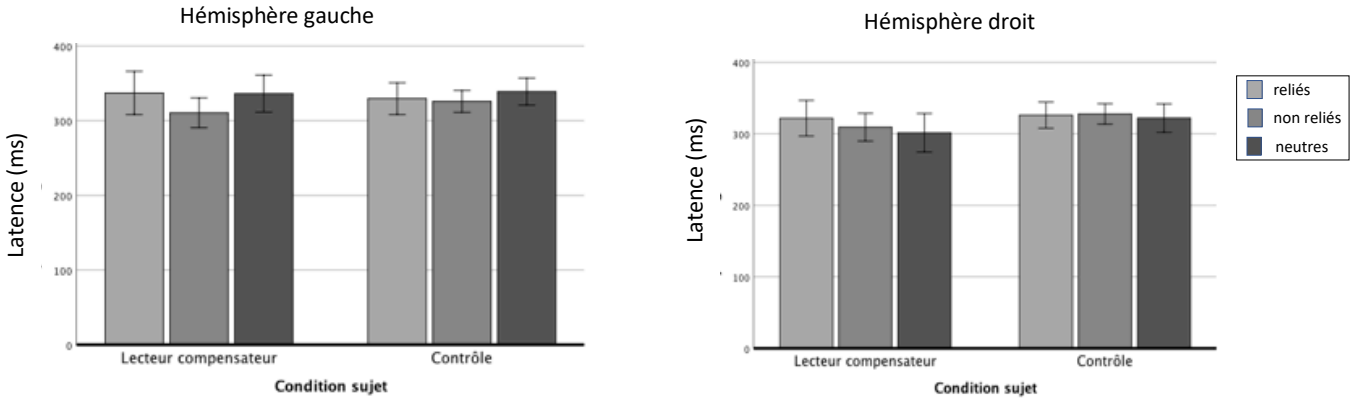


Figure 4.14 Résultats des latences (en ms) des potentiels évoqués sur la N400 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite.

4.3.2.2.3 Données d'amplitude de la P200 (SOA court)

Lorsque le SOA est court, les données concernant l'amplitude de la P200 témoignent d'un effet principal de la latéralité ($F(1,18) = 36,33$, $p < 0,01$, $\eta^2p = 0,67$), avec une amplitude plus positive à gauche. On retrouve également un effet principal de la sémantique ($F(2,36) = 4,03$, $p < 0,02$, $\eta^2p = 0,18$), avec une amplitude globalement plus positive pour les items reliés et non reliés que pour les items neutres (1,43 µV et 1,58 µV, respectivement). L'analyse de contraste indique effectivement une différence significative entre les items reliés et les items neutres ($F(1,18) = 6,74$, $p < 0,02$, $\eta^2p = 0,27$), ainsi qu'entre les items non reliés et les items neutres ($F(1,18) = 5,47$, $p < 0,03$, $\eta^2p = 0,23$), mais pas entre les items reliés et non reliés, chez l'ensemble des participants. Il existe un effet d'interaction entre la latéralité et le type de lecteur ($F(1,18) = 9,14$, $p < 0,01$, $\eta^2p = 0,34$), les lecteurs compensateurs ayant une plus grande différence d'amplitude entre l'hémisphère gauche et l'hémisphère droit (différence de 9,16 µV pour les lecteurs compensateurs et de 3 µV pour les lecteurs contrôles). Une interaction entre la latéralité et le lien sémantique est également mesurée ($F(2,36) = 5,35$, $p < 0,01$, $\eta^2p = 0,23$). Bien que globalement les amplitudes soient plus positives à gauche, on note de plus grandes différences entre l'hémisphère gauche et l'hémisphère droit sur les items reliés et les items neutres (différence de 5,01 µV et de 6,76 µV, respectivement) que sur les items non reliés (différence de 3,77 µV seulement). Aussi, dans l'hémisphère gauche, ce sont les items reliés qui occasionnent l'amplitude moyenne la moins négative (3,93 µV), alors que ce sont les items non reliés dans l'hémisphère droit (-0,3 µV). Finalement, il y a un effet d'interaction entre la latéralité, le lien sémantique et le type de lecteur ($F(2, 36) = 3,75$, $p < 0,03$, $\eta^2p = 0,17$). La figure 4.15 présente visuellement les amplitudes associées à cette interaction.

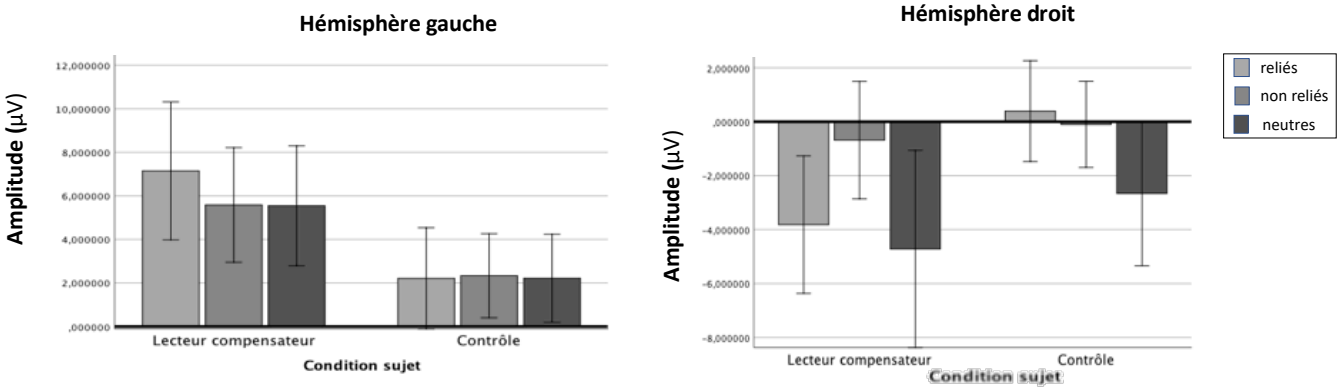


Figure 4.15 Résultats de l'amplitude des potentiels évoqués sur la P200 selon la latéralité. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite.

Dans l'hémisphère gauche, les lecteurs compensateurs présentent une amplitude moyenne plus positive pour les items reliés que pour les items non reliés et les items neutres qui présentent une amplitude similaire (différence de 1,61 µV). Les lecteurs contrôles, quant à eux, présentent une amplitude positive moindre que celle des lecteurs compensateurs, mais très similaire entre tous les types d'items (2,21 µV pour les items reliés, 2,33 µV pour les items non reliés et 2,21 µV pour les items neutres). Dans l'hémisphère droit, les contrôles présentent des amplitudes similaires pour les items reliés et les items non reliés (0,39 µV et -0,10 µV, respectivement). Ces amplitudes sont globalement plus positives comparées aux items neutres (-2,66 µV). Pour les lecteurs compensateurs, l'amplitude est plus négative pour les items neutres (-4,72 µV), et est presque similaire à l'amplitude des items reliés (-3,81 µV). L'amplitude des items non reliés est moins négative (différence de 3,13 µV). L'analyse de contraste indique que c'est la différence entre les items reliés et les items non reliés qui est significative au sein de cette interaction (F(1, 18) = 14,58, p < 0,01, $\eta^2 p$ = 0,45).

4.3.2.2.4 Données de latence de la P200 (SOA court)

En ce qui concerne la latence de la P200, les données obtenues ne relèvent aucun effet principal, ni de la sémantique, ni de la latéralité; aucun effet d'interaction n'atteint non plus le seuil de la significativité. En observant les tendances associées au lien sémantique, on remarque que les lecteurs compensateurs ont des latences légèrement plus courtes pour les items reliés (environ 210 ms) que pour les items non reliés et les items neutres (environ 217 ms et 215 ms). Pour les contrôles, la tendance est différente. Les items neutres témoignent de latences plus courtes (environ 203 ms) que les items reliés et les items non reliés, qui ont des latences similaires (environ 217 et 218 ms). Tel que mentionné, ces différences n'atteignent pourtant pas le seuil de la significativité.

4.3.2.2.5 Données d'amplitude de la N400 (SOA long)

Dans la tâche au SOA long, pour lequel 4 groupes d'électrodes sont retenus, les données d'amplitude de la N400 témoignent d'un effet principal de la localisation ($F(3,51) = 3,38$, $p < 0,05$, $\eta^2p = 0,17$), avec une amplitude globalement plus négative dans le groupe d'électrodes avant droit (– 5,76 µV) et globalement moins importante dans le groupe d'électrodes central gauche (– 2,88 µV). Les groupes d'électrodes avant gauche et central droit ont des amplitudes similaires, de -4,98 µV et de -4,06 µV, respectivement. On retrouve également un effet d'interaction entre la localisation et le type de lecteurs ($F(3,51) = 3,34$, $p < 0,03$, $\eta^2p = 0,16$), les lecteurs compensateurs ayant une activité clairement plus négative dans l'hémisphère droit (-7,32 µV à l'avant et -5,97 µV au centre) que dans l'hémisphère gauche (-3,31 µV à l'avant et -3,05 µV au centre), alors que les lecteurs contrôles ont une négativité similaire au centre (-2,78 µV à gauche et -2,95 µV à droite) et une amplitude plus négative à l'avant (-5,96 µV à gauche et -4,84 µV à droite).

Une interaction entre la localisation et le lien sémantique est également mesurée ($F(6,102) = 3,46$, $p < 0,01$, $\eta^2p = 0,17$). Les items neutres ont une amplitude plus négative à l'avant (-7,23 µV à gauche et -5,13 µV à droite), alors que les items non reliés ont une amplitude plus négative à l'avant, à droite (-6,64 µV). Les items reliés, quant à eux, présentent une négativité moins élevée à gauche (3,85 µV à l'avant et -3,16 µV au centre). Les analyses de contraste indiquent que ce sont particulièrement les différences entre les items reliés et les items neutres qui sont significatives.

Aucune interaction n'impliquant le type de lecteur n'atteint le seuil de signification statistique.

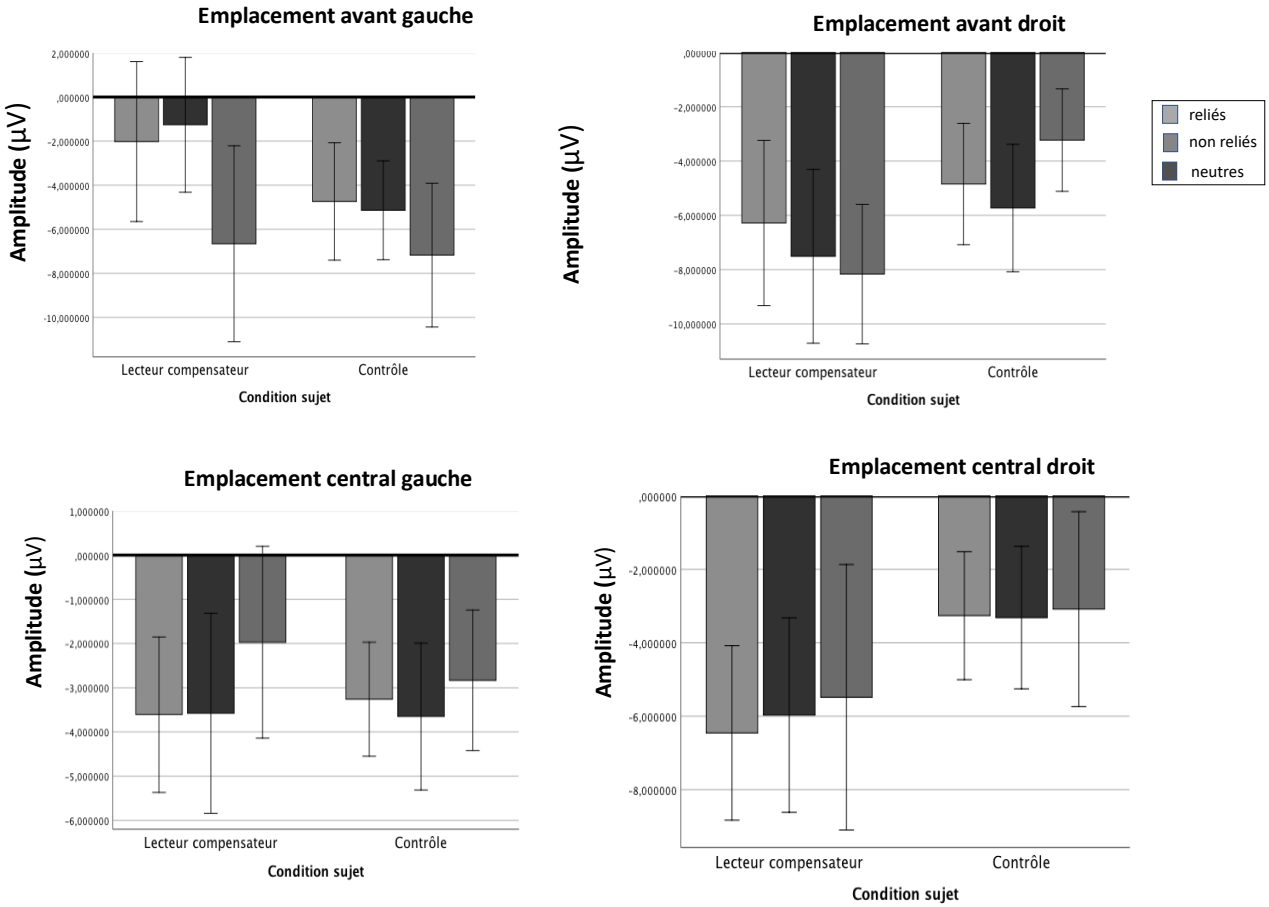


Figure 4.16 Résultats de l'amplitude des potentiels évoqués sur la N400 selon la latéralité au délai (SOA) long. Les données de l'hémisphère gauche sont présentées à gauche; celles de l'hémisphère droit sont présentées à droite; les données des électrodes avant sont présentées en haut; et celles des électrodes du centre sont présentées en bas.

4.3.2.2.6 Données de latence de la N400 (SOA long)

En ce qui concerne la latence de la N400, les données révèlent uniquement un effet principal de la localisation ($F(3,51) = 6,14$, $p < 0,01$, $\eta^2p = 0,27$), avec des latences en moyenne plus longues à gauche qu'à droite (385 ms à l'avant gauche, 323 ms au centre gauche et 373 ms à l'avant droit, 362 ms au centre droit). Aucune interaction n'entraîne de différence statistiquement significative dans les résultats de latence de la N400.

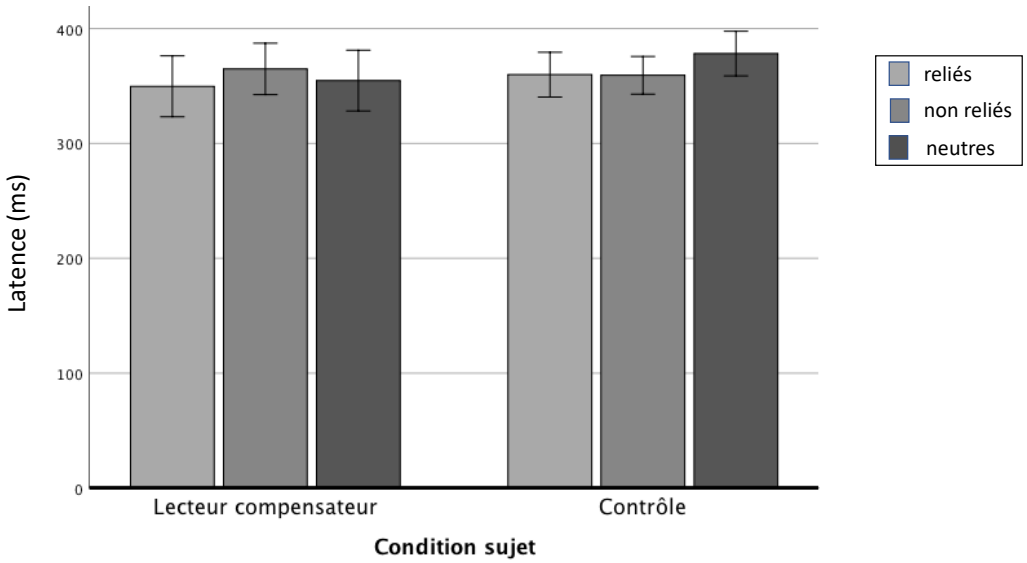


Figure 4.17 Résultats des latences (en ms) des potentiels évoqués sur la N400 au délai (SOA) long.

4.3.2.2.7 Données d'amplitude et de latence de la P200 (SOA long)

Dans la tâche au SOA long, les données d'amplitude de la P200 témoignent uniquement d'un effet principal de la localisation ($F(3,51) = 12,22$, $p < 0,01$, $\eta^2 p = 0,42$), avec une positivité plus importante à gauche (en moyenne 4,32 µV) qu'à droite (en moyenne 0,48 µV). Tous les autres effets ainsi que les interactions ne mènent pas à des différences significatives.

4.4 Discussion des résultats

L'objectif de cette étude est d'apporter un éclairage sur le fonctionnement du mécanisme de compensation en lecture chez des lecteurs compensateurs de 6e année du primaire. D'abord, puisque la faible puissance statistique a empêché d'utiliser les données à propos des composantes associées à la lecture comme variable contrôles dans les analyses statistiques, on documente ces composantes pour déterminer si on relève des particularités chez les lecteurs compensateurs : vocabulaire réceptif, fluence verbale, mémoire de travail phonologique, conscience phonologique et dénomination rapide. Ensuite, on cherche à vérifier les hypothèses suivantes concernant le traitement sémantique des lecteurs compensateurs : 1) il y a un effet facilitateur plus important chez les lecteurs compensateurs que chez les lecteurs contrôles, et ce, autant pour une amorce présentée rapidement que pour une amorce présentée longuement (recours à la sémantique malgré la contrainte de temps); 2) les lecteurs compensateurs activent plus tôt l'information sémantique pendant l'identification des mots écrits que ne le font les

lecteurs contrôles, et ce, malgré un déficit du traitement de la forme visuelle et phonologique des mots. Il semble pertinent de rappeler ici que l'échantillon étant relativement petit, plusieurs statistiques manquent de puissance. Les résultats et leurs interprétations représentent donc une étude exploratoire; certains résultats statistiquement non significatifs pourraient en réalité être le témoin de ce manque de puissance plutôt qu'une absence réelle d'effet. Par ailleurs, le petit nombre de sujets rend les statistiques plus sensibles à la variabilité interindividuelle. Ainsi, les performances de tous les lecteurs compensateurs ne sont peut-être pas similaires, mais sont tout de même considérées comme un ensemble. Par ailleurs, on sait également que le groupe de lecteurs compensateurs n'est pas homogène sur plusieurs variables (p.ex. présence d'un diagnostic formel ou non, présence d'un plan d'intervention à l'école, etc.), ce qui peut également jouer sur les résultats obtenus.

4.4.1 Composantes associées à la lecture-écriture

Il convient de rappeler que les données concernant les composantes associées à la lecture-écriture ont été initialement collectées afin d'être utilisées comme variables contrôles. Le nombre trop restreint de participants a entraîné une modification à l'utilisation de ces mesures; elles sont ainsi rapportées dans le but de mieux décrire l'échantillon. La caractérisation des lecteurs compensateurs en lien avec les composantes associées à la lecture-écriture indique que ceux-ci ne diffèrent pas beaucoup des contrôles dans les données obtenues.

D'abord, les lecteurs compensateurs ont une performance similaire aux contrôles quant au vocabulaire réceptif. Cela pourrait signifier que la capacité des lecteurs compensateurs à compenser leur difficulté d'identification des mots écrits ne serait pas uniquement une question de vocabulaire; les résultats obtenus ne nous permette pas d'avancer que les lecteurs compensent parce qu'ils ont un meilleur vocabulaire que leurs pairs. Il est toutefois possible que les lecteurs compensateurs utilisent d'une manière différente leurs connaissances lexicales en comparaison aux lecteurs contrôles, mais ces résultats ne nous permettent pas de vérifier cette hypothèse.

Dans une population d'élèves dyslexiques doués, van Viersen *et al.* (2016) ont obtenu un résultat légèrement différent concernant le vocabulaire expressif. Dans leur étude, les élèves dyslexiques doués (n=27, entre 7 et 9 ans) ont un vocabulaire expressif similaire aux élèves doués sans dyslexie (n=31), et significativement meilleur que celui des élèves au développement typique (n=31) et que celui des élèves dyslexiques non doués (n=33). Ces résultats amènent les auteurs à conclure que le vocabulaire est une

force chez les élèves dyslexiques doués, ce qui pourrait expliquer, en plus de leur bonne performance en mémoire de travail et en grammaire, que la compréhension écrite n'est pas aussi altérée qu'elle ne devrait l'être compte tenu de leur difficulté importante à identifier les mots écrits. Ouellette (2006) souligne effectivement que différentes mesures de vocabulaire sont liées à différentes composantes de la lecture. Ainsi, dans son étude, l'étendue du vocabulaire réceptif était la seule variable du vocabulaire oral qui prédisait la performance en décodage. L'étendue du vocabulaire expressif prédisait la reconnaissance visuelle de mots et la profondeur des connaissances du vocabulaire prédisait mieux la compréhension écrite. Il est donc fort possible que les mesures de vocabulaire réceptif et expressif ne soient pas liées de la même façon à la lecture. Par ailleurs, Cavalli *et al.* (2015), dans une étude portant sur des étudiants universitaires dyslexiques, montrent que c'est particulièrement sur la mesure de profondeur du vocabulaire réceptif que les dyslexiques sont forts (en comparaison à leurs pairs de même âge chronologique). Une mesure de profondeur du vocabulaire n'aurait ainsi peut-être pas donné le même résultat.

Quoi qu'il en soit, les résultats de la présente étude ne semblent pas indiquer une force au niveau de l'étendue du vocabulaire réceptif chez les lecteurs compensateurs. Si tel est vraiment le cas, en se référant au modèle *Simple View of Reading* (Gough et Tunmer, 1986), on peut envisager que le niveau similaire de compréhension écrite entre les lecteurs compensateurs et les lecteurs contrôles soit expliqué de deux manières distinctes. L'équation étant « Compréhension écrite = décodage X compréhension langagière », les lecteurs compensateurs ont un niveau moyen de compréhension écrite parce que bien qu'une grande partie de leurs ressources cognitives est dédiée aux processus spécifiques d'identification des mots écrits (ici appelés décodage), ils recourent à certains éléments provenant notamment de la compréhension langagière pour compenser. Les lecteurs contrôles pourraient avoir un niveau moyen en compréhension écrite parce que certains éléments de leur compréhension langagière seraient moyens ou faibles. Dans ces deux cas, il ne semble pas que l'étendue du vocabulaire réceptif soit pertinente, ni pour expliquer la compensation chez les lecteurs compensateurs, ni pour expliquer une faiblesse chez les lecteurs contrôles. Tel que mentionné précédemment, les différentes mesures de vocabulaire ne sont pas toutes liées à la lecture de la même manière. Celle de l'étendue du vocabulaire réceptif n'est peut-être tout simplement pas suffisamment sensible pour mettre en lumière la compensation en lecture. Il demeure finalement la possibilité que les lecteurs compensateurs aient une étendue de vocabulaire réceptif meilleure que celle d'élèves qui rencontrent également des difficultés d'identification des mots écrits mais qui ne compensent pas et qui, en conséquence, ont une performance en compréhension écrite faible. Malheureusement, la

présente étude ne comporte pas cette population d'élèves. Cette hypothèse demeure à vérifier et serait une manière de montrer une force particulière des lecteurs compensateurs en vocabulaire. Il semble pertinent de rappeler ici que les lecteurs compensateurs n'ont pas été sélectionnés sur la base de la présence d'un trouble spécifique d'apprentissage de la lecture-écriture (dyslexie-dysorthographie). Ce faisant, il est également possible que l'absence de résultats concluants quant au vocabulaire réceptif soit attribuable à cela.

La fluence verbale n'est pas non plus une composante qui distingue les lecteurs compensateurs des lecteurs contrôles, pas plus que ne l'est la mémoire de travail phonologique, telle que mesurée par l'empan de chiffres à l'endroit, à l'envers et en ordre croissant. Alors que les études documentent souvent un déficit de la mémoire phonologique chez les élèves dyslexiques (voir Wokuri *et al.*, 2018 pour une synthèse), certaines études documentent une supériorité des dyslexiques doués en mémoire de travail (van Viersen *et al.*, 2016; Berninger et Abbott, 2013) ou encore un déficit de mémoire de travail moins important chez des dyslexiques compensés que chez ceux qui ne le sont pas (Horowitz-Kraus et Breznitz, 2013). Les mesures utilisées par van Viersen *et al.* et par Berninger et Abbott sont légèrement différentes de celles de la présente recherche. Elles consistent en des scores composites incluant tant la mémoire de travail phonologique, la mémoire de travail visuelle, que la mémoire de travail visuo-spatiale. La mesure d'Horowitz-Kraus et Breznitz étant fortement similaire, il convient de noter que les résultats de la présente recherche ne mettent pas en évidence ce déficit moins important de la mémoire de travail phonologique chez les lecteurs compensateurs. Cependant, il convient de rappeler que les lecteurs compensateurs de l'échantillon ne sont pas identifiés comme étant dyslexiques; certains d'entre eux peuvent l'être, mais il ne s'agit pas d'un critère d'inclusion.

Les résultats obtenus en conscience phonologique confirment que, comme dans une multitude d'études rapportées dans le cadre théorique, les difficultés d'identification des mots écrits vécues par les lecteurs compensateurs sont hautement corrélées à un déficit de la conscience phonologique. En effet, les lecteurs compensateurs sont significativement plus faibles dans une tâche qui mobilise principalement des opérations de conscience phonémique que les lecteurs contrôles. Dans la mesure où les élèves sélectionnés pour faire partie du groupe de lecteurs compensateurs l'ont été sur la base notamment d'une faible performance en fluence de décodage de pseudomots, ce résultat est attendu, et confirme une bonne sélection des participants.

En ce qui concerne la dénomination rapide, seule la tâche de dénomination rapide de chiffres entraîne des résultats significativement différents (les performances sont similaires pour la dénomination de lettres). Ainsi, les lecteurs compensateurs sont plus lents que les lecteurs contrôles. La dénomination rapide étant souvent associée à l'identification de mots écrits (p. ex. Cohen *et al.*, 2018), il semble tout à fait attendu ici aussi que la dénomination soit significativement plus lente chez les lecteurs compensateurs en comparaison aux lecteurs contrôles puisqu'ils ont des difficultés d'identification de mots. Le pouvoir prédictif de cette habileté sur l'identification de mots écrits a été documenté (p. ex. de Jong, 2011) chez des lecteurs de tout âge et particulièrement après 9 ans, une fois que l'automatisation des processus spécifiques est atteinte. La dénomination rapide de lettres n'a pas fait ressortir de différence significative. Le nombre limité de stimuli présents dans la tâche pourrait expliquer cela.

En résumé, les données exploratoires obtenues dans le cadre de cette étude n'ont pas permis de faire ressortir une particularité des lecteurs compensateurs sur le plan des composantes associées à la lecture-écriture, que ce soit au niveau du vocabulaire réceptif, de la fluence verbale et de la mémoire de travail phonologique. Étant donné que les lecteurs compensateurs ont été identifiés sur la base d'une difficulté en fluence de décodage des pseudomots, des déficits sont attendus en conscience phonologique et en dénomination rapide. Ces déficits apparaissent bel et bien dans les données. Le petit nombre de participants ou le choix des mesures peuvent expliquer que les lecteurs compensateurs ne semblent pas avoir de force particulière expliquant la présence de compensation, mais les résultats soulèvent également la possibilité que la compensation en lecture ne repose pas sur des connaissances et habiletés particulières, mais bien sur la manière d'utiliser ou de mettre en œuvre ces connaissances et habiletés en situation de lecture.

4.4.2 Mobilisation de la sémantique lexicale en situation de lecture de mots isolés: amorçage sémantique

Toujours en considérant que les résultats de cette étude sont exploratoires, vu le petit nombre de participants, dans la tâche de décision lexicale avec amorçage sémantique, on remarque que les lecteurs compensateurs performent moins bien que les lecteurs contrôles sur le plan de l'identification de mots écrits, et ce, tant au niveau de la précision que de la latence. Cela semble confirmer ainsi leurs difficultés au niveau des processus spécifiques d'identification des mots écrits. Comme en témoignent les écarts-types, et comme cela est souvent rapporté au sein de groupes d'élèves en difficulté, les performances des lecteurs compensateurs présentent une plus grande variation que ce qui est observé pour les lecteurs

contrôles. Tous les résultats obtenus dans l'amorçage sémantique avec un délai (SOA) long semblent plus difficiles à interpréter, à cause de la moins bonne qualité des données.

4.4.2.1 Effet facilitateur de la sémantique dans la précision et les latences

Le tableau suivant présente un résumé des différents effets obtenus dans les données comportementales de précision et de latence.

Tableau 4.7 Présence d'un effet facilitateur de la sémantique

	SOA court		SOA long	
	Précision	Latence	Précision	Latence
Lecteurs compensateurs	effet facilitateur	---	Tendance, non statistiquement significative	effet facilitateur
Lecteurs contrôles	---	effet facilitateur	---	effet facilitateur

Dans l'amorçage sémantique au délai court, l'effet facilitateur du lien sémantique est présent chez les lecteurs compensateurs au niveau de la précision des réponses. Un tel effet n'est pas présent chez Les lecteurs contrôles. Ces résultats s'apparentent à ceux rapportés à la section 2.8.1 du présent cadre théorique concernant l'effet de contexte : ce sont les lecteurs en difficulté, ici les lecteurs compensateurs, qui bénéficient le plus d'un lien sémantique entre les mots au niveau de la précision lors de l'identification de mots écrits (pour une revue, voir Stanovich, 2000). Même si ces résultats sont attendus au regard de l'« effet de contexte » tant documenté dans la littérature scientifique, il est intéressant de noter que, dans la présente étude, les lecteurs compensateurs ont été ainsi identifiés sur la base de leur performance inattendue en compréhension écrite considérant leurs difficultés sur le plan de l'identification des mots écrits. Ainsi, la procédure qui a été utilisée pour repérer les lecteurs compensateurs met davantage en lumière leur capacité à compenser en situation de lecture de texte, alors que la tâche d'amorçage sémantique utilisée fait essentiellement appel à la compensation par la sémantique au niveau du mot. Les présents résultats confirment ainsi que cette capacité à compenser en situation de lecture de texte se traduit également, pour un même participant, par un plus grand recours au lien sémantique entre deux mots en situation de lecture de mots isolés. À notre connaissance, il s'agit de la première étude qui

documente cela, puisque généralement les études se centrent sur l'effet de contexte soit au niveau du mot, soit au niveau du texte (p. ex. Perfetti *et al.*, 1979 ; Schvaneveldt *et al.*, 1977).

Le fait que les contrôles ne présentent pas un tel effet facilitateur de la sémantique au niveau de la précision s'explique sans doute par leur niveau suffisant d'automatisation des processus d'identification des mots écrits. Les lecteurs compensateurs, eux, ont une plus grande dépendance au contexte puisqu'ils éprouvent des difficultés à identifier avec précision les mots écrits en recourant uniquement aux informations orthographiques et phonologiques. Par conséquent, les résultats de la présente étude vont dans le même sens que ceux obtenus dans toutes les études sur l'effet de contexte, c'est-à-dire que lorsque l'identification de mots écrits ne se fait pas automatiquement en recourant à la forme orthographique et phonologique, les lecteurs sont plus enclins à recourir à l'information sémantique disponible pour les soutenir dans cette identification. Habituellement, chez le lecteur plus avancé, ce sont les temps de réponse, ici appelés la latence, qui sont considérés (Ferrand, 2007) pour mesurer l'effet facilitateur de la sémantique parce que, souvent, un effet plafond est présent dans les données de précision des réponses. C'est exactement ce qui s'est produit chez les contrôles dans l'amorçage sémantique avec un délai (SOA) long. À cause de cela, aucune différence ne ressort comme étant statistiquement significative. Les lecteurs compensateurs semblent pourtant mieux réussir les items reliés que les items non reliés et les neutres; ce qui pourrait indiquer un effet facilitateur de la sémantique sur la précision de l'identification des mots écrits pour ces lecteurs même si les statistiques ne sont pas significatives.

Par ailleurs, toujours au niveau de la précision de l'identification des mots écrits, les amorces non reliées occasionnent chez les lecteurs contrôles un effet d'interférence lorsque le délai (SOA) est court, en ce sens que les items non reliés sont moins bien réussis que les items neutres. Dans la littérature scientifique, on documente cet effet chez des lecteurs experts lorsqu'une condition neutre est utilisée, à différents SOA (p. ex. Becker, 1980). La condition neutre est donc considérée comme un point zéro relatif; si une autre condition entraîne un gain dans la performance, la condition sera associée à un effet de facilitation; si une condition entraîne une diminution de la performance, la condition sera associée à un effet d'interférence. Tel que mentionné précédemment, ce sont surtout les latences qui sont mesurées chez l'adulte, les scores de précision atteignant souvent l'effet plafond. La présente étude se déroulant avec des élèves du primaire, il est possible d'envisager que cet effet d'interférence est présent sur la précision parce qu'il n'y a pas d'effet plafond sur les scores de précision de l'identification des mots écrits au SOA court. L'effet

d'interférence est généralement documenté comme étant le témoin de processus stratégiques (de Groot, 1986), c'est-à-dire des processus conscients et volontaires de la part des lecteurs (Ferrand, 2007). L'un de ces processus stratégiques pourrait entrer en jeu ici pour les lecteurs contrôles. Par exemple, la stratégie d'anticipation (*expectancy*) est un processus stratégique qui correspond à la génération active de candidats potentiels pour le mot cible (Neely, 1991). Le fait de générer des candidats facilite le traitement de mots sémantiquement reliés, mais interfère avec le traitement de mots non reliés, ce qui explique la moins bonne performance à ce dernier type d'items. L'effet d'interférence constaté chez les lecteurs contrôles pourrait être attribuable à cette stratégie consciente. Quoi qu'il en soit, cela semble indiquer que les lecteurs contrôles ont bel et bien le temps de traiter l'amorce malgré le délai (SOA) court. Cet effet d'interférence n'est cependant pas présent chez les lecteurs compensateurs. On pourrait suggérer que les lecteurs compensateurs n'ont pas totalement complété le traitement du mot amorce avec un SOA aussi court. D'une part, cela est très attendu vu leurs grandes difficultés en identification de mots écrits, et d'autre part, l'absence d'effet d'interférence chez eux pointe également vers l'impossibilité d'appliquer une procédure stratégique par manque de temps. Cette identification partielle de l'amorce (avec un SOA court) serait toutefois amplement suffisante pour que les lecteurs compensateurs bénéficient tout de même d'un lien sémantique entre les mots. Ainsi, on semble avoir des données qui appuient le fait que les lecteurs compensateurs recourent à l'information sémantique avant d'avoir complètement traité la forme du mot (forme orthographique et phonologique). Par ailleurs, probablement à cause de l'effet plafond, cet effet d'interférence n'est pas présent chez les lecteurs contrôles dans l'amorçage sémantique avec un délai (SOA) long.

Ces données sur l'effet d'interférence en précision de l'identification des mots écrits sont bien expliquées par le modèle de Plaut et Booth (2000) pour l'amorçage sémantique. Dans ce modèle, l'identification du mot-amorce active une certaine quantité d'information associée à ce mot jusqu'à ce que le réseau atteigne un niveau de stabilité représentant la configuration associée au sens du mot lu. Quand le réseau part de ce patron d'activation pour identifier le mot suivant et que ce mot est sémantiquement lié, le réseau est en quelque sorte déjà préparé pour activer un mot similaire, puisque sa configuration est déjà ressemblante. On considère alors que le réseau est « préactivé » pour traiter un mot similaire et donc que l'identification de ce dernier est facilitée. L'effet d'interférence, quant à lui, correspond au phénomène inverse, dans le sens où les unités activées par le mot-amorce éloignent le réseau du mot-cible non relié, puisque les unités activées par l'amorce ne sont pas pertinentes au traitement du mot-cible. Cette configuration du réseau fournit donc des indications qui sont non pertinentes (ni phonologiquement, ni

orthographiquement, ni sémantiquement) pour le traitement du mot cible, ce qui introduit du « bruit » dans le processus d'identification du mot-cible. Plutôt que de faciliter son traitement, éloigner le réseau en le dirigeant vers un mot non relié fait donc en sorte qu'il est plus difficile de faire « revenir » ce réseau sur la configuration associée au mot-cible.

En ce qui concerne les latences, on note que les lecteurs compensateurs sont globalement plus lents que les lecteurs contrôles, témoignant de leurs difficultés en identification des mots écrits. Par ailleurs, les résultats des lecteurs contrôles indiquent ici l'effet attendu, c'est-à-dire un effet facilitateur du lien sémantique, tant au SOA court qu'au SOA long. Cet effet facilitateur n'est pas toujours documenté chez des participants d'âge scolaire (en apprentissage) avec des SOA aussi courts (p. ex. Nievas et Justicias, 2004). Ces résultats indiquent pourtant qu'à un SOA de 150 ms, il est possible de constater un effet de la sémantique sur les temps de réponse d'élèves typiques. Pour ces lecteurs sans difficulté, l'effet facilitateur de la sémantique est donc uniquement présent sur les latences d'identification des mots écrits, et non sur la précision, comme c'est le cas chez les adultes lecteurs experts. On peut ainsi considérer que l'identification des mots écrits est suffisamment automatisée chez ces lecteurs pour que le lien sémantique ne change rien au niveau de la précision (ils peuvent reconnaître correctement les mots, qu'ils soient précédés ou non d'une amorce reliée), mais que des gains au niveau des temps de réponse sont observables. C'est d'ailleurs ce que l'on constate chez l'adulte (voir Lucas, 2000 et McNamara, 2005 pour des revues). De plus, aux deux SOA utilisés dans cette thèse, les lecteurs contrôles ne présentent pas, sur les latences, l'effet d'interférence documenté précédemment. Puisque les mesures de latence ne considèrent que le temps pris pour reconnaître adéquatement un mot écrit (seules les bonnes réponses sont conservées), il est possible que les amorces non reliées entraînent effectivement plus de confusion, affectant ainsi la précision (confondre un pseudomot avec un mot ou ne pas reconnaître un mot), mais ne changent rien dans les temps de réponse pour les mots adéquatement reconnus; ils ne prennent pas davantage de temps pour identifier les mots précédés d'une amorce non reliée que ceux précédés d'une amorce neutre. Ceci souligne l'importance de considérer tant la précision des réponses que les latences dans les études qui portent sur les enfants.

Pour les lecteurs compensateurs, par contre, l'effet facilitateur de la sémantique n'est pas présent au niveau de la latence de réponse au SOA court. Ceci signifie que le temps nécessaire pour reconnaître correctement un mot est aussi long lorsqu'il précédé d'une amorce sémantiquement liée que lorsqu'il

n'est pas précédé d'une telle amorce. Ce résultat est inattendu[30]. Considérant le caractère exploratoire de l'étude, on pourrait tout de même proposer que, à un délai (SOA) court, la sémantique fournisse aux lecteurs compensateurs un mécanisme efficace pour améliorer la précision de l'identification des mots écrits, mais pas pour améliorer la rapidité de celle-ci. Le temps nécessaire pour mettre en œuvre adéquatement le mécanisme de compensation par la sémantique serait donc aussi long que lorsque les lecteurs doivent identifier un mot sans aucun indice sémantique. Il est également possible de considérer que leurs difficultés importantes au niveau des processus d'identification des mots écrits fassent en sorte que toute identification de mots se fait lentement, lien sémantique ou non. Pourtant, en référence au modèle de Plaut et Booth (2000), la configuration sémantique du réseau sur une amorce devrait augmenter la rapidité de traitement d'un mot-cible sémantiquement relié. Par ailleurs, au SOA long, les résultats de latence indiquent un effet facilitateur de la sémantique chez les lecteurs compensateurs. On ne peut donc pas affirmer que ce sont leurs difficultés importantes qui font en sorte que toute identification de mots est longue.

On pourrait ainsi proposer que l'identification de l'amorce ne soit pas totalement complétée pour les lecteurs compensateurs à un SOA court (à cause de leur difficulté attestée à traiter la forme orthographique et phonologique des mots) et que, par conséquent, le réseau n'ait pas encore atteint un niveau de stabilité représentant la configuration associée au sens du mot-amorce. De ce fait, le réseau non stabilisé sur une seule configuration activerait en fait plusieurs unités sémantiques possibles. Toutes ces unités encore activées seraient donc des candidats possibles à utiliser comme indice pour identifier le mot-cible. Les lecteurs compensateurs devraient donc explorer toutes ces possibilités d'unités sémantiques encore activées pour tenter de faciliter l'identification du mot-cible et c'est cela qui prendrait du temps. Considérant l'effet facilitateur de la sémantique sur la précision des réponses et non sur les latences au SOA court, le système cognitif prendrait donc plus de temps pour évaluer les « candidats potentiels » mais parviendrait à porter un « jugement » adéquat sur la cible (précision). Au SOA long, laisser plus de temps aux lecteurs compensateurs pour traiter l'amorce leur permettrait ainsi de stabiliser complètement le réseau sur la configuration finale associée à la sémantique du mot amorce, et ceci constituerait donc une « préactivation » pour le traitement du mot cible, augmentant ainsi la rapidité de l'identification d'un mot sémantiquement relié.

[30] Il semble pertinent de souligner que les écarts-types sont importants pour cette catégorie de lecteur. Une hétérogénéité de la population pourrait expliquer ce résultat.

Quoi qu'il en soit, les résultats obtenus ici soulignent l'existence chez les lecteurs compensateurs d'un certain troc entre la justesse et la vitesse (dans le sens de « trade-off »; Betjemann, 2008) : un mécanisme de compensation efficace par la sémantique permet ainsi aux lecteurs compensateur un gain en précision, mais prend davantage de temps.

En somme, il semble que les lecteurs contrôles présentent des résultats comportementaux similaires à ceux documentés chez l'adulte lecteur expert, soit le fait que l'effet facilitateur de la sémantique se fait sentir uniquement sur les latences, et non pas sur la précision. Le patron des résultats obtenus par les lecteurs compensateurs est différent : au SOA court, l'effet facilitateur est documenté sur la précision; au SOA long, l'effet facilitateur semble présent sur la précision (sans que la différence ne soit statistiquement significative), mais également sur les latences. Tenant compte de leurs difficultés d'identification des mots écrits, il semble que l'effet facilitateur de la sémantique n'est pas davantage présent chez les lecteurs compensateurs que chez les lecteurs contrôles; il se manifeste simplement différemment, ce qui souligne l'importance de recourir tant aux mesures de précision qu'aux mesures de latence pour évaluer des élèves du primaire, surtout ceux en difficulté. Puisque la présente étude est exploratoire, davantage de données seraient nécessaires pour poursuivre en ce sens.

Par ailleurs, le temps de présentation de l'amorce ne semble pas entraîner de changement au niveau de l'influence de la sémantique pour les lecteurs compensateurs; ils bénéficient autant de la sémantique dans l'identification de mots écrits pour un SOA court que pour un SOA long, ce qui est conforme à l'hypothèse formulée : que ce soit une présentation brève ou longue de l'amorce, les lecteurs compensateurs utilisent la sémantique. Il semble toutefois pertinent de noter que le fait de recourir à la sémantique lexicale comme mécanisme de compensation ne permet toujours pas aux lecteurs compensateurs, même avec un SOA long, de rattraper la performance des contrôles au niveau des taux de réussite; les lecteurs compensateurs demeurent toujours globalement plus faibles.

4.4.2.2 Activation précoce de la sémantique lexicale chez les lecteurs compensateurs

Afin de vérifier si les lecteurs compensateurs activent l'information sémantique des mots plus tôt que les lecteurs contrôles, l'effet du lien sémantique entre l'amorce et la cible a été mesuré sur deux composantes de l'ERP, soit la N400 et la P200.

Généralement, dans la littérature scientifique, ce sont des incongruités sémantiques ou des éléments inattendus qui occasionnent les N400 les plus négatifs (voir la section 2.9.2 du cadre théorique). Dans cette étude, les lecteurs compensateurs ont une amplitude beaucoup plus négative pour les items neutres et beaucoup moins négative pour les items reliés (les items non reliés se situent environ entre les deux), ce qui témoigne d'un fort effet facilitateur du contexte sur l'amplitude de la N400 chez les lecteurs compensateurs dans l'hémisphère gauche. Pour les lecteurs contrôles, ce n'est pas le cas, puisque l'effet facilitateur de la sémantique n'est pas constaté dans ces données (il ne l'était pas non plus sur les taux de précision); les items neutres sont associés à davantage de négativité, mais les items sémantiquement reliés et les items non reliés provoquent des amplitudes similaires, moins négatives. Tout comme c'est le cas dans les données comportementales de précision, l'effet facilitateur du lien sémantique est présent pour les lecteurs compensateurs alors qu'il ne l'est pas pour les lecteurs contrôles. On associe fréquemment la N400 à une variété de processus qui concernent l'intégration sémantique des mots en contexte (Barber et Kutas, 2007). Certaines études documentent une absence de différenciation entre différentes catégories d'items chez des élèves en difficulté de lecture au primaire sur la N400, proposant que cette sensibilité à l'intégration sémantique n'apparait que plus tard pour ces élèves. D'autres études vont plutôt dans le sens de ce qui est mesuré dans la présente étude : les lecteurs compensateurs présentent des amplitudes distinctes pour chaque catégorie d'items. Ainsi, lorsqu'une nouvelle information est facilement intégrée sémantiquement dans son contexte immédiat, l'amplitude de la N400 est diminuée. Cette amplitude tend à diminuer énormément chez l'adulte lecteur expert, jusqu'à n'être plus du tout négative parfois lorsqu'il s'agit de deux mots sémantiquement reliés (voir références dans McNamara, 2005 : Bentin *et al.*, 1985 et Rugg, 1985). Les lecteurs contrôles de la présente étude présentent donc cette diminution de l'amplitude de la N400 autant sur des items reliés que sur des non reliés, ce qui témoigne d'une aisance similaire à « intégrer » les mots-cibles, qu'ils soient précédés ou non d'un mot sémantiquement relié. Les lecteurs compensateurs, quant à eux, intègrent beaucoup plus facilement des mots sémantiquement reliés dans le contexte immédiat que des mots non reliés. Les amplitudes moyennes de la N400 chez les lecteurs compensateurs et chez les lecteurs contrôles appuient donc, encore une fois, l'idée selon laquelle ce sont les lecteurs en difficulté qui bénéficient le plus du contexte sémantique dans l'identification des mots écrits (ceci est conforme aux données sur l'effet de contexte). Au SOA long, les résultats sont similaires à ceux décrits pour le SOA court. Cependant, le fort effet facilitateur de la sémantique apparaît dans l'hémisphère droit chez les lecteurs compensateurs. Il n'est pas rare de constater dans des études recourant à différentes techniques d'imagerie cérébrale que des élèves en grande difficulté d'identification des mots écrits activent davantage des régions de l'hémisphère droit alors que leurs pairs sans difficulté activent

surtout des régions de l'hémisphère gauche (p. ex. Costanzo *et al.*, 2013; Pugh *et al.*, 2000; Shaywitz *et al.*, 2003). Cette distinction dans la latéralisation n'était pas présente au SOA court. Les contrôles, quant à eux, présentent une négativité plus grande pour les items neutres dans l'hémisphère gauche sans toutefois présenter un effet facilitateur clair de la sémantique. Il semble donc que le délai (SOA) ne change pas le patron des résultats obtenus sur l'amplitude de la N400; la seule différence réside dans l'endroit où ces potentiels évoqués sont mesurés.

Concernant les latences des potentiels évoqués de la N400, les données obtenues dans le cadre de cette étude exploratoire n'ont pas permis de démontrer une différence significative entre les lecteurs compensateurs et les lecteurs contrôles. Cette incapacité à démontrer une différence peut être attribuable au manque de puissance statistique de la présente étude. Cependant, si, avec un plus grand échantillon, on arrivait à montrer que la latence de la N400 est similaire chez les lecteurs compensateurs et chez les lecteurs contrôles, on documenterait un phénomène surprenant. En effet, souvent, on documente un retard dans les potentiels évoqués de la N400 chez des élèves qui présentent de sévères difficultés en identification des mots écrits, comme c'est le cas chez des élèves dyslexiques (p. ex. Jednorog *et al.*, 2010; Schulz, 2008). Dans ces études portant sur des élèves de 10-11 ans, qui réalisent également des tâches d'amorçage sémantique, les latences de la N400 sont plus grandes chez les élèves dyslexiques que chez les normo-lecteurs, et les auteurs attribuent cette plus grande latence soit à une difficulté à traiter la forme (orthographique et phonologique) du mot – ce qui entraîne un délai dans l'activation de la sémantique – ou encore à un déficit de la sémantique chez les élèves dyslexiques. Quoi qu'il en soit, ce marqueur caractéristique des élèves en difficulté d'identification des mots écrits n'a pas été relevé chez les lecteurs compensateurs de l'échantillon. Cette différence pourrait également être expliquée par le fait que les lecteurs compensateurs de l'échantillon ne sont pas tous dyslexiques, comme c'est le cas dans plusieurs études rapportées.

Les lecteurs compensateurs ont pourtant une difficulté marquée et attestée à traiter la forme des mots. Puisque le décours temporel largement accepté dans la littérature scientifique (voir la section 2.6 pour les détails) indique que les premières étapes de l'identification des mots écrits concernent la forme du mot, on aurait pu penser que les difficultés documentées chez les lecteurs compensateurs auraient retardé le traitement de l'information sémantique, et donc auraient augmenté la latence de la N400. Pourtant, cela ne semble pas être le cas.

Dans le cas où la N400 n'aurait effectivement pas une latence plus grande chez les lecteurs compensateurs en comparaison aux lecteurs contrôles – alors que les études attestent que c'est le cas chez des élèves dyslexiques –, on pourrait également proposer que cela arrive parce que les lecteurs compensateurs compensent leur déficit du traitement de la forme orthographique et phonologique du mot en recourant très tôt à la sémantique. Cette compensation précoce permettrait aux lecteurs compensateurs d'arriver à une intégration sémantique (mesurée par la N400) au même moment que les lecteurs contrôles. Cette idée est effectivement appuyée par les résultats obtenus sur la composante P200. Il est toutefois primordial de noter qu'il faudra vérifier si cette explication est appuyée par des études ultérieures.

Dans l'hémisphère gauche, les lecteurs compensateurs sont les seuls à présenter un fort effet facilitateur de la sémantique sur la composante P200. Cet effet facilitateur se traduit par une amplitude beaucoup plus positive pour les mots précédés d'une amorce reliée en comparaison aux mots neutres (sans amorce) ou ceux précédés d'une amorce non reliée. Cette influence de la sémantique sur une composante aussi précoce dans l'identification des mots écrits n'est pas attendue. En effet, la composante P200 est généralement associée à l'accès à la forme orthographique (visuelle) des mots pendant l'identification des mots écrits, et non à la sémantique. La P200 présente habituellement une amplitude plus positive pour les mots fréquents en comparaison aux mots peu fréquents (p. ex. Bermudez *et al.*, 2010) ou encore pour un mot précédé d'un autre mot qui a une forme orthographique similaire (p. ex. Petit *et al.*, 2006). On y associe également l'interaction entre l'information phonologique et orthographique (Kong *et al.*, 2012), et on documente qu'elle est sensible au traitement de la phonologie (voir dans Landi et Perfetti, 2006). Bref, tous ces éléments documentés pour la P200 indiquent que cette composante est plutôt liée au traitement de la forme du mot qu'à celui de sa sémantique, et qu'une positivité plus importante est associée à une familiarité ou une facilité de traitement. D'ailleurs, plusieurs études ne documentent pas d'influence de la sémantique sur la P200 dans des tâches d'amorçage, et ce, tant en français que dans d'autres langues (p. ex. Beyersmann *et al.*, 2014; Cavalli *et al.*, 2016). Le fait que les lecteurs compensateurs présentent un effet facilitateur de la sémantique sur l'amplitude de la P200 (amplitude plus positive pour les items reliés) peut indiquer une influence précoce de la sémantique lexicale dans l'identification des mots écrits, qui est présente uniquement chez les lecteurs compensateurs et non chez les lecteurs contrôles. Une activation aussi précoce de la sémantique est déjà documentée chez l'adulte lecteur expert (p. ex. Bentin *et al.*, 1985). En effet, on documente l'influence de certains facteurs sémantiques, par des manipulations bien particulières, sur les premiers 250 ms de l'identification des mots écrits (Martin-Loeches *et al.*, 2001, Hauk *et al.*, 2012; Segalowitz *et al.*, 2009). Il est donc possible d'envisager qu'une telle activation précoce de la

sémantique puisse également être observée chez les lecteurs compensateurs sans que les manipulations mentionnées précédemment aient été réalisées. Cela signifie que le mécanisme de compensation par la sémantique lexicale interviendrait en toutes circonstances, dès que le traitement des différents aspects de la forme du mot débute, soit autour de 200 ms après la présentation du mot écrit; ce mécanisme de compensation viendrait effectivement soutenir les traitements de la forme du mot, qui sont déficitaires chez les lecteurs compensateurs.

Le fait que la sémantique influence le traitement de la forme (orthographique et phonologique) du mot à 200 ms après la présentation d'un mot écrit vient, quoi qu'il en soit, appuyer les modèles cognitifs de l'identification des mots écrits qui permettent l'intervention de la sémantique lexicale au cours du processus d'identification des mots écrits. Les modèles développementaux présentés dans le cadre théorique de cette thèse (et qui sont ceux les plus utilisés en évaluation-intervention orthopédagogique) ne précisent pas du tout le rôle de la sémantique lexicale dans l'identification des mots écrits. Ces modèles semblent ainsi considérer que le traitement de la forme du mot peut avoir lieu sans le recours à la sémantique. Les modèles connexionnistes, comme le modèle PDP, prévoient quant à eux un mécanisme qui lie la sémantique au traitement de la forme du mot. Ce sont ces derniers modèles qui semblent le mieux rendre compte de cette intervention précoce de la sémantique dans l'identification des mots. En effet, il semble que ces données soulignent l'importance de considérer le traitement de la sémantique pendant l'identification des mots écrits.

En somme, les principaux constats peuvent être résumés ainsi :
Au SOA court, les lecteurs compensateurs n'ont possiblement pas le temps de traiter entièrement l'amorce, mais cette identification partielle est suffisante pour occasionner un gain de la sémantique en précision, gain qui se traduit également par l'amplitude de la N400. Le mécanisme de compensation par la sémantique, même s'il fonctionne bien pour améliorer la précision de l'identification des mots écrits, n'est pourtant pas plus rapide qu'identifier un mot sans indice sémantique, probablement à cause de l'activation de plusieurs candidats étant donné que l'identification de mots écrits n'est pas complétée (le réseau d'activation ne s'étant pas encore stabilisé). Les latences de la N400 ne sont pourtant pas différentes entre les lecteurs compensateurs et les lecteurs contrôles, ce qui est un peu surprenant considérant la difficulté attestée des lecteurs compensateurs dans le traitement de la forme des mots, lequel a lieu dans les composantes plus précoces que la N400, et considérant également les nombreuses études qui attestent d'une latence plus grande chez les élèves qui ont des difficultés d'identification des

mots écrits. Notons cependant que le petit échantillon pourrait expliquer cette absence de différence, ainsi que la variabilité qui pourrait exister sur plusieurs variables à l'intérieur même du groupe de lecteurs compensateurs. Par ailleurs, la possibilité que la sémantique intervienne plus tôt que la N400 est appuyée par ce qui a été constaté sur la composante P200 (composante habituellement associée au traitement de la forme du mot). Alors que des mots familiers ou fréquents engendrent généralement une positivité plus grande dans cette composante, les lecteurs compensateurs ont une positivité plus grande pour les mots sémantiquement reliés alors que ce n'est pas le cas chez les lecteurs contrôles. Les lecteurs compensateurs semblent ainsi recourir très tôt à la sémantique dans le décours temporel de l'identification de mots écrits, puisqu'une influence de la sémantique est constatée chez ces lecteurs uniquement dès la composante P200. De plus, les lecteurs contrôles ne présentent pas d'effet facilitateur de la sémantique au niveau de la précision, ce qui tend à confirmer que leurs processus d'identification des mots écrits sont déjà bien automatisés. Les données d'amplitude de la N400 corroborent ces résultats comportementaux. Les lecteurs contrôles présentent simplement un effet d'interférence (d'inhibition) occasionné par les items non reliés, ce qui est plutôt un indicateur de processus stratégiques (conscients et volontaires) mis en œuvre par ces lecteurs. Ceci confirme donc que les lecteurs contrôles ont le temps, même avec un SOA court, de traiter complètement l'amorce. L'absence d'influence de la sémantique sur la composante P200 vient également appuyer le fait que les lecteurs contrôles n'ont pas autant besoin de recourir au lien sémantique pour les soutenir dans l'identification de la forme des mots, alors que c'est le cas pour les lecteurs compensateurs.

Au SOA long, les données sont moins nettes. On note en effet que les lecteurs contrôles présentent un effet plafond dans leurs taux de réussite, donc l'effet d'interférence n'apparaît plus. Pourtant, l'effet facilitateur de la sémantique, hormis l'effet plafond chez les lecteurs contrôles, apparaît tant dans la précision des lecteurs compensateurs que dans les latences de réponses pour les deux types de lecteurs. Ainsi, au SOA long, il semble que les lecteurs compensateurs ont le temps de traiter l'entièreté de l'amorce; contrairement au SOA court, les lecteurs compensateurs présentent un effet facilitateur au niveau des latences également ici. Pour les données de la N400, il convient de noter que l'on capte les effets chez les lecteurs compensateurs dans l'hémisphère droit alors que c'est dans l'hémisphère gauche pour les lecteurs contrôles, ce qui est également documenté dans d'autres études portant sur les difficultés d'identification des mots écrits.

La sémantique lexicale semble donc jouer un rôle dans la compensation d'un déficit d'identification des mots écrits lors d'une tâche de lecture de mots isolés chez des lecteurs qui, par ailleurs, comprennent relativement bien un texte qu'ils lisent malgré leur déficit attesté d'identification des mots écrits. Un SOA de 150 ms est suffisant pour que ces lecteurs puissent bénéficier de la sémantique, malgré leur difficulté à traiter la forme du mot. Finalement, les lecteurs compensateurs semblent recourir plus tôt à la sémantique pendant l'identification d'un mot écrit que leurs pairs de même âge chronologique et de même niveau de compréhension écrite.

CONCLUSION

En conclusion, cette recherche avait pour objectif d'éclairer le phénomène de la compensation en lecture, compensation mise en œuvre par une certaine proportion d'élèves qui ont une difficulté sévère d'identification des mots écrits. Cette compensation fait en sorte que leur compréhension écrite n'est pas drastiquement affectée par leur difficulté d'identification des mots écrits, alors que c'est généralement le cas dans les données empiriques du domaine. Cette catégorie particulière de lecteurs qui ont des difficultés que l'on peut qualifier de non apparentes sont appelés lecteurs compensateurs.

Puisque peu d'écrits scientifiques sont disponibles pour documenter la situation des lecteurs compensateurs, une première étude avait pour but de déterminer les mesures pertinentes au repérage de lecteurs compensateurs, et de procéder à ce repérage dans des classes de 5^{e} année du primaire. Les résultats indiquent que la sélection d'une mesure de compréhension écrite (réponses à des questions ouvertes de compréhension suite à la lecture autonome d'un texte narratif) ainsi qu'une mesure de fluence de décodage de pseudomots (nombre de pseudomots lus correctement en une minute) permettent de repérer des lecteurs compensateurs. Selon les distributions observées à ces deux mesures, les critères ont ainsi été définis: les lecteurs compensateurs doivent obtenir un score de compréhension écrite dans la moyenne supérieure ou mieux, ainsi qu'un score de fluence de décodage de pseudomots à au moins une écart-type sous la moyenne. Ces seuils ont permis d'identifier 12 élèves lecteurs compensateurs dans l'échantillon de 165 élèves, ce qui représente environ 7% de l'échantillon.

Dans la deuxième étude, qui s'est avérée être exploratoire, on a tenté de répondre à la question suivante pour tenter de décrire plus en détails le mécanisme de compensation par la sémantique lexicale : le traitement sémantique des lecteurs compensateurs est-il mobilisé de manière particulière en identification de mots écrits présentés de manière isolée? Pour ce faire, différentes composantes associées à la lecture-écriture ont été mesurées, puis une tâche d'amorçage sémantique avec décision lexicale a été proposée. Deux différents délais (SOA) de présentation de l'amorce ont été utilisés, soit un court et un long. Les résultats obtenus sur les composantes associées à la lecture-écriture indiquent que seules les mesures de conscience phonologique et de dénomination rapide de chiffres distinguent les lecteurs compensateurs des lecteurs contrôles; la performance des premiers étant plus faible que celle des seconds. Ces résultats sont conformes à ce qui est attendu, puisque ces deux mesures sont connues pour être reliées à l'habileté à identifier des mots écrits, alors que les lecteurs compensateurs ont

justement été sélectionnés sur la base de leur difficulté importante à identifier les mots écrits. Par ailleurs, aucune différence significative n'est ressortie sur les mesures de vocabulaire, de mémoire de travail phonologique et de fluence verbale. Ceci peut être expliqué par le petit échantillon ou la variabilité interindividuelle des lecteurs compensateurs, mais indiquent tout de même à ce stade-ci que les lecteurs compensateurs n'ont pas de force particulière (parmi ces mesures) qui expliquerait la présence de compensation.

Dans la tâche d'amorçage sémantique, on a tenté de vérifier l'hypothèse selon laquelle l'effet facilitateur d'un lien sémantique entre l'amorce et la cible devrait se manifester de manière plus forte chez les lecteurs compensateurs que chez les contrôles, et ce, autant lorsque l'amorce est présentée rapidement que lorsque l'amorce est présentée plus longuement. Il s'est avéré que l'effet facilitateur de la sémantique ne se fait pas sentir de la même manière chez les lecteurs compensateurs que chez les contrôles. En effet, les lecteurs compensateurs présentent un effet facilitateur plus clair sur le nombre de réponses correctes, c'est-à-dire sur la précision de l'identification des mots écrits, alors que chez les lecteurs contrôles, l'effet facilitateur est présent uniquement sur les latences, ce qui souligne l'importance de considérer tant les latences que les scores de précision dans les études portant sur les enfants. Cette différence semble refléter le niveau d'expertise de chacun des types de lecteurs; les lecteurs contrôles ont le même patron de résultats que les adultes lecteurs experts. Ceux-ci présentent également un effet d'interférence pour les items non reliés, comme les adultes experts, alors que cet effet d'interférence n'est pas présent chez les lecteurs compensateurs, probablement à cause de leurs difficultés importantes à identifier les mots écrits. Par ailleurs, les résultats montrent que, malgré un SOA court, les lecteurs compensateurs sont tout de même en mesure d'utiliser le lien sémantique entre les mots pour soutenir leur identification de mots déficitaire, et ce, bien que l'on ait certains indices à l'effet que les lecteurs compensateurs n'ont peut-être pas eu le temps de traiter entièrement l'amorce (absence d'effet d'interférence). Ainsi, l'hypothèse mentionnée plus haut a été partiellement vérifiée. L'effet facilitateur n'est pas plus important chez les lecteurs compensateurs que chez les lecteurs contrôles – la seule particularité pour les lecteurs compensateurs est que cet effet se manifeste sur la précision plutôt que sur les latences –, mais qu'il est présent tant à un SOA court qu'à un SOA long.

Les résultats obtenus éclairent également la deuxième hypothèse selon laquelle les lecteurs compensateurs pourraient activer plus tôt l'information sémantique pendant l'identification des mots écrits que ne le font les lecteurs contrôles, et ce, malgré un déficit du traitement de la forme visuelle et

phonologique des mots. Bien que les résultats ne soient qu'exploratoires, il semble qu'il soit possible que la sémantique puisse être mobilisée plus tôt chez les lecteurs compensateurs, en comparaison aux lecteurs contrôles, puisqu'ils sont les seuls à présenter un effet de la sémantique sur la composante P200. Par ailleurs, les latences de la N400 chez les lecteurs compensateurs ne sont pas retardées, comme c'est habituellement le cas dans les études réalisées auprès d'élèves qui présentent des difficultés d'identification des mots écrits. Ceci pourrait également être une particularité des lecteurs compensateurs.

Ces deux études se centrent sur la compensation en lecture chez des élèves du primaire. Il s'agit sans doute de la première étude à vérifier des hypothèses quant au fonctionnement du mécanisme de compensation chez des enfants. Par ailleurs, le fait que ces études portent sur le français constitue une retombée importante, tant pour l'éducation que pour la psycholinguistique. L'échantillon testé dans le cadre de l'étude 1 est conséquent, ce qui a permis d'obtenir des données solides pour ensuite déterminer les seuils pour les critères de sélection des lecteurs compensateurs. Dans l'étude 2, un soin particulier a été apporté à la sélection des participants, et ce, tant dans le groupe de lecteurs compensateurs que dans celui des lecteurs contrôles. Les difficultés de recrutement pour l'étude 2 (nécessitant un déplacement à l'université) ont fait en sorte que certains critères d'inclusion/exclusion n'ont pu être maintenus, mais un soin minutieux a été apporté à la consignation de plusieurs informations aux fins de sélection des participants. Finalement, le recours à l'EEG en plus des données comportementales permet de soulever de nouvelles questions de recherche et permet d'envisager des données additionnelles à des questions pertinentes à l'éducation, ce qui constitue un exemple pertinent pour les neurosciences éducationnelles.

Bien évidemment, ces études comportent plusieurs limites qu'il convient de discuter ici. La principale limite réside sans doute dans le manque de puissance statistique de l'étude 2. Ce manque de puissance peut s'expliquer, d'une part, par le nombre restreint de participants. En effet, le profil de lecteur compensateur est un profil atypique, et correspond donc à une prévalence faible dans la population des élèves du primaire. Le recrutement de ces participants est également rendu difficile par le caractère parfois non manifeste de leurs difficultés à identifier les mots écrits alors que la compréhension écrite est relativement normale Malgré un échantillon conséquent dans l'étude 1, trop peu de lecteurs compensateurs y ont été repérés, il a donc fallu déployer des efforts pour en trouver un nombre suffisant. Le manque de puissance statistique peut aussi s'expliquer par la possibilité que la population des lecteurs compensateurs soit peut-être une population hétérogène. Les écarts-types systématiquement plus grands obtenus chez les lecteurs compensateurs en comparaison aux lecteurs contrôles semblent appuyer ce

phénomène. Peu de résultats (comportementaux et d'électroencéphalographie) dans la tâche avec le délai (SOA) long démontrent des effets significatifs. Il est possible d'envisager que des différences dans les caractéristiques linguistiques des deux listes d'items soient à l'origine de ce phénomène. La mise à l'essai des tâches n'avait pourtant pas révélé cela, mais cela reste à vérifier. Dans les deux études, il aurait été pertinent, voire primordial, de comparer les lecteurs compensateurs à la fois à des lecteurs contrôles de même niveau de compréhension écrite, comme cela a été le cas dans la présente recherche, mais aussi à des lecteurs contrôles de même niveau d'identification des mots écrits. Comme cela a été le cas dans la littérature scientifique sur la dyslexie (voir section 2.7), la considération de l'âge de lecture a été primordiale dans la sélection des contrôles pour faire ressortir les particularités des dyslexiques. De la même façon, il est possible d'envisager que les résultats quant aux différences de performance entre les lecteurs compensateurs et les contrôles de même niveau d'identification des mots écrits puissent révéler d'autres caractéristiques particulières des lecteurs compensateurs. Par ailleurs, il n'est pas possible de savoir si tous les lecteurs compensateurs repérés vont effectivement présenter des difficultés manifestes de lecture plus tard. Il serait intéressant de faire des études longitudinales sur le phénomène afin de documenter cet aspect.

Finalement, malgré les limites de ces deux études, les retombées pour l'éducation sont présentes, surtout en ce qui concerne l'évaluation de la lecture, dans un contexte de dépistage des élèves à risque ou en difficulté, ou d'évaluation orthopédagogique approfondie pour dégager un portrait des forces et lacunes au niveau des différents processus cognitifs impliqués dans la compréhension écrite. Les résultats de l'étude 1 montrent que parmi les 12 élèves repérés comme étant des lecteurs compensateurs, 5 n'étaient pas du tout identifiés comme présentant une difficulté. Le fait que l'on ait tendance à évaluer uniquement la compréhension écrite en classe au-delà de la 3^{e} année du primaire contribue au fait que ces élèves aient réussi à faire en sorte que leurs difficultés passent inaperçues, à cause de leur caractère non apparent en situation de lecture de texte. Ceci confirme l'importance de bien réfléchir aux pratiques d'évaluation de la lecture. Cela implique également qu'il est important, tout le long du primaire, d'intégrer des évaluations des processus spécifiques d'identification des mots écrits pour tous les élèves, particulièrement des mesures de fluence (combinaison des taux de réussite et du temps pris pour faire la tâche) de décodage des pseudomots, même si cela n'est pas indiqué dans le Programme de formation de l'école québécoise (section Progression des apprentissages) au-delà du 1^{e} cycle du primaire. Éventuellement, les recherches futures portant sur ce sujet pourront permettre de proposer des hypothèses quant aux interventions à réaliser auprès des lecteurs compensateurs, mais pas maintenant, par souci de rigueur.

Tel que suggéré par certaines études sur l'effet de contexte, on pourrait proposer que la compensation est une propension naturelle, utilisée par n'importe quel lecteur qui se retrouve dans une situation où ses capacités ou le contexte ne lui permettent pas d'actualiser de manière automatique ses processus d'identification des mots écrits. Ainsi, le lecteur expert adulte compensera si les stimuli sont dégradés ou s'il doit lire un texte dans une langue seconde. Le lecteur en apprentissage compensera également même s'il n'a pas de difficulté particulière en identification des mots écrits dans un contexte où le texte est d'un niveau trop avancé pour lui ou parce qu'il s'agit d'un sujet moins connu. En conséquence, ce ne sont peut-être pas les lecteurs compensateurs qui sont atypiques, mais bien ceux qui n'arrivent pas à compenser. Il pourrait ainsi être intéressant d'étudier ces deux populations (lecteurs en difficulté qui compensent et lecteurs en difficulté qui ne compensent pas) de manière parallèle pour en faire ressortir les similarités et distinctions.

LISTE DE RÉFÉRENCES

Aaron, P. G. (1989). *Dyslexia and hyperlexia*. The Netherlands: Kluwer Academic Publishers.

Aaron, P.G., Joshi, M. et Williams, K.A. (1999). Not all reading disabilities are alike. *Journal of Learning Disabilities*, 32, 120-137.

Ackerman, P. T., Dykman, R. A., et Oglesby, D. M. (1994). Visual event-related potentials of dyslexic children to rhyming and non-rhyming stimuli. *Journal of Clinical and Experimental Neuropsychology*. 16, 138–154. doi: 10.1080/01688639408402624

Adelman, J., Sabatos-DeVito, M., Marquis, S. J., et Estes, Z. (2014). Individual differences in reading aloud: A mega- study, item effects, and some models. *Cognitive Psychology*, 68, 113–160.

Alegria, J. (1988). L'acquisition de la lecture : aspects cognitifs. *Questions de logopédie,* 19(3), 99-122.

Alegria, J. et Mousty, P. (1996). The development of spelling procedures in French-speaking, normal and reading-disabled children: effects of frequency and lexicality. *Journal of Experimental Child Psychology*, *63*(2), 312–338. https://doi.org/10.1006/jecp.1996.0052

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders DSM-5* (5e éd.). Arlington, VA : American Psychiatric Publishing.

Andrews, S. (2008). Lexical expertise and reading skill. *Psychology of Learning and Motivation*, 49, 249-281.

Andrews, S., et Bond, R. (2009). Lexical expertise and reading skill: Bottom-up and top-down processing of lexical ambiguity. *Reading and Writing*, *22*(6), 687–711. https://doi.org/10.1007/s11145-008-9137-7

Association des orthopédagogues du Québec (2003). *L'acte orthopédagogique*, mémoire préparé par l'Association des orthopédagogues du Québec, Montréal.

Association des orthopédagogues du Québec (2018). *Le référentiel des compétences professionnelles liées à l'exercice de l'orthopédagogue au Québec*. Récupéré de https://www.ladoq.ca/

Baars, B.J. et Gage, N.M. (2013). *Fundamentals of Cognitive Neuroscience : A beginner's guide*. Oxford: Academic Press.

Bakos S, Landerl K, Bartling J, Schulte-Körne G, Moll K. (2018). Neurophysiological correlates of word processing deficits in isolated reading and isolated spelling disorders. Clinical Neurophysiology. 129(3), 526–40. https://doi.org/10.1016/

Barber, H. A., et Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*(1), 98–123. https://doi.org/10.1016/j.brainresrev.2006.07.002

Beauchamp, C. et Beauchamp, M. H. (2013). Boundary as Bridge: An Analysis of the Educational Neuroscience Literature from a Boundary Perspective. *Educational Psychology Review*, *25*(1), 47–67. https://doi.org/10.1007/s10648-012-9207-x

Becker, C. A. (1982). The Development of Semantic Context Effects : Two Processes or Two Strategies ? *Reading Research Quartely*, *17*(4), 482–502.

Becker, C.A. (1980) Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition*, 8, 493-512.

Bentin, S., McCarthy, G., et Wood, C. C. (1985). Event-related potentials, lexical decision, and semantic priming. Electroencephalography. *Clinical Neurophysiology*, 60, 353–355.

Benuck, M. B. et Peverly, S. T. (2004). The effect of orthographic depth on reliance upon semantic context for oral reading in English and Hebrew. *Journal of Research in Reading*, *27*(3), 281–299. https://doi.org/10.1111/j.1467-9817.2004.00232.x

Bermúdez-Margaretto, B., Beltrán, D., Shtyrov, Y., Dominguez, A., et Cuetos, F. (2020). Neurophysiological correlates of top-down phonological and semantic influence during the orthographic processing of novel visual word-forms. *Brain Sciences*, *10*(10), 1–19. https://doi.org/10.3390/brainsci10100717

Berninger, V. W. et Abbott, R. D. (2013). Differences between children with dyslexia who are and are not gifted in verbal reasoning. *Gifted Child Quarterly*, *57*(4), 223–233. https://doi.org/10.1177/0016986213500342

Betjemann, R. S., et Keenan, J. M. (2008). Phonological and semantic priming in children with reading disability. *Child Development*, *79*(4), 1086–1102. https://doi.org/10.1111/j.1467-8624.2008.01177.x

Bianco, M., Lima, L, Megherbi, H., Nardy, A., Rocher, T. et Colé, P. (2018). Évaluation diagnostique de la compréhension en lecture – ANR DEVCOMP. Manuscrit non publié.

Binder, K. S., & Lee, C. S. (2012). Reader Profiles for Adults with Low Literacy Skills: A Quest to Find Resilient Readers. *Journal of Research and Practice for Adult Literacy, Secondary, and Basic Education*, 1(2), 78–90.

Bonin, P., Méot, A., Aubert, L., Malardier, N., Niedenthal, P. et Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. L'année psychologique, 103-4, 655-694.

Bonnotte, I. et Casalis, S. (2010). Semantic priming in French children with varying comprehension skills. *European Journal of Developmental Psychology*, *7*(3), 309–328. https://doi.org/10.1080/17405620802114546

Bowers, J. S. (2016). Psychology, Not Educational Neuroscience, Is the Way Forward for Improving Educational Outcomes for All Children: Reply to Gabrieli (2016) and Howard-Jones *et al.* (2016). *Psychological Review*, *123*(5), 628–635. https://doi.org/10.1037/rev0000043

Braibant, J.-M. (1994). Le décodage et la compréhension: deux composantes essentielles de la lecture en 2e primaire. Dans J. Grégoire et B. Piérart (dir.), *Évaluer les troubles de la lecture* (p. 173–194). De Boeck Supérieur. https://doi.org/10.3917/dbu.grego.1994.01.0173

Briggs, P., Austin, S. et Underwood, G. (1984). The Effects of Sentence Context in Good and Poor Readers : A Test of Stanovich's Interactive-Compensatory Model A. *Reading Research Quarterlyly*, *20*(1), 54–61.

Brousseau, A.-M. et Roberge, Y. (2000). *Syntaxe et sémantique du français*. Québec : Fides.

Bruck, M. (1992). Persistence of Dyslexics' Phonological Awareness Deficits. *Developmental Psychology*, *28*(5), 874–886. https://doi.org/10.1037/0012-1649.28.5.874

Caillot, M. (1992). Vers une didactique cognitive? *Intellectica*, *1/2*(13–14), 273–289.

Caravolas, M. (2018). Growth of Word and Pseudoword Reading Efficiency in Alphabetic Orthographies: Impact of Consistency. *Journal of Learning Disabilities*, *51*(5), 422–433. https://doi.org/10.1177/0022219417718197

Caravolas, M. et Samara (2015). Learning to read and spell words in different writing systems. Dans A. Pollatsek et R. Treiman (dir.), *The Oxford handbook of reading* (p. 326– 343). New York, NY: Oxford University Press.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., …, Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, 23, 678– 686. doi:10.1177/0956797611434536

Catts, H. W. et Kamhi, A. (2005). *Language and reading disabilities*. Boston: Pearson.

Catts, H. W., Adlof, S. M. et Weismer, S. E. (2006). Language Deficits in Poor Comprehenders: A Case for the Simple View of Reading. *Journal of Speech Language and Hearing Research*, *49*(2), 278. https://doi.org/10.1044/1092-4388(2006/023)

Catts, H. W., Compton, D., Tomblin, J. B. et Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166–181. https://doi.org/10.1037/a0025323

Catts, H. W., Hogan, T. P. et Adlof, S. M. (2005). Developing changes in reading and reading disabilities. Dans H. W. Catts et A. G. Kahmi (dir.), *The connections between language and reading disabilities* (p. 50-71). Mahwah, NJ : Lawrence Erlbaum Associates.

Cavalli, E., Colé, P., Badier, J.-M., Zielinski, C., Chanoine, V. et Ziegler, J. C. (2016). Spatiotemporal Dynamics of Morphological Processing in Visual Word Recognition. *Journal of Cognitive Neuroscience*, *28*(8), 1228–1242. https://doi.org/10.1162/jocn

Chartrand, S.G., Aubin, D, Blain, R. et Simard, C. (1999). *Grammaire pédagogique du français d'aujourd'hui*. Graficor.

Chevrier-Muller, C., Fournier, S., Maillart, C. et Simon, A.-M. (2010). *La Batterie Langage oral, langage écrit, mémoire, attention* – 2e édition (L2MA2). Pearson.

Coch, D. et Ansari, D. (2009). Thinking about mechanisms is crucial to connecting neuroscience and education. *Cortex*, *45*(4), 546–547. https://doi.org/10.1016/j.cortex.2008.06.001

Cohen, H. et Lefebvre, C. (dir.) (2005). *Handbook of categorization in cognitive science*. Oxford: Elsevier.

Collins, A.M. et Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 407-428.

Collins, A.M. et Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Learning and Verbal Behavior*, 8, 240-247.

Collins, T., Andler D., et Tallon-Baudry, C. (2018). *La cognition : Du neurone à société*. Gallimard.

Coltheart, M. (1978). Lexical access in simple reading tasks. Dans G. Underwood (dir.), *Strategies of information processing*, p. 151-216. New York: Academic Press.

Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. Dans M. J. Snowling et C. Hulme (dir.), *The science of reading: A handbook* (p. 6–23). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch1

Coltheart, M., Patterson, K.E et Marshall, J.C. (1980). *Deep Dyslexia*. London: Routhledge.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. et Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Conners, F. et Olson, R. (1990). Comprehension processes in reading. Dans D. A. Balota, G. B. Flores d'Arcais et K. Rayner (dir.), *Reading comprehension in dyslexic and normal readers: A component skills analysis* (p. 557–579). Hillsdale, NJ: Lawrence Erlbaum.

Content, A. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological research*, 53, 3-12.

Costanzo, F., Menghini, D., Caltagirone, C., Oliveri, M. et Vicari, S. (2013). How to improve reading skills in dyslexics: The effect of high frequency rTMS. *Neuropsychologia*, *51*(14), 2953–2959. https://doi.org/10.1016/j.neuropsychologia.2013.10.018

Creswell, J. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: SAGE Publications.

Cromley, J. G. et Azevedo, R. (2007). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, *1*(3), 229–247. https://doi.org/10.1007/s11409-006-9002-5

De Groot, A.M.B., Thomassen, A.J.W.M. et Hudson, P.T.W. (1986). Primed-lexical decision : The effect of varying the stimulus-onset asynchrony of prime and target. *Acta Psychologica*, 61, 17-36.

De Jong, P. F. et van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, 95(1), 22–40. doi:10.1037/0022-0663.95.1.22

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.

Debruille, J. B. (2007). The N400 potential could index a semantic inhibition. *Brain Research Reviews*, 56, 472–477.

Delahaie, M., Sprenger-Charolles, L., et Serniclaes, W. (2007). Effet de la lexicalité chez des faibles et très faibles lecteurs comparativement à des normolecteurs de même âge et de même niveau de lecture. *Annee Psychologique*, *107*(3), 361–396. https://doi.org/10.4074/S0003503307003028

Demont, E. et Gombert, J. E. (1996). Phonological awareness as a pre- dictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology*, 66, 315–332.

Desrochers, A. (2009). *Apprendre à lire à l'âge adulte*. Sudbury, Ontario : Éditions du Centre FORA.

Desrochers, A., Desgagné, L. et Kirby, J. R. (2011). L'évaluation de la lecture orale. DansM. J. Berger et A. Desrochers (dir.), *L'évaluation de la littératie* (p. 177–214). Ottawa: Les Presses de l'Université d'Ottawa.

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: A model of converging pathways. *Biological Psychology*, *80*(1), 10–22. https://doi.org/10.1016/j.biopsycho.2008.04.013

Ducrot, O. et Schaeffer, J.-M. (1995). Nou*veau dictionnaire encyclopédique des sciences du langage*. Paris : Éditions du Seuil.

Duncan, L. G. et Seymour, P. H. K. (2000). Socio-economic differences in foundation-level literacy. *British Journal of Psychology*, *91*(2), 145–166.

Dunn, D., Lloyd, M., Leota, M. et Theriault-Whalen, C.M. (1993). *Échelle de vocabulaire en images Peabody (EVIP)*. Toronto : Pearson.

Durante, R., et Hirshman, E. (1994). Retrospective priming and masked semantic priming: The interfering effects of prime activation. Journal of Memory and Language, 33, 112-127.

Ecalle, J. et Magnan, A. (2010). *L'apprentissage de la lecture et ses difficultés*. Dunod.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. Dans C. A. Perfetti, L. Rieben et M. Fayol (dir.), *Learning to spell : Research, theory, and practice across languages* (p. 237-269). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (2001). *Reports of the Subgroups*. National Reading Panel.

Ehri, L. C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2007). Development of sight word reading: Phases and findings. Dans M. J. Snowling et C. Hulme (dir.), *The science of reading* (p. 135-154). Oxford, Royaume-Uni: Blackwell

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Ehri, L. C., Nunes, S. R., Stahl, S. A. et Willows, D. (2001). Systematic phonies instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71, 393-447.

Faísca, L., Reis, A., et Araújo, S. (2019). Early brain sensitivity to word frequency and lexicality during reading aloud and implicit reading. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00830

Fayol, M. et Jaffré, J.-P. (2008). *Orthographier*. Paris: Presses universitaires de France.

Ferrand, L. (2007). *Psychologie cognitive de la lecture*. Bruxelles: Éditions De Boeck.

Ferrand, L. et Alario, F.-X. (1998). Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année psychologique*, 98, 659-709.

Ferrand, L. et Grainger, J. (2004). *Psycholinguistique cognitive. Essais en l'honneur de Juan Segui*. Bruxelles : De Boeck Université.

Ferrand, L., et New, B. (2003). Syllabic length effects in visual word recognition and naming. Acta Psychologica, 113, 167–183.Fletcher *et al.*, 1994;

Fodor,J. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.

Forster, K. I., et Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments et Computers*, 35(1), 116–124. https://doi.org/10.3758/BF03195503

Fox, M. C., Ericsson, K. A. et Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Fraenkel, J.R. et Wallen, N.E. (2000). *How to design and evaluate research in education* (4e éd.). United States of America : The McGraw-Hill Companies.

Francis, D. J., Fletcher, J. M., Catts, H. W. et Tomblin, B. (2005). Dimensions affecting the assessment of reading comprehension. Dans S. G. Paris et S. A. Stahl (dir.), *Current issues in reading comprehension and assessment* (p. 369-394). Mahwah, NJ : Erlbaum.

Frederiksen, J. R. et Warren, B. M. (1987). A cognitive framework for developing expertise in reading. Dans R. Glaser (Dir.), *Advances in instructional psychology* (p. 1–39). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frith, U. (1985). Beneath the surface of developmental dyslexia. Dans K. E. Patterson,]. C. Marshall et M. Coltheart (dir.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (p. 301-330). Londres: Erlbaum.

Frith, U. (1986). A Developmental Framework for Developmental Dyslexia. *Annals of Dyslexia*, *36*, 69–81.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Frost, R., Katz, L. et Bentin, S. (1987). Strategies for visual word recognition and orthographic depth. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 104-115.

Gabrieli, J. D. E. (2009). Dyslexia: a new synergy between education and cognitive neuroscience. *Science*, *325*(July), 280–283. https://doi.org/10.1126/science.1171999

Goodman, K. S. (1967). Reading: a psycholinguistic guessing game. *Journal of Reading Specialist*.

Goodman, K.S. (1976). Reading: A Psycholinguistic Guessing Game. Dans H., Singer et R.B. Rudell, (dir.), *Theoretical Models and Processes of Reading* (p. 497–508). International Reading Association.

Goswami, U. (2008). Reading, dyslexia and the brain. *Educational Research*, *50*(2), 135–148. https://doi.org/10.1080/00131880802082625

Goswami, U., Gombert, J. E. et Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French and Spanish. *Applied Psycholinguistics*, 19, 19-52.

Gough, P. B. et Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Grainger, J. et Ferrand, L. (1994). Phonology and orthography in visual word recognition : Effects of masked homophone primes. *Journal of Memory and Language*, 33, 218-233.

Grainger, J. et Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*(April), 54. https://doi.org/10.3389/fpsyg.2011.00054

Grainger, J., et Holcomb, P. J. (2009). Watching the Word Go by: On the Time-course of Component Processes inVisual Word Recognition. *Language and Linguistics Compass*, *3*(1), 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x.Watching

Gregory, H. (1999). *Semantics : an introductory workbook*. New York and London : Routledge.

Haager, D., Klingner, J. et Vaughn, S. (2007). *Evidence-based reading practices for response to intervention*. Baltimore: Brookes, MA.

Haberlandt, K. (1994). Methods in reading research. Dans M.A. Gernsbacher (dir.), *Handbook of Psycholinguistics* (p. 1-31). San Diego: Academic Press.

Hahne, A., Eckstein, K., et Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, 16(7), 1302–1318.

Hauk, O., Coutout, C., Holden, et Chen, Y. (2012). The time-course of single-word reading: evidence from fast behavioral and brain responses. *NeuroImage*, *60*(2), 1462–77. https://doi.org/10.1016/j.neuroimage.2012.01.061

Herrmann, J. A., Matyas, T., et Pratt, C. (2006). Meta-analysis of the nonword-reading deficit in specific reading disorder. *Dyslexia*, 12, 195–221. http://dx.doi.org/10.1002/dys.324 Kessler,

Hillyard SA, Kutas M. (1983). Electrophysiology of cognitive processing. Annual Review of Psychology. 34:33– 61.

Hoover, W. A. et Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2, 127–160.

Howard-Jones, P. A. (2011). A multiperspective approach to neuroeducational research. *Educational Philosophy and Theory*, *43*(1), 24–30. https://doi.org/10.1111/j.1469-5812.2010.00703.x

Howard-Jones, P. A. et Fenton, K. D. (2012). The need for interdisciplinary dialogue in developing ethical approaches to neuroeducational research. *Neuroethics*, *5*(2), 119–134. https://doi.org/10.1007/s12152-011-9101-0

Huey, E.B. (1908). *The Psychology and Pedagogy of Reading*. Reproduit par: Cambridge, Mass.: M.I.T. Press.

Huot, H. (2005). *La morphologie : forme et sens des mots du français*. Dans M. Perret (Dir.) Paris : Armand Colin.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Inserm (2007). Dyslexie Dysorthographie Dyscalculie. *Institut National de La Santé et de La Recherche Médicale INSERM*, *43*(104), 1350–1360. https://doi.org/10.1177/0091270003258666

Jackson, N. E., et Doellinger, H. L. (2002). Resilient readers? University students who are poor recoders but sometimes good text comprehenders. *Journal of Educational Psychology*, *94*(1), 64–78. https://doi.org/10.1037/0022-0663.94.1.64

Janosz, M., Pascal, S., Belleau, L., Archambault, I. et Parent, S. (2013). Les élèves du primaire à risque de décrocher au secondaire : caractéristiques à 12 ans et prédicteurs à 7 ans. *Institut de La Statistique Du Québec*, *7*, 24.

Jucla, M., Nenert, R., Chaix, Y., et Demonet, J. F. (2009). Remediation effects on N170 and P300 in children with developmental dyslexia. *Behavioural Neurology*, *22*(3–4), 121–129. https://doi.org/10.3233/BEN-2009-0257

Kazdin, A.E. (2003). *Research design in clinical psychology* (4e éd.). Boston: Allyn & Beacon.

Kieffer, M. J., Vukovic, R. K. et Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly*, 48(4), 333-348. doi:10.1002/rrq.54

Kintsch, W. (1988). the Role of Knowledge in Discourse Comprehension - a Construction Integration Model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037//0033-295X.95.2.163

Kintsch, W. et Rawson, K. A. (2005). Comprehension. Dans M. Snowling (dir.), *The science of reading: A handbook* (p. 209–226).

Kirby, J. R. et Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, *42*(2), 75–82. https://doi.org/10.1111/j.1741-4369.2008.00487.x

Knox, R. (2016). A Transdisciplinary Field. *Mbe*, *10*(1), 4–9. https://doi.org/10.1111/mbe.12102

Kutas, M. et Hillyard, S.A. (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 207:203–5.

Kutas, M. et Hillyard, S.A. (1984). Brain potentials during reading reflect word expentancy and semantic association. *Nature*, 307, 161-163.

Kutas, M., et Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.Thirty

LaBerge, D. et Samuels, J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.

Landerl, K. et Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, 100(1), 150–161. doi:10.1037/0022-0663.100.1.150

Landi, N., et Perfetti, C. a. (2006). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language*, *102*(1), 30–45. https://doi.org/10.1016/j.bandl.2006.11.001

Landi, N., Perfetti, C. a, Bolger, D. J., Dunlap, S. et Foorman, B. R. (2006). The role of discourse context in developing word form representations: a paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, *94*(2), 114–33. https://doi.org/10.1016/j.jecp.2005.12.004

Laplante, L. (1998). *Dyslexie développementale et système de reconnaissance des mots écrits* [Thèse de doctorat, Université de Montréal].

Laplante, L. (2011). L'évaluation diagnostique des difficultés d'apprentissage de la lecture. Dans M. J. Berger et A. Desrochers (dir.), *L'évaluation de la littératie* (p. 139–174). Ottawa: Les Presses de l'Université d'Ottawa.

Laplante, L. (2012). L'historique de l'orthopédagogie au Québec. *Vie Pédagogioque*, *160*(Février), 10–14.

Leach, J. M., Scarborough, H. S. et Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211–224.

Lee, H., Lee, Y., Tae, J. et Kwon, Y. (2019). Advantage of the go/no-go task over the yes/no lexical decision task: ERP indexes of parameters in the diffusion model. *PloS one*, *14*(7).

Leloup, G (2011). Lecture et Compétences reliées à la lecture chez des adultes dyslexiques universitaires [Mémoire, Université de Paris].

Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, and Computers*, 36, 156-166.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*, *7*(4), 618–630. https://doi.org/10.3758/BF03212999

Luck, S.J. (2005). *An Introduction to the Event-related Potentials Technique*. Cambridge: The MIT Press.

Lyon, G. R., Shaywitz, S. E., et Shaywitz, B. A. (2003). Defining Dyslexia. *Annals of Dyslexia*, *53*(1), 1–14.

Maeder, C. (2010). *La forme noire : Test de compréhension écrite de récits de 9 à 12 ans*. Paris : Ortho Éditions.

Marcel, A.J. (1983). Conscious and unconscious perception : Experiment on visual masking and word recognition. Cognitive Psychology, 15, 197-237.

Martín-Loeches, M., Hinojosa, J. A., Gómez-Jarabo, G., et Rubia, F. J. (2001). An early electrophysiological sign of semantic processing in basal extrastriate areas. *Psychophysiology*, *38*(1), 114–124. https://doi.org/10.1017/S004857720199105X

McCauly, C., Weil, C. M., et Sperber, R. S. (1976). The development of memory structure as reflected by semantic-priming effects. *Journal of Experimental Child Psychology*, 22, 511– 518.

McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287-300.

McGill-Franzen, A. et Allington, R. L. (1991). The gridlock of low-achievement: Perspectives on policy and practice. *Remedial and Special Education*, 12(3), 20-30.

McNamara, T. P. (2005). *Semantic Priming : Perspectives from Memory and Word Recognition*. New York, NY: Taylor et Francis Group.

Mehlhase, H., Bakos, S., Bartling, J., Schulte-Körne, G. et Moll, K. (2020). Word processing deficits in children with isolated and combined reading and spelling deficits: An ERP-study. *Brain Research*. https://doi.org/10.1016/j.brainres

Mercier, J. et Charland, P. (2013). An agenda for neuroeducation: relating psychophysiological and behavioral data across time scales of learning. *Neuroéducation*, *2*(1), 71–86.

Mercier, J., Bédard, M. et Laplante, L. (2016). Les neurosciences éducationnelles et les TA : faut-il s'intéresser au cerveau pour mieux intervenir? *Le Rendez-vous* (revue de l'Institut des troubles d'apprentissage/ITA), 30-33.

Meyer, D.E. et Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words : Evidence of a dependance between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.

Mialaret, G. (2017). *Les sciences de l'éducation*. 12e édition, Paris : PUF.

Milne, D. (2005). *Teaching the brain to read*. Artarmon : SK Publishing. Récupéré de : https://www.dyslexia-international.org/content/Articles/Teaching%20the%20brain_EN.pdf

Mimouni, Z. et King, L. (2007). *Troubles de lecture au collégial : deux mesures de soutien*. Rapport de recherche PAREA :Montréal. 80p.

Ministère de l'Éducation et de l'Enseignement supérieur (2020). Guide pour la mise en œuvre de la réponse à l'intervention dans le milieu scolaire, Québec, Ministère de l'Éducation et de l'Enseignement supérieur. https://recit.qc.ca/nouvelle/conference-de-consensus-reponse-a-lintervention-rai/

Ministère de l'Éducation et de l'enseignement supérieur (MEES) (2017). Politique de la réussite éducative : Le plaisir d'apprendre, la chance de réussir. Récupéré de http://www.education.gouv.qc.ca/

Ministère de l'Éducation et de l'enseignement supérieur (MEES) (2016). Document de consultation *Pour une politique de la réussite éducative*. Récupéré de http://www.education.gouv.qc.ca/

Ministère de l'Éducation, du Loisir et du Sport (MELS) (1999). *Politique de l'adaptation scolaire*. Récupéré de http://www.mels.gouv.qc.ca/fileadmin/site_web/documents/publications/EPEPS/Formation_jeunes/Adaptation_scolaire/politi00F.pdf/

Morais, J. (1994). Perception et traitement du langage écrit. Dans M. Richelle, J. Requin et M. Robert (dir.), *Traité de psychologie expérimentale* (p. 271-331). Paris : PUF.

Morais, J. (1999). *L'art de lire*. Paris : Odile Jacob. Office québécois de la langue française. Récupéré de http://gdt.oqlf.gouv.qc.ca/

Morais, J., Pierre, R. et Kolinsky, R. (2003). Du lecteur compétent au lecteur débutant: implications des recherches en psycholinguistique cognitive et en neuropsychologie pour l'enseignement de la lecture. *Revue Des Sciences de l'?Ducation*, *29*(1), 51. https://doi.org/10.7202/009492ar

Moret-Tatay, C., et Perea, M. (2011). Is the go/no-go lexical decision task preferable to the yes/no task with developing readers? *Journal of experimental child psychology*, *110*(1), 125-132.

Mosenthal, P., Walmsley, S. et Allington, R. (1978). Word Recognition Reconsidered : Toward a Multi-Context Model. *Visible Language*, *12*(4), 448–468.

Mousty, P, Leybaert, J., Alegria, J. Content, A. et Morais, J. (1994) BELEC : une batterie d'évaluation du langage écrit et de ses troubles. Dans J. Grégoire et B. Piérart (dir.) *Evaluer les troubles de la lecture : les nouveaux modèles théoriques et leurs implications diagnostiques*, p. 127-145. Bruxelles : De Boeck.

Nation, K. (2007). Children's reading comprehension difficulties. Dans M. J. Snowling et C. Hulme (dir.), *The science of reading. A handbook* (p. 248-265). Malden, États-Unis: Blackwell.

Nation, K. et Snowling, M. J. (1998). Individual differences in contextual facilitation: evidence from dyslexia and poor reading comprehension. *Child Development*, *69*(4), 996–1011. Récupéré de http://www.ncbi.nlm.nih.gov/pubmed/9768483

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Role of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology*, 106, 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recogni- tion:A selective review of current findings and theories. Dans D. Besner et G. W. Humphreys (dir.), *Basic processes in reading: Visual word recognition* (p. 264-336). Hillsdale, NJ: Erlbaum.

Neely, J.H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, 4, 646-654.

New B., Pallier C., Ferrand L. et Matos R. (2001) Une base de données lexicales du français contemporain sur internet: LEXIQUE, L'Année Psychologique, 101, 447-462.

New, B., Pallier, C., Brysbaert, M. et Ferrand, L. (2004) Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, and Computers*, 36 (3), 516-524

Newell, A. (1990). *Unified Theories of Cognition*. Boston: Harvard University Press.

Nievas, F. et Justicia, F. (2004). A cross-sectional study about meaning access processes for homographs. *Cognitive Development*, 19, 95–109.

Ouellette, G. et Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23*(2), 189–208. https://doi.org/10.1007/s11145-008-9159-1

Peereman R, Lété B, Sprenger-Charolles L. Manulex-infra: Distributional characteristics of infra-lexical and lexical units in child-directed written material. *Behavior Research Methods, Instruments and Computers*. 2007; 39: 593–603.

Perea, M., Rosa, E., et Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory and cognition*, *30*(1), 34-45.

Perfetti, C. (1985). *Reading ability*. New York : Oxford University Press.

Perfetti, C. A. (1999). Comprehending written langage: A blueprint of the reader. Dans C. M. Brown et P. Hagoort (dir.), *The neurocognition of language* (p. 167–208). Oxford University Press.

Perfetti, C. A. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C. A., Goldman, S. R. et Hogaboeam, T.W. (1979). Reading skill and identification of words in discourse context. *Memory & Cognition*, 7(4), 273-282.

Perfetti, C. et Hart, L. (2002). The lexical quality hypothesis. Dans L. Verhoeven, C. Elbro et P. Reitsma (dir.), *Precursors of Functional Literacy* (p. 189–213). Amsterdam: John Benjamins Publishing Company.

Perfetti, C. et Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Perfetti, C.A. (1988). Verbal efficiency theory in reading ability. Dans M. Daneman, G.E. MacKinnon, et T.G. Waller (dir.), *Reading research: Advances in theory and practice* (p. 109-143). New York: Academic Press.

Perry, C., Ziegler, J. C. et Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Petit, J.P., Midgley, K.J., Holcomb, P.J., Grainger, J., 2006. On the time-course of letter perception: a masked priming ERP investigation. Psychonomic Bulletin and Review, 13 (4), 674–681.

Pinker, S. (1994). *L'instinct du langage*. Paris : Odile Jacob.

Plaut, D. C. et Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*(4), 786–823. https://doi.org/10.1037/0033-295X.107.4.786

Plaut, D. C., McClelland, J. L., Seidenberg, M. S. et Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. Récupéré de http://www.ncbi.nlm.nih.gov/pubmed/8650300

Politis, M. (2007). L'apport de la psychologie cognitive à la didactique de la traduction. *Meta*, *52*(1), 156–163.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, N.J.: Erlbaum.

Posner. M. I., et Snyder, C. R. R. (1975). Attention and cognitive control. In R.L. Solso (dir). Information processing and cognition: The Loyola symposium. Hillsdale : Erlbaum.

Powell, D., Plaut, D. et Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child? *Journal of Research in Reading*, *29*(2), 229–250. https://doi.org/10.1111/j.1467-9817.2006.00300.x

Préfontaine, R. R., et Préfontaine, G. C. (1968). *Échelle du vocabulaire oral des enfants de 5 à 8 ans au Canada français*. Montréal: Beauchemin

Prud'homme, J. (2018). *Instruire, corriger, guérir? Les orthopédagogues, l'adaptation scolaire et les difficultés d'apprentussage au Québec 1950-2017*. Québec, Presses de l'Université du Québec.

Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., … Gore, J. C. (2000). The angular gyrus in developmental dyslexia: Task-specific differences in functional connectivity within posterior cortex. *Psychological Science*, *11*(1), 51–56. https://doi.org/Doi 10.1111/1467-9280.00214

Pulvermüller, F., Shtyrov, Y., et Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, *110*(2), 81–94. https://doi.org/10.1016/j.bandl.2008.12.001

Rack, J. P., Snowling, M. J., et Olson, R. (1992). The nonword reading deficit in developmental dyslexia: A review. Reading Research Quar- terly, 27, 29–53.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., et Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865. https://doi.org/10.1093/brain/awg076

Rayner, K., (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124 (3), 372–422.

Ricketts, J., Davies, R., Masterson, J., Stuart, M. et Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *Journal of Experimental Child Psychology*, *150*, 330–345. https://doi.org/10.1016/j.jecp.2016.05.013

Rodrigue, A. (2006). *Étude des représentations orthographiques chez deux types de scripteurs en trouble spécifique d'acquisition du langage* [Mémoire de maîtrise, Université du Québec à Montréal].

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology*, 104, 192-233.

Rosch, E. (1978). Principles of categorization. Dans E. Rosh et B.B. Lloyd (dir.), *Cognition and Categorization* (p. 27-48). Hillsdale: Erlbaum.

Rumelhart, D. E. (1977). Toward an interactive model of reading. Dans S. Dornic (Dir.), *Attention and performance*, (p. 573-603). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.

Rumelhart, D.E. (1994). Toward an Interactive Model of Reading. Dans R.B. Ruddell et M.R. Ruddell (Dir.), Theo*retical Models and Processes of Reading*, 4e édition, Newark : International Reading Association.

Schulz, E., Maurer, U., van der Mark, S., Bucher, K., Brem, S., Martin, E., et Brandeis, D. (2009). Reading for meaning in dyslexic and young children: Distinct neural pathways but common endpoints. *Neuropsychologia*, *47*(12), 2544–2557. https://doi.org/10.1016/j.neuropsychologia.2009.04.028

Schulz, E., Maurer, U., van der Mark, S., Bucher, K., Brem, S., Martin, E., et Brandeis, D. (2008). Impaired semantic processing during sentence reading in children with dyslexia: Combined fMRI and ERP evidence. *NeuroImage*, *41*(1), 153–168. https://doi.org/10.1016/j.neuroimage.2008.02.012

Schvaneveldt, R., Ackerman, B. P. et Semlear, T. (1977). The Effect of Semantic Context on Children's Word Recognition. *Child Development*, 48, 612–616.

Schwantes, F. M. (1981). Effect of Story Context on Children's Ongoing Word Recognition. *Journal of Literacy Research*, *13*(4), 305–311. https://doi.org/10.1080/10862968109547419

Schwantes, F. M. (1982). Text readability level and developmental differences in context effects. *Journal of Literacy Research*, *14*(1), 5–12. https://doi.org/10.1080/10862968209547431

Schwantes, F. M., Boesl, S. L. et Ritz, E. G. (1980). Children's Use of Context in Word Recognition : A Psycholinguistic Guessing Game. *Child Development*, *51*(3), 730–736.

Segalowitz, S. J. et Zheng, X. (2009). An ERP study of category priming: Evidence of early lexical semantic access. *Biological Psychology*, *80*(1), 122–129. https://doi.org/10.1016/j.biopsycho.2008.04.009

Seidenberg, M. S. et McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Semel, E., Wiig, E. H. et Secord, W. (2003). Clinical evaluation of language fundamentals (4th ed.). San Antonio, TX: Harcourt Assessment

Seymour, P. H. K. (1986). *A cognitive analysis of dyslexia*. Londres: Routhledge et Kegan Paul.

Seymour, P. H. K. (1997). Les fondations du développement orthographique et morphographique. Dans C. A. Perfetti, L. Rieben et M. Fayol (dir.), *Des orthographes et leur acquisition* (p. 385–403). Delachaux et Niestlé.

Seymour, P. H. K. (2008). Continuity and discontinuity in the development of single-word reading: Theoretical speculations. Dans E. L. Grigorenko et A. J. Naples (dir.), *Single-word reading: Behavioral and biological perspective* (p. 1–24). New York: Erlbaum.

Seymour, P. H. K., Aro, M. et Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Seymour, P. H. K., et Evans, H. M. (1999). Foundation-Level Dyslexia Assessment and Treatment. *Journal of Learning Disabilities*, *32*(5), 394–406.

Shankweiler,D. *et al.* (1999). Comprehension and decoding : Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, 3(1), 69-94.

Share; David L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, *55*(2), 126–151. https://doi.org/10.1016/0010-0277(94)00645-2

Siegel. L.S, et Faux. D. (1989). Acquisition of certain grapheme-phoneme correspondences in normally achieving and disabled readers. *Reading and Writing: An Interdisciplinary Journal*, 1, 37-52.

Silva-Pereyra, J., Rivera-Gaxiola, M., Aubert, E., Bosch, J., Galan, L. et Salazar, A. (2003). N400 during lexical decision tasks: a current source localization study. *Clinical Neurophysiology*. 114 (12), 2469–2486.

Simpson, G. B. et Lorsbach, T. C. (1983). The development of automatic and conscious components of contextual facilitation. *Child Development*, 54, 760–772.

Simpson, G. B., et Foster, M. R. (1986). Lexical Ambiguity and Children's Word Recognition. *Developmental Psychology*, *22*(2), 147–154. https://doi.org/10.1037/0012-1649

Simpson, G., Lorsbach, T. et Whitehouse, D. (1983). Encoding and contextual components of word recognition in good and poor readers. *Journal of Experimental Child Psychology*, 35, 161-171.

Smith, E.E., Shoben, E.J. et Ripps, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.

Smith, F. (1999). Why Systematic Phonics and Phonemic Awareness Instruction Constitute an Educational Hazard. *Language Arts*, 77(2), 150-161.

Sprenger-Charolles, L. (1989). L'apprentissage de la lecture et ses difficultés approche psycho-linguistique. *Revue Française de Pédagogie*, *87*(avril-mai-juin), 77–106.

Sprenger-Charolles, L. (2012). L'apprentissage de la lecture dans les écritures alphabétiques. *L'Information Grammaticale*, *133*, 15–21.

Sprenger-Charolles, L. (2016). L ’apprentissage de la lecture (du comportement aux corrélats neuronaux) : un bilan de 30 ans de recherche. *Pratiques*, 169–170.

Sprenger-Charolles, L., et Colé, P. (2013). L’apprentissage typique de la lecture. Dans *Lecture et dyslexie* 53–136.Sprenger-Charolles, L., Colé, P. et Serniclaes, W. (2006). *Reading acquisition and developmental dyslexia*. Hove : Psychology Press.

Sprenger-Charolles, L., Colé, P., Piquard-Kipffer, A. et Leloup, G. (2010). *EVALEC, Batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture*. Paris : Ortho Éditions.

Sprenger-Charolles, L. et Leloup, G. (2013). Lecture à haute voix de mots familiers et de mots inventés : Test 1 minute. Manuscrit non publié.

St-Pierre, M. (2012). *Ce que d’autres ont réussi, on peut le réussir aussi*. Communication présentée au 37e congrès annuel de l’AQETA.

Stanovich, K. E. (1980). Toward an Interactive-Compensatory Model of Individual Differences in the Development of Reading Fluency. *Reading Psychology*, *16*(1), 32–71.

Stanovich, K. E. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. *Remedial and Special Education*, *5*(3), 11–19. https://doi.org/10.1177/074193258400500306

Stanovich, K. E. (1988). Explaining the Differences Between the Dyslexic and the Garden-Variety Poor Reader: The Phonological-Core Variable-Difference Model. *Journal of Learning Disabilities*, *21*(10), 590–604.

Stanovich, K. E., Nathan, R. G., West, R. F. et Vala-Rossi, M. (1985). Children's word recognition in context: Spreading activation, expectancy, and modularity. *Child Development*, 56, 1418-1428.

Stanovich, K., West, R. et Feeman, D. (1981). A longitudinal study of sentence context effects in second-grade children: Tests of an interactive-compensatory model. *Journal of Experimental Child Psychology*, 32, 185-199.

Stanovich, K.E. (2000). *Progress in Undrstanding Reading*. New York: Guilford Press.

Stuart, M., Stainthorp, R. et Snowling, M. (2008). Literacy as a complex activity: deconstructing the simple view of reading. *Literacy*, *42*(2), 59–66. https://doi.org/10.1111/j.1741-4369.2008.00490.x

Table interuniversitaire de formation en orthopédagogie, TIFO (2015) *Référentiel de compétences pour une maîtrise professionnelle en orthopédagogie*.

Taylor, J. S. H., Rastle, K. et Davis, M. H. (2012). Can Cognitive Models Explain Brain Activation During Word and Pseudoword Reading? A Meta-Analysis of 36 Neuroimaging Studies. *Psychological Bulletin*, *139*(4), 766–791. https://doi.org/10.1037/a0030266

Thibault, Lenfant et Helloin, 2012 Batterie informatisée d'examen du Langage Oral, Langage écrit, Mémoire, Attention, Compétences transversales (Exalang 8-11).

van der Kleij, S. W., Groen, M. A., Segers, E., et Verhoeven, L. (2019). Enhanced semantic involvement during word recognition in children with dyslexia. *Journal of Experimental Child Psychology*, *178*, 15–29. https://doi.org/10.1016/j.jecp.2018.09.006

van Dijk, T. A. et Kintsch, W. (1983). Strategies of Discourse Comprehension New York : Academic Press. *New York: Academic Press*, 413. https://doi.org/10.2307/415483

van Viersen, S., Kroesbergen, E. H., Slot, E. M. et de Bree, E. H. (2016). High Reading Skills Mask Dyslexia in Gifted Children. *Journal of Learning Disabilities*, *49*(2), 189–199. https://doi.org/10.1177/0022219414538517

Vellutino, F. R. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., et Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology et Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., et 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. https://doi.org/10.1080/10888430709336632

Walczyk, J. (2000). The Interplay Between Automatic and Control Processes in Reading. *Reading Research Quarterly*, *35*(4), 554–566. https://doi.org/10.1598/RRQ.35.4.7

Walczyk, J. J., Marsiglia, C. S., Amanda, K., Bryan, K. S. et Bryan, J. (2004). Children's Compensations for Poorly Automated Reading Skills. *Discourse Processes*, *37*(1), 47–66. https://doi.org/10.1207/s15326950dp3701

Walczyk, J. J., Marsiglia, C. S., Bryan, K. S. et Naquin, P. J. (2001). Overcoming inefficient reading skills. *Journal of Educational Psychology*, *93*(4), 750–757. https://doi.org/10.1037//0022-0663.93.4.750

Ward, J. (2015). *The Student's Guide to Cognitive Neuroscience*. London: Psychology Press.

Welcome, S. E., Chiarello, C., Halderman, L. K., et Leonard, C. M. (2009). Lexical processing skill in college-age resilient readers. *Reading and Writing*, *22*(3), 353–371. https://doi.org/10.1007/s11145-008-9120-3

West, R. et Stanovich, K. (1978). Automatic contextual facilitation in readers of three ages. *Child Development*, 49, 717-727.

West, R. P., Stanovich, K. E., Feeman, D. et Cunningham, A. (1983). The effect of sentence context on word recognition in second- and sixth- grade children. *Reading Research Quarterly*, 19, 6-15.

Ying, X. (2013). *Approche cognitive en didactique des langues : analyse et interprétation d'erreurs écrites prototypiques en français langue étrangère par des apprenants chinois et remédiation*. [Thèse de doctorat, Université Michel de Montaigne - Bordeaux III].

Ying, X. (2016). Introduction d'une approche cognitive en didactique des langues. *Pensées Vives*, 227–251.

Yuill, N. et Oakhill, J. V. (1991). *Children's problems in reading comprehension*. Cambridge: Cambridge University Press.

Zheng, X. (2008). *What happens in the first 200 ms of word reading: ERP studies on visual word recognition with top-down and bottom-up approaches*. Brock University.

Ziegler, J. C. et Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C. et Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*, *9*(5), 429–436. https://doi.org/10.1111/j.1467-7687.2006.00509.x

Zorman, M., Valdois, S., Jacquier-Roux, M. (2002). Un outil de dépistage des dyslexies développementales (ODEDYS). Diffusé en accès libre sur le site Internet du laboratoire Cogni-Sciences et Apprentissage de l'IUFM de Grenoble : http://www.grenoble.iufm.fr/recherch/cognisciences/index.html

ANNEXE A

Phonèmes du français et symboles de l'alphabet phonétique international

LES PRINCIPAUX GRAPHÈMES DU FRANÇAIS

PHONÈMES	GRAPHÈMES	
VOYELLES		
[i]	**i**	*part**i**r*
	y	*c**y**gne*
	hi	***hi**ver*
	hy	***hy**giène*
	ï	*maï**s***
	î	***î**le*
[e]	**é**	*prédire*
	e (+ cons.)	*pi**e**d*
	hé	***hé**ron*
	aï	*qu**ai***
	œ	*fœtus*
	æ	*et cætera*
[ɛ]	**è**	*grè**v**e*
	e	*m**e**r*
	aï	*m**ai**re*
	ê	*f**ê**te*
	ei	*p**ei**ne*
	he	*h**e**rbe*
	ë	*Noël*
	aî	*n**aî**tre*
[a]	**a**	*p**a**pa*
	à	*à*
	ha	*h**a**bit*
	e + m	*f**e**mme*
[y]	**u**	*u**n**ité*
	hu	*h**u**ron*
	û	*m**û**r*
	eu	*j'ai **eu***
[ø]	**eu**	*j**eu***
	œu	*v**œu***
	eû	*j**eû**ne*
[œ]	**eu**	*p**eu**r*
	œu	*c**œu**r*
	heu	*h**eu**re*
[ə]	**e**	*v**e**nir*
	on	*m**on**sieur*
	ai	*f**ai**sons*
[u]	**ou**	*f**ou***
	oû	*g**oû**t*
	où	*où*
	hou	*h**ou**blon*
	aoû	***aoû**t*
[o]	**o**	*r**o**se*
	au	*p**au**me*
	eau	*châ**teau***
	ô	*c**ô**te*
	hau	*h**au**teur*
[ɔ]	**o**	*p**o**rter*
	ho	*h**o**ckey*
	u (+m)	*alb**u**m*
[ɑ]	**â**	*p**â**te*
	hâ	*h**â**te*
	a	*b**a**s*
[ɛ̃]	**in**	*p**in**te*
	im	*im**p**oser*
	en	*r**ien***
	ain	*m**ain***
	ein	*fr**ein***
	yn	*s**yn**thèse*
	ym	*s**ym**pathique*
[œ̃]	**un**	*br**un***
	hum	*h**um**ble*
[ɔ̃]	**on**	*s**on**ge*
	hon	*h**on**te*
	om	*t**om**ber*
[ɑ̃]	**an**	*bl**an**che*
	am	*am**p**le*
	en	*s**en**tir*
	em	*r**em**plir*
	han	*h**an**che*
CONSONNES		
[p]	**p**	*p**a**ge*
	pp	*a**pp**el*
[b]	**b**	*b**ou**le*
	bb	*a**bb**é*
[m]	**m**	*m**er***
	mm	*so**mm**eil*
[t]	**t**	*t**a**ble*
	tt	*so**tt**ise*
	th	*th**é**orie*
[d]	**d**	*d**r**oit*
	dd	*a**dd**ition*
[n]	**n**	*n**ei**ge*
	nn	*so**nn**er*
[k]	**qu**	*qu**a**nd*
	c	*c**o**rde*
	cc	*a**cc**ord*
	k	*k**i**lo*
	cqu	*a**cqu**itter*
	ch	*te**ch**nologie*
	ck	*ti**ck**et*
[g]	g	*g**oû**t*
	gg	*a**gg**raver*
	gu	*ba**gu**ette*
	c	*se**c**onder*
	gh	***gh**etto*
[ɲ]	**gn**	*a**gn**eau*
	ign	*o**ign**on*
[f]	**f**	*f**u**sion*
	ff	*e**ff**usion*
	ph	***ph**are*
[v]	**v**	*v**e**nin*
	w	*w**a**gon*
[s]	**s**	*s**a**chet*
	ss	*la**ss**e*
	c	*c**i**ller*
	ç	*ran**ç**on*
	t	*atten**t**ion*
	sc	*s**c**ience*
	x	*si**x***
[z]	**s**	*ru**s**e*
	z	*z**é**ro*
	x	*di**x**ième*
[ʃ]	**ch**	*ch**a**t*
	sch	*sch**é**ma*
[ʒ]	**j**	*j**a**mais*
	g	*g**e**rbe*
	ge	*pi**ge**on*
[ʁ]	**r**	*r**u**e*
	rr	*a**rr**êt*
[l]	**l**	*l**a**gune*
	ll	*a**ll**ure*
[j]	**i**	*l**i**en*
	il	*bi**ll**et*
	ill	*feu**ill**u*
	y	***y**oga*
	ï	*a**ï**eul*
	hi	*h**i**érarchie*
[ɥ]	u	*br**u**it*
	hu	*caca**hu**ète*
[w]	**oi** [w]+[a]	*r**oi***
	ou (+ voy.)	*o**ua**te*
	oin [w]+[ɛ̃]	*l**oin***
	oê [w]+[a]	*p**oê**le*

Chartrand *et al.*, 1999, p. 16