

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LE TRAITEMENT MORPHOGRAPHIQUE DES MOTS COMPOSÉS FORMÉS DE RACINES LATINES ET
GRECQUES CHEZ DES ÉLÈVES DU 3E CYCLE DU PRIMAIRE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN ÉDUCATION

PAR

KATHLEEN WHISSELL-TURNER

OCTOBRE 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Le doctorat et, par le fait même la rédaction de cette thèse, est l'accomplissement d'un sinueux cheminement professionnel et personnel. Malgré les embuches, c'est aussi un des plus beaux cadeaux que j'aie pu m'offrir. Néanmoins, ce projet ambitieux n'aurait pas pu se concrétiser sans le soutien de plusieurs personnes.

Mes remerciements les plus sincères et chaleureux vont à ma directrice de recherche, Anila Fejzo, qui m'accompagne depuis mes tout premiers pas en recherche, alors que j'étais encore au baccalauréat. Merci d'avoir décelé en moi ce petit je ne sais quoi et de m'avoir toujours soutenue dans mon parcours universitaire, dans les hauts comme dans les bas. Votre rigueur et vos attentes élevées ont fait de moi la chercheuse que je suis aujourd'hui. Surtout, merci pour votre confiance, votre disponibilité, votre écoute, votre sagesse et vos précieux conseils. Je tiens aussi à remercier chaleureusement ma codirectrice de recherche, Laura Gonnerman, qui a su rapidement reprendre le flambeau et nous permettre de terminer ce projet en beauté. D'autres belles collaborations nous attendent. Merci aussi à Jill Vandermeerschen, statisticienne au Service de consultation en analyse de données (SCAD) de l'UQAM pour sa patience et ses exemples extravagants, mais extrêmement éclairants.

Je tiens également à remercier la présidente de mon jury de thèse, Catherine Turcotte, ainsi que mes trois évaluatrices, Catherine Mimeau, Pauline Quémart et Phaedra Royle, pour la lecture attentive de cette thèse doctorale et leurs précieux commentaires qui me permettent de présenter une thèse bonifiée et qui nourrissent mes réflexions dans le domaine.

Aussi, cette étude n'aurait jamais pu être possible sans l'implication de mes participants et de leurs parents. Chers élèves, merci! Continuez d'être aussi curieux et avides de nouvelles expériences.

Sur une note plus personnelle, je souhaite remercier mes amis doctorants, plus particulièrement Ariane, Rihab et Ivan, pour leur écoute et pour nos discussions des plus animées. Vous seuls savez! Merci aussi à ma famille et à mes amies, Maude et Ariane, qui me rappelaient que j'avais une vie sociale à entretenir.

Enfin, merci à mon mari, Fred, mon ancre et mon complice au quotidien. Tous ces moments, tu les as vécus à mes côtés et tu m'as toujours soutenue et encouragée. J'aimerais tant me voir à travers tes yeux. Merci de croire en moi et de toujours m'amener à me surpasser. Merci d'être fier de moi.

DÉDICACE

À moi.

TABLE DES MATIÈRES

REMERCIEMENTS	II
DÉDICACE	IV
LISTE DES FIGURES.....	VIII
LISTE DES TABLEAUX	X
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	XIV
LISTE DES SYMBOLES ET DES UNITÉS	XV
RÉSUMÉ	XVI
ABSTRACT	XVII
INTRODUCTION	1
CHAPITRE 1 PROBLÉMATIQUE	3
1.1 La réussite en lecture	3
1.2 Les textes courants et leurs particularités	4
1.3 L'apprentissage du vocabulaire	6
1.4 Le traitement morphographique des mots polymorphémiques	11
1.5 Les objectifs de recherche	13
1.6 La pertinence scientifique de la recherche	14
1.7 La pertinence sociale de la recherche.....	15
CHAPITRE 2 CADRE CONCEPTUEL	17
2.1 La construction des mots en français.....	17
2.1.1 Une définition du concept <i>mot</i>	18
2.1.2 Une définition des concepts <i>morphologie</i> et <i>morphèmes</i>	18
2.1.3 Une définition du concept <i>racines latines ou grecques</i>	21
2.1.4 Les procédés de formation des mots en français	25
2.2 Les différentes catégorisations des types de vocabulaire	29
2.3 Le développement de la connaissance du vocabulaire	32
2.4 L'identification des mots écrits	35
2.4.1 Une définition du concept <i>identification des mots écrits</i>	35
2.4.2 Les modèles de l'identification des mots écrits.....	36
2.4.2.1 Les modèles de décomposition morphologique	37
2.4.2.2 Les modèles holistiques.....	40

2.4.2.3	Les modèles hybrides.....	42
2.4.2.4	Les modèles connexionnistes distribués	46
2.4.2.5	Synthèse des modèles de l'identification des mots écrits.....	47
2.4.3	Le traitement morphographique	48
2.5	Les mesures de l'identification des mots écrits	49
2.5.1	La tâche de décision lexicale.....	49
2.5.2	Les mesures psychophysiques	51
2.6	Les facteurs qui influencent le développement du vocabulaire	57
2.7	Les facteurs qui influencent l'identification des mots écrits polymorphémiques.....	60
2.7.1	Les études portant sur les effets de la structure morphologique des items dans le traitement des mots composés.....	61
2.7.2	Les études portant sur les effets de fréquence dans le traitement des mots composés.....	68
2.7.3	Les études portant sur les effets de transparence sémantique dans le traitement des mots composés	73
2.7.4	Les études portant sur les effets de la position de la tête dans le traitement des mots composés	74
2.7.5	Les études portant sur les effets du type de racine dans le traitement des mots dérivés.....	77
2.7.6	La synthèse des facteurs qui influencent l'identification des mots écrits polymorphémiques ..	84
2.8	Les questions et les hypothèses de recherche.....	86
CHAPITRE 3 MÉTHODOLOGIE.....		90
3.1	Le devis de recherche.....	90
3.2	Les participants	91
3.2.1	La description de l'échantillon et de la méthode d'échantillonnage	91
3.2.2	Les critères et les mesures d'inclusion et d'exclusion	92
3.3	Les instruments de mesure	95
3.3.1	La tâche de décision lexicale : le traitement morphographique	95
3.3.1.1	Les stimulus de la tâche de décision lexicale.....	95
3.3.1.2	Le déroulement de la tâche de décision lexicale.....	102
3.3.2	Le test dynamique pour mesurer la connaissance du vocabulaire, la connaissance des racines latines et grecques et les habiletés morphologiques.....	107
3.3.2.1	La validation du test dynamique.....	111
3.4	La procédure de la collecte de données	111
3.5	La préparation des données électroencéphalographiques	113
3.6	Le plan d'analyse des résultats	122
CHAPITRE 4 RÉSULTATS.....		128
4.1	Les statistiques descriptives des résultats de la tâche de décision lexicale	128
4.2	Les statistiques descriptives des résultats du test dynamique	130
4.3	L'influence des facteurs linguistiques sur le traitement des mots écrits.....	132
4.3.1	L'effet de la lexicalité	132

4.3.2	L'effet de la structure morphologique des items	139
4.3.3	L'effet du type de racine	148
4.3.4	L'effet de l'origine des racines	158
4.3.5	Les effets de fréquence du mot entier et des racines	167
4.4	L'influence des facteurs individuels sur le traitement des mots écrits.....	183
4.5	Résumé des résultats de recherche	184
CHAPITRE 5 DISCUSSION		190
5.1	L'influence des facteurs linguistiques sur le traitement des mots écrits.....	190
5.1.1	L'effet de la lexicalité sur le traitement des mots écrits.....	191
5.1.2	L'effet de la structure morphologique des items sur le traitement des mots écrits.....	196
5.1.3	L'effet du type de racine sur le traitement des mots écrits	200
5.1.4	L'effet de l'origine des racines sur le traitement des mots écrits	203
5.1.5	Les effets de fréquence du mot entier et des racines sur le traitement des mots écrits.....	205
5.2	L'influence des facteurs individuels sur le traitement des mots écrits.....	208
5.3	Les résultats à la lumière des modèles de l'identification des mots écrits polymorphémiques	211
CONCLUSION		217
ANNEXE A CERTIFICAT D'APPROBATION ÉTHIQUE		228
ANNEXE B STIMULUS POUR LA TÂCHE DE DÉCISION LEXICALE		229
ANNEXE C LISTES FRANCOPHONES DE VOCABULAIRE		232
ANNEXE D PROPRIÉTÉS PSYCHOLINGUISTIQUES DES STIMULUS DE LA TÂCHE DE DÉCISION LEXICALE ET DE LEURS RACINES		238
ANNEXE E PROTOCOLE DE PASSATION DE LA TÂCHE DE DÉCISION LEXICALE		254
ANNEXE F ITEMS DU TEST DYNAMIQUE		257
ANNEXE G STATISTIQUES DESCRIPTIVES DE L'AMPLITUDE ET DE LA LATENCE DE 200 MS À 350 MS ET DE 350 MS À 500 MS		259
APPENDICE A MODÈLES D'IDENTIFICATION DES MOTS ÉCRITS QUI N'ABORDENT PAS EXPLICITEMENT LA MORPHOLOGIE.....		268
APPENDICE B RÉSULTATS DES ANALYSES DE COMPARAISON DES MOYENNES DES DIFFÉRENTES PROPRIÉTÉS PSYCHOLINGUISTIQUES DES CATÉGORIES DES ITEMS DE LA TÂCHE DE DÉCISION LEXICALE		274
RÉFÉRENCES		278

LISTE DES FIGURES

Figure 2.1 Conceptualisation de la structure des mots en français	21
Figure 2.2 Conceptualisation de la structure du mot <i>inspection</i>	24
Figure 2.3 Conceptualisation de la structure du mot <i>hydrographie</i>	24
Figure 2.4 Décours temporel des principaux marqueurs de potentiels évoqués d'identification des mots écrits	55
Figure 3.1 Plan factoriel des stimulus pour la tâche de décision lexicale	97
Figure 3.2 Présentation des stimulus durant la tâche de décision lexicale	107
Figure 3.3 Représentation schématique des électrodes dans chacune des six régions topographiques d'intérêt	114
Figure 3.4 Étapes de préparation réalisées sur les enregistrements électroencéphalographiques bruts de chacun des participants	122
Figure 4.1 Groupements des catégories d'items pour vérifier l'effet de la lexicalité	133
Figure 4.2 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la lexicalité	135
Figure 4.3 Groupements des catégories d'items pour vérifier l'effet de la structure morphologique des items	139
Figure 4.4 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la structure morphologique des mots.....	143
Figure 4.5 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la structure morphologique des pseudomots et des non-mots	144
Figure 4.6 Groupements des catégories d'items pour vérifier l'effet du type de racine.....	149
Figure 4.7 Tracés des potentiels évoqués aux six régions d'intérêt en fonction du type de racine des mots	153
Figure 4.8 Tracés des potentiels évoqués aux six régions d'intérêt en fonction du type de racine des pseudomots et des non-mots.....	154
Figure 4.9 Groupements des catégories d'items pour vérifier l'effet de l'origine des racines.....	159
Figure 4.10 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de l'origine des racines des mots	162

Figure 4.11 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de l'origine des racines des pseudomots	163
Figure 4.12 Distribution des données initiales pour la latence de 200 ms à 350 ms et de 350 ms à 500 ms	171
Figure 4.13 Distribution des données filtrées sans les bornes inférieures et supérieures pour la latence de 200 ms à 350 ms et de 350 ms à 500 ms.....	172

LISTE DES TABLEAUX

Tableau 3.1 Niveau de scolarité le plus élevé obtenu par un des parents des participants.....	93
Tableau 3.2 Moyennes et écarts types des fréquences des items et de leurs racines pour chacune des catégories d'items de la tâche de décision lexicale.....	101
Tableau 3.3 Modalités temporelles de tâches de décision lexicale dans des recherches mesurant le traitement morphographique	104
Tableau 3.4 Modalités visuelles utilisées lors de la présentation des stimulus dans les tâches de décision lexicale de recherches mesurant le traitement morphographique	105
Tableau 3.5 Structure du test dynamique avec l'item <i>ovipare</i> comme exemple	108
Tableau 3.6 Nombre de segments utilisés et perdus pour chacune des conditions expérimentales.....	115
Tableau 3.7 Nombre de segments utilisés et perdus par participant pour chacune des conditions expérimentales	116
Tableau 4.1 Pourcentage moyen, écart type et intervalle de confiance des bonnes réponses pour chaque catégorie d'items à la tâche de décision lexicale	129
Tableau 4.2 Temps de réaction moyen (ms) pour chaque catégorie d'items réussis à la tâche de décision lexicale	130
Tableau 4.3 Moyennes, écarts types et intervalles de confiance des habiletés mesurées au test dynamique.....	131
Tableau 4.4 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon la lexicalité	133
Tableau 4.5 Résultats de l'ANOVA à mesures répétées pour l'effet de la lexicalité sur la précision (%) à la tâche de décision lexicale.....	133
Tableau 4.6 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon la lexicalité.....	134
Tableau 4.7 Résultats de l'ANOVA à mesures répétées pour l'effet de la lexicalité sur le temps de réaction (ms) à la tâche de décision lexicale	134
Tableau 4.8 Moyennes et écarts types de l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la lexicalité pour chaque région d'intérêt.....	136
Tableau 4.9 Résultats des ANOVAs à mesures répétées pour l'effet de la lexicalité sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	137

Tableau 4.10 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la lexicalité pour chaque région d'intérêt	138
Tableau 4.11 Résultats des ANOVAs à mesures répétées pour l'effet de la lexicalité sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	138
Tableau 4.12 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon la structure morphologique des items	140
Tableau 4.13 Résultats de l'ANOVA à mesures répétées pour l'effet de la structure morphologique des items sur la précision (%) à la tâche de décision lexicale	140
Tableau 4.14 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon la structure morphologique des items	141
Tableau 4.15 Résultats de l'ANOVA à mesures répétées pour l'effet de la structure morphologique des items sur le temps de réaction (ms) à la tâche de décision lexicale	141
Tableau 4.16 Moyennes et écarts types de l'amplitude (μ V) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la structure morphologique des items pour chaque région d'intérêt	145
Tableau 4.17 Résultats des ANOVAs à mesures répétées pour l'effet de la structure morphologique des items sur l'amplitude (μ V) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	146
Tableau 4.18 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la structure morphologique des items pour chaque région d'intérêt	147
Tableau 4.19 Résultats des ANOVAs à mesures répétées pour l'effet de la structure morphologique des items sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	148
Tableau 4.20 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon le type de racine	149
Tableau 4.21 Résultats de l'ANOVA à mesures répétées pour l'effet du type de racine sur la précision (%) à la tâche de décision lexicale	150
Tableau 4.22 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon le type de racine	151
Tableau 4.23 Résultats de l'ANOVA à mesures répétées pour l'effet du type de racine sur le temps de réaction (ms) à la tâche de décision lexicale	151
Tableau 4.24 Moyennes et écarts types de l'amplitude (μ V) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon le type de racine pour chaque région d'intérêt	155

Tableau 4.25 Résultats des ANOVAs à mesures répétées pour l'effet du type de racine sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	156
Tableau 4.26 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon le type de racine pour chaque région d'intérêt	157
Tableau 4.27 Résultats des ANOVAs à mesures répétées pour l'effet du type de racine sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	158
Tableau 4.28 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon l'origine des racines	159
Tableau 4.29 Résultats de l'ANOVA à mesures répétées pour l'effet de l'origine des racines sur la précision (%) à la tâche de décision lexicale.....	159
Tableau 4.30 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon l'origine des racines	160
Tableau 4.31 Résultats de l'ANOVA à mesures répétées pour l'effet de l'origine des racines sur le temps de réaction (ms) à la tâche de décision lexicale	160
Tableau 4.32 Moyennes et écarts types de l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon l'origine des racines pour chaque région d'intérêt	164
Tableau 4.33 Résultats des ANOVAs à mesures répétées pour l'effet de l'origine des racines sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	165
Tableau 4.34 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon l'origine des racines pour chaque région d'intérêt.....	166
Tableau 4.35 Résultats des ANOVAs à mesures répétées pour l'effet de l'origine des racines sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale	167
Tableau 4.36 Résultats du modèle de régression logistique multiniveaux pour la précision (%) à la tâche de décision lexicale.....	170
Tableau 4.37 Résultats du modèle de régression linéaire multiniveaux pour le temps de réaction (ms) des participants à fournir une réponse correcte à la tâche de décision lexicale.....	173
Tableau 4.38 Résultats du modèle de régression linéaire multiniveaux pour l'amplitude maximale (μV) de 200 ms à 350 ms durant la tâche de décision lexicale	175
Tableau 4.39 Résultats du modèle de régression linéaire multiniveaux pour la latence (ms) de 200 ms à 350 ms durant la tâche de décision lexicale.....	177
Tableau 4.40 Résultats du modèle de régression linéaire multiniveaux pour l'amplitude maximale (μV) de 350 ms à 500 ms durant la tâche de décision lexicale	179
Tableau 4.41 Résultats du modèle de régression linéaire multiniveaux pour la latence (ms) de 350 ms à 500 ms durant la tâche de décision lexicale.....	181

Tableau 4.42 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la précision des participants à la tâche de décision lexicale	187
Tableau 4.43 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur le temps de réaction des participants à la tâche de décision lexicale	187
Tableau 4.44 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur l'amplitude des potentiels évoqués de 200 ms à 350 ms à la tâche de décision lexicale	188
Tableau 4.45 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur l'amplitude des potentiels évoqués de 350 ms à 500 ms à la tâche de décision lexicale	188
Tableau 4.46 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la latence des potentiels évoqués de 200 ms à 350 ms à la tâche de décision lexicale	189
Tableau 4.47 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la latence des potentiels évoqués de 350 ms à 500 ms à la tâche de décision lexicale	189

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ANOVA	Analyse de variance
ANG	Marqueur antérieur négatif gauche
BELEC	Batterie d'évaluation du langage écrit et de ses troubles
BOLD	<i>Blood oxygen level-dependant</i>
<i>ddl</i>	Degrés de liberté
DEL	Diode électroluminescente
EEG	Électroencéphalographie, électroencéphalographique
ÉLDEQ	Étude longitudinale du développement des enfants du Québec
IC	Intervalle de confiance
ICA	<i>Independant component analysis</i>
ICC	Coefficient de corrélation intraclasse
IRMf	Imagerie par résonance magnétique fonctionnelle
LAN	<i>Left anterior negativity</i>
MAPL	Modèle d'activation parallèle lexicomorphologique
max.	Maximum
MCM	Modèle de compétition morphologique
MEG	Magnétoencéphalographie
<i>n</i>	Nombre de participants
p. ex.	Par exemple
PISA	<i>Programme for International Student Assessment</i>
TDA/H	Trouble du déficit de l'attention avec ou sans hyperactivité
TLFi	Trésor de la langue française informatisé
vs	Versus

LISTE DES SYMBOLES ET DES UNITÉS

*	Marque un pseudomot ou un non-mot
d	Mesure de la taille d'effet de Cohen
$ÉT$	Écart type
F	Distribution F
Hz	Hertz
M	Moyenne
ms	Millisecondes
p	Probabilité
t	Distribution t de Student
α	Alpha de Cronbach
μV	Microvolts
η^2p	Eta-carré partiel
σ^2	Variance de la population

RÉSUMÉ

La présente étude vise à vérifier comment les élèves du 3^e cycle du primaire traitent les mots polymorphémiques du vocabulaire disciplinaire formés de racines latines et grecques et à dégager les facteurs linguistiques et individuels qui interviennent dans le traitement de ces mots. Pour ce faire, 29 élèves de 5^e ou de 6^e année du primaire ont participé à cette étude. Les participants ont réalisé une tâche de décision lexicale durant laquelle ils devaient lire silencieusement des séries de mots et de non-mots variant selon la structure morphologique (polymorphémique ou monomorphémique), le type de racine (libre ou liée) et l'origine des racines (latine ou grecque). Durant la tâche de décision lexicale, l'activité cérébrale des participants a été recueillie grâce à l'électroencéphalographie dans le but d'extraire les potentiels évoqués aux fenêtres temporelles de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus typiquement associées aux marqueurs N250 et N400. Les participants ont aussi passé un test dynamique sous forme d'entrevue semi-dirigée permettant de mesurer leur connaissance du vocabulaire disciplinaire, leurs habiletés morphologiques à reconnaître une racine dans un mot polymorphémique et à décomposer un mot en morphèmes, leur connaissance du sens de racines latines et grecques ainsi que leur habileté à déduire le sens d'un mot polymorphémique par l'analyse morphologique, soit la restitution du sens du mot par le sens de chacun de ses morphèmes. Les résultats indiquent que le traitement des mots composés formés de racines latines et grecques est influencé au niveau de l'activité cérébrale par la lexicalité et la structure morphologique des items. Au niveau comportemental, l'identification des mots écrits demeure influencée par la lexicalité et la structure morphologique des items, mais elle l'est aussi par le type de racine, l'origine des racines, la fréquence du mot, la fréquence des racines et les habiletés individuelles de décomposition morphologique des participants. Les résultats suggèrent que les élèves du 3^e cycle du primaire traitent les informations morphologiques lors de l'identification des mots et pseudomots composés formés de racines latines et grecques. Ces résultats pourraient nourrir les réflexions quant à la conception de dispositifs didactiques visant à développer le vocabulaire disciplinaire des élèves.

Mots-clés : identification des mots écrits, traitement morphographique, racines latines et grecques, connaissances morphologiques, N250, N400.

ABSTRACT

The purpose of this study is to investigate how students in the third cycle of elementary school process polymorphemic domain-specific academic vocabulary words formed from Latin and Greek roots and to identify the linguistic and individual factors involved in the processing of these words. To do so, 29 elementary school students in Grade 5 or 6 participated in this study. Participants completed a lexical decision task in which they were asked to silently read sets of words and non-words that varied according to their morphological structure (polymorphemic or monomorphemic), the type of root (free or bound) and the origin of the roots (Latin or Greek). During the lexical decision task, participants' brain activity was collected using electroencephalography in order to extract evoked potentials at the 200 ms to 350 ms and 350 ms to 500 ms poststimulus time windows typically associated with N250 and N400 markers. Participants also completed a dynamic assessment in the form of a semi-structured interview to measure their knowledge of domain-specific academic vocabulary, their morphological skills in recognizing a root in a polymorphemic word and in decomposing a word into morphemes, their knowledge of the meaning of Latin and Greek roots, and their ability to deduce the meaning of a morphologically complex word through morphological analysis, or the rendering of the meaning of the word by the meaning of each of its morphemes. The results indicate that the processing of compound words formed from Latin and Greek roots is influenced at the level of brain activity by the lexicality and the morphological structure of the items. At the behavioral level, the identification of written words remains influenced by the lexicality and the morphological structure of the items, but is also influenced by the type of roots, the origin of the roots, the frequency of the word, the frequency of the roots, and the individual morphological decomposition skills of participants. The results suggest that Grade 5 and 6 students process morphological information when identifying written compound words and pseudowords formed from Latin and Greek roots. These results could support reflections on designing didactic tools aimed at developing students' domain-specific academic vocabulary.

Keywords : written words identification, morphographic processing, Greek and Latin roots, morphological knowledge, N250, N400.

INTRODUCTION¹

Les élèves du 3^e cycle du primaire rencontrent une grande quantité de mots composés formés de racines latines et grecques dans les textes de leurs manuels scolaires (p. ex. *manuscrit, omnivore*). Alors qu'une connaissance de ces racines latines et grecques pourrait les mener à accéder plus facilement au sens des mots et à mieux comprendre les textes qui les contiennent, on ne sait pas si les élèves utilisent ces connaissances et habiletés lors de l'identification des mots écrits, s'ils accèdent directement au sens entier des mots ou s'ils les décomposent en unités porteuses de sens pendant le traitement lexical. Peu de recherches ont été effectuées sur le traitement des mots composés formés de racines latines et grecques, encore moins chez des lecteurs intermédiaires. Les résultats des études antérieures ayant porté sur le traitement des mots composés auprès d'adultes, sans toucher spécifiquement à ces racines particulières que sont les racines latines et grecques, laissent envisager que les lecteurs décomposeraient les mots composés en racines avant d'accéder à leur sens et que ce traitement s'appliquerait particulièrement aux mots fréquents et transparents sémantiquement.

Dans ce contexte, la présente étude vise à vérifier comment les élèves du 3^e cycle du primaire traitent les mots polymorphémiques du vocabulaire disciplinaire formés de racines latines et grecques et à identifier les facteurs linguistiques et individuels qui interviennent lors de l'identification de ces mots. Plus spécifiquement, l'effet des facteurs linguistiques de la lexicalité (mots, pseudomots/non-mots), de la structure morphologique des items (polymorphémique, monomorphémique), du type de racine (racine libre, racine liée), de l'origine des racines (latine, grecque), de la fréquence du mot entier et de la fréquence des racines sur le traitement morphographique (la précision et le temps de réaction à la tâche de décision lexicale ainsi que l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus) seront mesurés chez des élèves du 3^e cycle du primaire. Pour ce faire, une tâche de décision lexicale permettra d'activer les processus cognitifs mis en œuvre lors de l'identification des mots écrits. Aussi, à l'aide d'un test dynamique, cette étude permettra de mesurer l'effet des facteurs individuels de connaissance du vocabulaire disciplinaire, de connaissance des racines latines et grecques et des habiletés morphologiques sur le traitement morphographique des élèves du 3^e cycle du primaire.

¹ Cette thèse adopte l'orthographe rectifiée.

Pour répondre aux objectifs de recherche, le premier chapitre de la thèse dresse un bref état des lieux sur la problématique de recherche en passant notamment par la réussite en lecture des élèves québécois, les difficultés qui les guettent ainsi que la place accordée à l'enseignement et à l'apprentissage du vocabulaire en milieu scolaire. Le deuxième chapitre, le cadre conceptuel, définit les concepts centraux de l'étude avant de présenter une recension des principaux modèles d'identification des mots écrits, en portant une attention particulière sur ceux qui accordent une place à la morphologie. Ce chapitre se termine par un tour d'horizon des recherches causales qui portent sur le traitement morphographique puis clôt sur les questions et les hypothèses de recherche. Le troisième chapitre présente la méthodologie privilégiée pour répondre aux questions de recherche tout en présentant différents aspects méthodologiques comme le profil des participants, les instruments de mesure et leur constitution, la procédure de la collecte de données ainsi que la préparation des données électroencéphalographiques en vue de l'analyse des résultats. Le quatrième chapitre, soit celui dédié aux résultats, rapporte d'abord les statistiques descriptives des résultats de la tâche de décision lexicale et du test dynamique. Puis, les résultats des analyses permettant d'identifier l'influence des facteurs linguistiques (lexicalité, structure morphologique des items, type de racine, origine des racines, fréquence du mot entier, fréquence des racines) et individuels (connaissance du vocabulaire disciplinaire, reconnaissance des racines latines et grecques, habiletés de décomposition morphologique, connaissance du sens des racines latines et grecques, habiletés d'analyse morphologique) sur le traitement morphographique seront présentés. Enfin, le cinquième chapitre est consacré à la discussion des résultats. Dans ce chapitre, les résultats de la présente étude sont discutés à la lumière des études antérieures et au regard des modèles de traitement morphographique, dans le but de faire ressortir l'apport scientifique et l'unicité des résultats. Enfin, les retombées scientifiques et sociales ainsi que les limites de la recherche sont présentées dans la conclusion, afin d'alimenter la réflexion des lecteurs sur des pistes de recherches futures dans le domaine du traitement morphographique lors de l'identification des mots écrits auprès des lecteurs intermédiaires.

CHAPITRE 1

PROBLÉMATIQUE

Ce chapitre permet de cerner le problème de recherche de la présente étude à partir d'enjeux relevant des milieux scientifique et scolaire. Dans la première section, un état des lieux de la réussite en lecture au primaire est dressé (1.1). Puis, suit une deuxième partie sur les textes courants puisqu'ils représentent un défi majeur en lecture pour les élèves du primaire (1.2). Ensuite, dans la troisième section, les enjeux relatifs à l'apprentissage du vocabulaire sont présentés (1.3). Finalement, la quatrième partie porte sur l'identification des mots écrits qui est au cœur de cette étude, particulièrement le traitement morphographique des mots polymorphémiques (1.4). Cette présentation mènera à proposer des objectifs de recherche (1.5). Ce chapitre se termine en démontrant la pertinence scientifique (1.6) et la pertinence sociale de la présente étude (1.7).

1.1 La réussite en lecture

La *lecture*, dont l'objectif ultime est la compréhension du texte, se définit comme « l'ensemble des mécanismes cognitifs qui permettent l'accès au code écrit » (Bianco, 2015, p. 11). En effet, le lecteur doit réaliser « un ensemble d'opérations qui vont de la reconnaissance du mot écrit à la construction d'une représentation cohérente de la situation décrite, en passant par l'analyse syntaxique et la mise en relation des référents et idées énoncés dans les phrases successives » (Bianco, 2015, p. 11). La réussite scolaire (Janosz *et al.*, 2013; Pagani *et al.*, 2011; Snow, 1991) et sociale des individus (Statistique Canada, 2012) étant fortement liée à leur réussite en lecture, cette compétence constitue une des pierres angulaires des programmes de formation de l'école québécoise du préscolaire au secondaire (Ministère de l'Éducation, 2006a, 2006b; Ministère de l'Éducation, du Loisir et du Sport, 2007). Néanmoins, le développement de la lecture représente un défi pour les apprenants puisque cette habileté fait appel à de nombreux processus cognitifs qui œuvrent en parallèle. Un processus cognitif se définit comme « toute fonction mentale impliquée dans l'acquisition, le stockage, l'interprétation, la manipulation, la transformation et l'utilisation des connaissances »² [ma traduction] (American Psychological Association, s.d.).

² « Any of the mental functions assumed to be involved in the acquisition, storage, interpretation, manipulation, transformation, and use of knowledge. »

Les résultats de l'étude longitudinale du développement des enfants du Québec (ÉLDEQ) de 1998 à 2010 indiquent qu'un élève sur quatre n'atteint toujours pas le seuil de réussite attendu en lecture au terme de sa 6^e année du primaire (Desrosiers et Tétreault, 2012). Par ailleurs, certains élèves qui réussissaient bien en lecture depuis le début de leur scolarité éprouvaient soudainement de la difficulté en lecture en 4^e année du primaire (Snow *et al.*, 2005). Cette difficulté semble se perpétuer au-delà de l'école primaire. Effectivement, bien que les résultats de l'étude du PISA (*Programme for International Student Assessment*) de 2018 permettent au Canada de se qualifier parmi les pays des élèves de 15 à 16 ans ayant obtenu les meilleurs résultats en lecture, en mathématiques et en sciences, les résultats québécois demeurent inquiétants. En ne focalisant que sur la compétence en lecture, 12,3 % des élèves québécois évalués n'atteignent pas le seuil de compétence minimal (Organisation for Economic Co-operation and Development [OECD], 2019). Au seuil suivant, 20,5 % des élèves québécois atteignent de justesse la note de passage (OECD, 2019). Ces données permettent de constater que la compréhension en lecture représente un défi pour plusieurs élèves. Enfin, il est inquiétant de constater que, de façon générale, la proportion d'élèves canadiens qui n'atteint pas le seuil minimal de compétence en lecture a augmenté de 10,3 % en 2009 à 13,8 % en 2018 (OECD, 2019). Bien que ces résultats ne soient pas disponibles par province et territoire canadiens, on constate que le score moyen en lecture obtenu au Québec était à 532 en 2015 et qu'il a chuté à 519 en 2018 (OECD, 2019)³. Ces difficultés en lecture rencontrées chez les élèves à la fin de leur scolarité primaire (Desrosiers et Tétreault, 2012; OECD, 2019; Snow *et al.*, 2005) pourraient en partie s'expliquer par la présence accrue de textes courants et des racines latines et grecques qui forment les mots du vocabulaire disciplinaire au cœur de ces textes à partir de la fin du primaire.

1.2 Les textes courants et leurs particularités

Si la lecture représente un défi pour de nombreux élèves même après plusieurs années de scolarité obligatoire, cette difficulté est d'autant plus marquée lors de la lecture de textes courants comparativement aux textes littéraires (Bar-Ilan et Berman, 2007; Cartier, 2007; Desrosiers et Tétreault, 2012; Fang, 2006). De plus en plus nombreux dans les dernières années du primaire (Dymock, 2005), les

³ Il est impossible de fournir le score maximum du test en lecture puisque les résultats de chaque matière sont échelonnés pour correspondre à des distributions normales où la moyenne se situe autour de 500 points (OECD, s.d.).

textes courants comptent en effet pour 70 % à 80 % des textes présentés aux élèves (Daniels, 2002). Pour comprendre la difficulté supplémentaire qu'entraînent les textes courants lors de la lecture, ils sont comparés aux textes littéraires. Les textes courants se distinguent des textes littéraires par différents aspects, tous susceptibles d'influencer la compréhension des élèves : leur objectif, leur syntaxe, leur macrostructure, les formes sous lesquelles ils se présentent ou encore les mots qui les composent (Bar-Ilan et Berman, 2007; Cartier, 2007; Fang, 2006).

D'abord, contrairement aux *textes littéraires* qui servent à raconter une histoire, le principal objectif des *textes courants* est de communiquer une information et de permettre des apprentissages (Cartier, 2007). Aussi, les textes courants se distinguent des textes littéraires par leur syntaxe habituellement plus dense (Bar-Ilan et Berman, 2007; Nagy et Townsend, 2012). Ensuite, contrairement aux textes littéraires qui adoptent souvent une structure narrative prévisible à laquelle l'apprenant a été en contact dès le plus jeune âge, les textes courants peuvent présenter différentes structures informatives (p. ex. cause à effet, problème-solution, etc.) auxquelles les enfants sont exposés plus tard (Bar-Ilan et Berman, 2007; Cartier, 2007). Enfin, les textes courants se distinguent des textes littéraires par la fréquence des mots rares qui les composent. Les textes littéraires peuvent compter un plus grand nombre de mots rares (p. ex. *éthéré, vétuste*) que les textes courants. Cependant, ces mots rares ne sont pas souvent répétés dans un même texte littéraire, alors que, dans un texte courant, on peut retrouver jusqu'à six répétition du même mot rare (p. ex. *géologie, chromosome*), témoignant de son importance pour accéder au sens du texte (Hiebert et Cervetti, 2012).

Ainsi, les textes courants se caractérisent par la présence accrue de termes techniques issus du vocabulaire disciplinaire (Fang, 2006; Henry, 2010). Le *vocabulaire disciplinaire* est un vocabulaire spécifique à un domaine d'apprentissage (p. ex. écologie, géographie) et réfère à un concept précis (p. ex. *photosynthèse, hydrographie*) (Baumann et Graves, 2010). La compréhension de ce vocabulaire disciplinaire dans les textes courants est souvent inhérente à la compréhension globale des textes (Bar-Ilan et Berman, 2007; Blachowicz et Fisher, 2010; Nagy et Townsend, 2012), contrairement aux textes littéraires. En anglais, les mots du vocabulaire disciplinaire sont majoritairement formés de *racines latines ou grecques* (Nagy et Townsend, 2012), de petites unités lexico-sémantiques (morphèmes) d'origine latine ou grecque porteuses du sens principal du mot. La proportion de ces mots dans les textes courants destinés aux élèves de la fin du primaire est estimée entre 60 % et 90 % par des études menées en contexte anglophone (Bar-Ilan et Berman, 2007; Green, 2008). À notre connaissance, une telle

estimation n'est pas disponible en français. Cependant, sachant que le vocabulaire français recense 87 % de mots d'origine latine ou grecque (Bertrand, 2011) versus 60 % à 75 % pour le vocabulaire anglophone (Nagy et Anderson, 1984; Rasinski *et al.*, 2008; Stockwell et Minkova, 2001), il serait possible de penser que les élèves francophones rencontrent au moins une proportion semblable, sinon supérieure, de mots du vocabulaire disciplinaire issus du latin ou du grec que les élèves anglophones. Comme la compréhension d'un texte est tributaire à la connaissance des mots qui le composent, il est possible que la méconnaissance d'une grande proportion de mots formés de racines latines ou grecques affecte négativement la compréhension des textes courants.

Encore aujourd'hui, la productivité des racines latines et grecques est sans égal dans la formation des mots savants (Lesaux et Kieffer, 2010), mais aussi des mots nouveaux. Par exemple, avec la pandémie de la COVID-19, de nombreux mots formés d'au moins une racine latine ou grecque ont fait leur entrée dans le dictionnaire. C'est notamment le cas de *télétravail* qui s'est ajouté au dictionnaire, ou encore *aéroporté* qui s'est vu ajouter un nouveau sens (Le Robert, 2021). Ces mots ont la particularité d'être rapidement intégrés dans les textes courants. Or, la compréhension de leur sens pourrait être facilitée par la connaissance de leurs composantes, soit des racines latines et grecques. Par exemple, le lecteur qui connaît le sens de *télé* (racine grecque qui signifie « loin ») n'éprouverait pas de difficulté à comprendre le sens de *télétravail*.

Comme les racines latines et grecques sont fréquentes dans les textes utilisés pour l'enseignement des disciplines (Hiebert et Cervetti, 2012), il est possible que cela soit une des causes des difficultés en lecture liées aux textes courants chez les élèves du primaire. La méconnaissance des racines latines et grecques pourrait nuire non seulement à la compréhension du vocabulaire disciplinaire et des concepts sous-jacents (Fazio et Gallagher, 2014), mais aussi à celle des textes dans lesquels ils apparaissent (Nation, 2006), et par conséquent, aux apprentissages qui devraient être réalisés. Dans ce contexte, il convient de souligner la nécessité d'outiller les élèves à mieux connaître les mots présents dans leurs manuels scolaires, notamment ceux relatifs au vocabulaire disciplinaire composés de racines latines et grecques.

1.3 L'apprentissage du vocabulaire

Au Québec, les élèves du primaire devraient développer des connaissances sur le sens des racines latines et grecques, appelées *éléments savants*, les plus fréquentes dès la 4^e année du primaire avec l'aide de

leur enseignante ou de leur enseignant (Ministère de l'Éducation, du Loisir et du Sport, 2009). Au terme de leur 3^e cycle du primaire, soit en 6^e année du primaire, les élèves devraient être en mesure d'identifier le sens de ces racines latines et grecques fréquentes sans aide (Ministère de l'Éducation, du Loisir et du Sport, 2009). Cependant, on ne sait pas à quel point cette habileté est réellement développée en salle de classe. On sait toutefois que les enseignantes et les enseignants du primaire déclarent consacrer peu de temps à l'enseignement explicite sur les contenus lexicaux et sur les stratégies lexicales pour déduire le sens de mots inconnus lors de la lecture (Anctil *et al.*, 2018). Une connaissance des racines latines et grecques rencontrées lors de leurs lectures pourrait permettre aux élèves d'acquérir du vocabulaire disciplinaire nécessaire à la compréhension des textes.

Le modèle théorique de Perfetti et Stafura (2014) accorde une place centrale au vocabulaire lors de la lecture. En effet, selon ce modèle (Perfetti et Stafura, 2014), le vocabulaire (sens, morphologie et syntaxe – structure des arguments et rôles thématiques) assure le lien entre le système permettant l'identification des mots écrits et celui permettant la compréhension du texte. Comme une méconnaissance de seulement 2 % des mots d'un texte peut mener à une incompréhension en lecture (Nation, 2006; Observatoire National de la Lecture, 2000), le vocabulaire joue un rôle crucial pour bien comprendre un texte. Par exemple, dans la phrase « Il est interdit de *rouber à cet endroit. », la non-compréhension d'un seul mot, ici un non-mot, ne permet pas de comprendre l'entièreté du message. Il en va de même avec les mots inconnus, d'autant plus si le contexte ne fournit pas suffisamment d'indices pour déduire même partiellement le sens du mot, comme dans la phrase suivante : « J'ai développé une agranulocytose. » Confronté à de telles situations, un lecteur se voit dans l'impossibilité de faire les inférences nécessaires pour comprendre la phrase, ou le texte, afin d'accéder au sens véhiculé par l'auteur ou encore, en contexte éducatif, afin de réaliser un apprentissage.

Hiebert (2011) propose un classement du vocabulaire utilisé dans des textes scolaires anglophones afin de tenter d'estimer le nombre de mots que les élèves rencontrent dans leurs lectures. Le vocabulaire de base [ma traduction de *core vocabulary*], composé des 5 600 mots les plus communs en anglais et leurs flexions grammaticales [ma traduction de *core morphology*], forme environ 85 à 90 % des textes scolaires. Les 10 à 15 % restants sont composés de 300 000 à 600 000 mots rares. Il s'agit du vocabulaire littéraire et spécifique. Malgré la forte proportion de mots du vocabulaire de base dans les textes (jusqu'à 90 %), la connaissance de ceux-ci n'est pas suffisante pour accéder à leur compréhension. Selon Nation (2006), le seuil minimal de connaissance du vocabulaire d'un texte est plutôt estimé à 98 %. À

titre d'exemple, considérons qu'un individu connaît la totalité des mots du vocabulaire de base, soit 90 % du texte. À cela, doivent s'ajouter les 8 % supplémentaires constitués de mots moins fréquents afin d'atteindre le seuil minimal indiqué par Nation (2006). Ce lecteur hypothétique doit donc connaître un total de 24 000 à 48 000 mots littéraires et spécifiques pour combler cet écart de 8 % avec le vocabulaire de base. Donc, la connaissance d'une grande quantité de mots issus du vocabulaire spécifique, bien que représentant une plus faible proportion des textes, est indispensable. D'autres auteurs estiment que pour comprendre les textes qu'ils lisent dans leurs manuels scolaires et respecter une trajectoire de développement du vocabulaire optimale, les élèves du primaire doivent acquérir de 2 000 à 3 000 nouveaux mots par année (Anglin, 1993; Beck *et al.*, 2002). Cela correspond en moyenne à 17 mots par jour scolaire (Baumann *et al.*, 2003). Bien que ces estimations demeurent approximatives et varient d'un auteur à l'autre, il est inconcevable de penser qu'un enseignement systématique de chacun des mots puisse être possible pour développer le vocabulaire des élèves.

L'apprentissage des mots se réalise de deux façons : explicite et implicite. D'une part, *l'apprentissage explicite* se réalise lorsque les élèves prennent consciemment des moyens pour développer leur vocabulaire. Cela peut se faire en consultant le dictionnaire à la recherche d'une définition, par exemple, tout comme par l'enseignement direct des mots par l'enseignante ou l'enseignant, ou encore par les pairs (Nation, 2001). L'enseignement direct du vocabulaire permet aux élèves d'apprendre environ 400 mots annuellement (Beck *et al.*, 2002). Il est inconcevable de prévoir un enseignement systématique pour tous les mots susceptibles d'être rencontrés lors de la lecture. Ainsi, d'autre part, *l'apprentissage implicite* du vocabulaire se réalise de façon inconsciente, par le contact avec les mots. Les élèves qui rencontrent des mots à l'oral ou à l'écrit développent des connaissances sur ces mots de façon incidente (Nation, 2001). Or, l'apprentissage implicite du vocabulaire disciplinaire s'avère difficile puisque les élèves ont peu de chances de rencontrer ces mots à l'extérieur d'un contexte éducatif (Worthington et Nation, 1996).

Toutefois, deux stratégies lexicales pourraient se révéler efficaces pour rendre les élèves autonomes dans leur apprentissage des mots qu'ils rencontrent de façon incidente : l'analyse contextuelle et l'analyse morphologique. *L'analyse contextuelle* consiste en l'utilisation d'indices typographiques et linguistiques (syntaxiques ou sémantiques) dans le texte pour accéder au sens d'un mot inconnu (Baumann *et al.*, 2012). Par exemple, les indices typographiques comme l'utilisation de l'italique ou du gras peuvent être utilisés pour référer à la définition du mot ciblé dans un glossaire. Les indices

syntaxiques s'appuient sur l'ordre des mots dans la phrase alors que les indices sémantiques comprennent l'utilisation du sens des autres mots de la phrase afin d'attribuer un sens au mot inconnu. Les types d'indices contextuels sont nombreux : la définition, la synonymie, l'antonymie, l'exemple, le recours au champ lexical, etc. Cependant, cette stratégie d'analyse contextuelle est susceptible d'échouer fréquemment, particulièrement avec les mots du vocabulaire disciplinaire pour lesquels 85 % à 95 % des tentatives se révèlent infructueuses (Nagy *et al.*, 1985; Swanborn et de Glopper, 1999). Effectivement, contrairement au vocabulaire courant qui peut être plus facilement déduit à partir du contexte, les indices contextuels sont souvent insuffisants pour accéder au sens des mots appartenant au vocabulaire disciplinaire (Blachowicz et Fisher, 2010; Blachowicz et Fisher, 2000). Ceux-ci sont habituellement plus abstraits et nécessitent une compréhension approfondie du concept pour permettre la compréhension des idées principales du texte (Bar-Ilan et Berman, 2007; Blachowicz et Fisher, 2010; Nagy et Townsend, 2012).

La deuxième stratégie, c'est-à-dire l'analyse morphologique, semble une avenue plus prometteuse dans le cas de l'apprentissage du vocabulaire disciplinaire. En effet, si l'analyse contextuelle ne permet de déduire correctement le sens que d'environ 15 % des mots inconnus dans un texte (Swanborn et de Glopper, 1999), l'analyse morphologique serait efficace pour 60 % des nouveaux mots rencontrés (Nagy *et al.*, 1989), du moment que ceux-ci sont polymorphémiques, c'est-à-dire composés d'au moins deux morphèmes. L'*analyse morphologique* consiste en l'analyse de la structure du mot, et de chacun des morphèmes, pour déduire le sens d'un mot inconnu (Crosson *et al.*, 2021). Par exemple, un élève qui ne connaît pas le sens du mot *inconsolable* peut recourir à l'analyse morphologique en restituant son sens grâce au sens de chacun des morphèmes (*in-* + *console* + *-able* = « négation » + « console » + « pouvoir » = « qui ne peut pas être consolé »). Aussi, la connaissance d'un morphème (p. ex. *in-*) permet d'établir des relations entre les mots polymorphémiques partageant ce morphème (p. ex. le sens de « négation » dans *insupportable* et *innombrable*).

Combinée à l'analyse contextuelle, l'analyse morphologique serait une alternative intéressante pour l'enseignement du vocabulaire puisque l'apprentissage de seulement quatre racines par semaine permettrait aux élèves de déduire le sens de 800 à 3200 mots par année, soit des mots majoritairement du vocabulaire disciplinaire (Cabral, 2015; Rasinski *et al.*, 2011). L'analyse morphologique fournirait aussi des indices importants pour déduire le sens des mots inconnus du vocabulaire disciplinaire formés de racines latines et grecques, bien que ces derniers soient souvent abstraits (Blachowicz et Fisher, 2010;

Padak *et al.*, 2008). Habituellement, le sens des racines latines et grecques est transcrit par un mot du vocabulaire courant, donc plus accessible pour les élèves que le mot du vocabulaire disciplinaire (*hexagone* = « six » + « angle »). En plus, les racines latines et grecques ont la particularité de conserver leur sens d'un mot à l'autre (Palumbo *et al.*, 2015) : *hexagone*, *polygone*, *pentagone*, etc. Il est d'ailleurs possible de dériver de cinq à vingt mots à partir d'une seule racine latine ou grecque (Rasinski *et al.*, 2008), une proportion non négligeable pour la croissance du vocabulaire des élèves (White *et al.*, 1989).

D'ailleurs, différentes études montrent que l'analyse morphologique des mots composés de racines latines et grecques favorise le développement du vocabulaire (p. ex. Anglin, 1993; Bowers et Kirby, 2010; Crosson et McKeown, 2016; Crosson *et al.*, 2019; Crosson et Moore, 2017; Daigle et Gagnon, 2014; Goodwin, 2016). Une telle analyse permettrait de contribuer au développement de représentations lexicales de qualité puisqu'elle relie différents aspects de la connaissance des mots – orthographiques, phonologiques et sémantiques (Bowers *et al.*, 2010; Kirby et Bowers, 2017). Par exemple, il serait plus facile pour un lecteur qui connaît le sens et la forme de la racine grecque *hydr* d'établir les différentes représentations – sémantiques, orthographiques et phonologiques – des mots *hydrique* ou *hydrologie* même s'il les rencontre pour la première fois. La qualité de ces représentations contribue à son tour à une meilleure compréhension des textes dans lesquels se trouvent ces mots (Perfetti et Stafura, 2014).

Or, les études portant sur l'apport des racines latines et grecques en lecture de phrases et de textes ainsi que pour le développement du vocabulaire ont essentiellement été menées auprès de populations anglophones (Crosson et McKeown, 2016; Crosson *et al.*, 2019; Crosson et Moore, 2017; McKeown *et al.*, 2018). À notre connaissance, une seule étude a été menée en contexte francophone québécois (Whissell-Turner et Fejzo, 2021). Cette étude a permis d'observer que la connaissance de ces morphèmes entretient des relations étroites avec le vocabulaire et la compréhension en lecture chez des élèves de 6^e année du primaire (Whissell-Turner et Fejzo, 2021). Elle corrobore les résultats d'autres études empiriques anglophones qui ont permis de démontrer que la connaissance de racines latines et grecques permettait de mieux comprendre une phrase contenant un mot inconnu (Crosson et McKeown, 2016; Crosson *et al.*, 2019) et permettait de développer le vocabulaire disciplinaire des élèves du primaire et du secondaire (Crosson et Moore, 2017).

Si ces études en contextes francophone et anglophone convergent pour établir un lien étroit entre la connaissance des racines latines et grecques, le vocabulaire (notamment disciplinaire) et la

compréhension en lecture, les processus cognitifs inconscients qui sous-tendent ces relations sont encore méconnus. En d'autres termes, il est important de savoir si, à la lecture du mot *octogone*, l'élève accède à son sens dans sa globalité en l'associant au concept appris en classe, soit une forme géométrique rappelant celle d'un panneau d'arrêt, ou encore s'il le décompose automatiquement en unités lexico-sémantiques pour récupérer son sens (*octo* + *gone* = « forme à huit angles »). Contrairement à l'analyse morphologique qui est une stratégie lexicale qui peut être mise en place consciemment par l'élève pour déduire le sens d'un mot, ces processus d'identification des mots menant au traitement morphographique sont inconscients. L'élève ne peut pas les verbaliser. Cependant, une meilleure compréhension du traitement morphographique et des processus sous-jacents chez des élèves du 3^e cycle du primaire serait nécessaire, dans une perspective didactique, pour mieux orienter les futures recherches en didactique de la lecture et, ultimement, nourrir les réflexions didactiques sur l'enseignement du vocabulaire disciplinaire.

1.4 Le traitement morphographique des mots polymorphémiques

Le concept de *traitement morphographique* réfère aux processus mis en œuvre pour identifier les mots polymorphémiques à l'écrit (Amenta et Crepaldi, 2012). En psycholinguistique, plusieurs modèles ont été proposés pour tenter d'expliquer le traitement morphographique. À un extrême, il est suggéré que tous les mots sont décomposés en morphèmes avant l'accès lexical (p. ex. Taft et Forster, 1975; Taft, 1994), à l'autre, aucun mot n'est décomposé – tous les mots sont stockés indépendamment dans le lexique mental (p. ex. Butterworth, 1983; Manelis et Tharp, 1977). Cependant, chacune de ces propositions pose des limites importantes que les modèles hybrides et les modèles connexionnistes distribués tentent de résoudre. Les modèles hybrides (Caramazza *et al.*, 1985; Schreuder et Baayen, 1995) expliquent le traitement des mots polymorphémiques en proposant un accès lexical en deux voies où certains mots sont traités de manière globale alors que d'autres sont décomposés en morphèmes. Cette sélection est établie sur la base de divers facteurs comme la fréquence du mot, la fréquence des morphèmes constituant le mot ou encore sa *transparence sémantique*, soit la propriété d'un mot à être entièrement défini par le sens respectif de chacun de ses morphèmes. Les modèles connexionnistes distribués, quant à eux, proposent un traitement identique des mots qu'ils soient monomorphémiques ou polymorphémiques (Plaut et Gonnerman, 2000; Seidenberg et Gonnerman, 2000). Selon ces modèles, la morphologie, bien que procédé important dans la formation des mots, n'est pas un facteur influençant l'accès au lexique mental. Le mot est activé dans le lexique mental grâce aux interactions entre ses représentations phonologiques, orthographiques et sémantiques.

Bien que les modèles hybrides et connexionnistes distribués offrent des hypothèses plausibles, ils s'appuient sur des résultats comportementaux, comme les temps de lecture ainsi que des simulations computationnelles des processus cognitifs. Les processus d'identification des mots écrits sont inconscients, donc impossibles à identifier par le lecteur, et automatiques, c'est-à-dire qu'ils se produisent très rapidement, habituellement en deçà de 200 ms (Hruby et Goswami, 2011). L'utilisation unique de mesures comportementales ne permet pas de renseigner aussi précisément sur le déroulement temporel du traitement des mots puisqu'elles ne permettent de recueillir que deux points de mesure après l'identification des mots, soit le temps de réaction ainsi que la réponse du participant. L'ajout de mesures psychophysiologiques, comme l'analyse des potentiels évoqués, pourrait compléter les données comportementales en inférant sur les processus sous-jacents à l'identification des mots écrits, comme le traitement morphographique, avant qu'une réponse comportementale ne soit fournie. Les études en neurosciences éducationnelles n'en sont qu'à leurs balbutiements dans le domaine de l'apprentissage de la lecture, en particulier dans le domaine des connaissances morphologiques. Cependant, leur modeste contribution permettrait une meilleure compréhension des mécanismes d'apprentissage pour améliorer certains aspects de la réussite des élèves (Thomas et Ansari, 2020).

Les recherches sur le traitement morphographique ont surtout porté sur les mots fléchis et dérivés (Leminen *et al.*, 2018). Très peu de recherches visent à mieux comprendre le traitement des mots composés, soit le procédé de formation des mots privilégié avec les racines latines et grecques (Dal et Amiot, 2008; Dressler, 2006). À notre connaissance, aucune n'a encore porté sur le traitement des mots composés formés de racines latines et grecques. On ne sait donc pas encore par quelles étapes ou par quels processus passe le lecteur lorsqu'il identifie des mots polymorphémiques composés formés de racines latines et grecques.

De plus, les chercheurs qui se sont intéressés au traitement des mots polymorphémiques se sont concentrés sur les habiletés du lecteur expert. Il est possible que les processus cognitifs impliqués lors du traitement morphographique, ou leur latence, soient différents chez les lecteurs intermédiaires, notamment chez les élèves du 3^e cycle du primaire pour lesquels les processus d'identification des mots écrits ne sont pas encore tout à fait automatisés. Les recherches sur le traitement morphographique des mots en français ne convergent pas sur un moment précis d'activation des représentations morphologiques lors de l'identification des mots polymorphémiques. Selon certaines études, les apprentis lecteurs tiendraient compte des représentations morphologiques des mots dérivés vers le 2^e

cycle du primaire (Casalis *et al.*, 2009; Quémart et Casalis, 2015; Quémart *et al.*, 2011). Selon d'autres études, ce ne serait pas avant la 6^e année du primaire (Quémart et Casalis, 2015; Quémart *et al.*, 2012). Comme il a été montré que les locuteurs francophones développent d'abord une sensibilité pour la morphologie dérivationnelle plutôt que compositionnelle en production de néologismes (Clark, 1993), compte tenu de la dominance de la dérivation en français comparativement à la composition (Fejzo *et al.*, 2018), on pourrait penser que cette progression suit un parcours similaire durant l'identification des mots écrits. Comme les mots formés de racines latines et grecques relèvent principalement de la composition, il est pertinent d'observer le traitement de ces mots chez les élèves plus âgés du primaire.

La question est de savoir si les mots composés formés de racines latines et grecques sont traités et représentés en unités lexicales entières ou plutôt comme des unités lexico-sémantiques combinées et selon quels facteurs s'effectue ce traitement (Leminen *et al.*, 2018). La connaissance de ce traitement éclairerait les modèles actuels sur l'accès lexical des mots composés, plus particulièrement ceux formés de racines latines et grecques. De plus, les résultats de cette étude permettraient d'alimenter les réflexions sur l'enseignement à mettre en place pour soutenir le développement de ces processus. Ainsi, si les élèves décomposent les mots en morphèmes lors de la lecture, l'enseignement du vocabulaire disciplinaire pourrait s'orienter davantage vers un enseignement morphologique des racines latines et grecques productives pour favoriser l'accès lexical et la compréhension en lecture des élèves. En revanche, si les élèves traitent les mots du vocabulaire disciplinaire dans leur entièreté pour accéder à leur sens, les recommandations viseraient à vérifier si les élèves accèdent davantage au sens du mot et du texte à la suite d'un enseignement global du vocabulaire disciplinaire (présentation de la définition) ou d'un enseignement morphologique et si l'enseignement prodigué peut influencer l'accès lexical.

1.5 Les objectifs de recherche

Considérant qu'aucune recherche visant à explorer le traitement de mots polymorphémiques formés de racines latines et grecques ni des facteurs pouvant influencer ce traitement n'a encore été menée en contexte francophone ou en contexte anglophone, et que cette connaissance pourrait soutenir la conception de dispositifs d'enseignement et d'apprentissage du vocabulaire pour favoriser la compréhension de textes, la présente étude vise à répondre aux deux objectifs de recherche suivants :

- 1) Vérifier comment les élèves du 3^e cycle du primaire traitent les mots polymorphémiques du vocabulaire disciplinaire formés de racines latines et grecques.

2) Dégager les facteurs qui influencent le traitement des mots polymorphémiques du vocabulaire disciplinaire formés de racines latines et grecques auprès des élèves du 3^e cycle du primaire.

1.6 La pertinence scientifique de la recherche

Cette étude est la première recherche à porter sur le traitement des mots composés formés de racines latines et grecques. Il est d'ailleurs pertinent de s'intéresser à ce traitement en contexte francophone puisque le français est une langue qui se caractérise par une morphologie plus productive et plus riche que l'anglais (Casalis *et al.*, 2015). De plus, compte tenu de la plus grande quantité de mots d'origine latine et grecque en français qu'en anglais (Bertrand, 2011), on peut estimer que la quantité de mots formés de racines latines et grecques est de proportion similaire à l'anglais, sinon plus grande. Par conséquent, étudier le traitement des mots composés formés de racines latines et grecques en contexte francophone permet de situer la place de la morphologie compositionnelle durant l'identification des mots écrits, notamment pour les mots du vocabulaire disciplinaire.

Mieux comprendre le traitement des mots polymorphémiques formés de racines latines et grecques permet d'orienter les didacticiens dans le développement de programmes d'intervention appuyés par la recherche. D'abord, la présente recherche permet de situer les connaissances des élèves quant aux racines latines et grecques et d'identifier si de telles connaissances influencent l'identification de mots écrits formés de ces racines. Aussi, l'identification de facteurs linguistiques au sein de mots composés formés de racines latines et grecques pourrait permettre une classification de ces derniers pour l'enseignement et l'apprentissage des mots. En effet, savoir que la présence de racines fréquentes, par exemple, faciliterait l'identification des mots aiderait à la conception de programmes gradués. Aussi, l'identification de facteurs individuels tels que la connaissance des racines latines ou grecques ou la connaissance du vocabulaire pourrait permettre aux didacticiens de prévoir une progression adéquate dans le développement des programmes d'intervention en littératie, plus particulièrement du vocabulaire disciplinaire, en tenant compte des connaissances antérieures et préalables des élèves.

D'une perspective psycholinguistique, la prise de mesures psychophysiologiques conjointement aux mesures comportementales lors de l'identification de mots écrits formés de racines latines et grecques permet de renseigner sur les processus neurocognitifs d'identification des mots composés afin d'alimenter les modèles théoriques actuellement muets à cet effet. Les résultats permettent de mieux capturer les processus d'identification des mots écrits afin d'en définir les principales étapes. Ils

permettent aussi d'identifier avec précision le décours temporel de ces processus complexes. À notre connaissance, aucune recherche n'a, à ce jour, vérifié le traitement morphographique des mots composés formés de racines liées, plus particulièrement de racines latines et grecques. Contrairement aux racines libres (p. ex. *bon* et *soir*) qui sont autonomes dans la langue et qui correspondent aussi à une entrée lexicale dans le lexique mental, les racines liées ne sont pas autonomes dans la langue. On ne sait pas si ce type de racines liées, les racines latines et grecques, sont stockées indépendamment dans le lexique mental. Ainsi, il est possible que le traitement morphographique des mots composés formés de racines latines et grecques soit différent des mots composés formés de deux racines libres (p. ex. *bonsoir*).

D'une perspective développementale, cette recherche permet aussi de mieux comprendre l'apport des racines latines et grecques dans le développement du vocabulaire disciplinaire et dans l'accès lexical chez les élèves de la fin du primaire. Comme les recherches antérieures ont surtout été menées auprès de lecteurs experts adultes, les résultats de la présente recherche pourrait servir de pistes pour alimenter ou nuancer les conclusions des recherches antérieures afin de proposer une vision développementale des processus neurocognitifs en identification des mots écrits, du lecteur intermédiaire au lecteur adulte. Aussi, obtenus dans un nouveau contexte linguistique, soit en contexte francophone, les résultats permettent d'approfondir la compréhension du rôle des racines latines et grecques dans l'identification des mots du vocabulaire disciplinaire français, qu'ils soient connus ou inconnus des élèves.

1.7 La pertinence sociale de la recherche

La recherche proposée revêt un intérêt particulier dans la mesure où elle répond à certains besoins du milieu scolaire quant à l'apprentissage et à l'enseignement du vocabulaire disciplinaire à partir du 3^e cycle du primaire. D'abord, la recherche permet de dissiper le flou terminologique qui persiste quant au concept de racines latines et grecques. En enseignement, il convient d'utiliser des termes précis pour ne pas nuire à l'apprentissage des élèves qui construisent leur connaissance de ces concepts qui peuvent être complexes et abstraits.

La recherche permet aussi d'identifier si les élèves décomposent en morphèmes les mots formés de racines latines et grecques pour accéder à leur sens ou s'ils y accèdent dans leur globalité. La

connaissance de cet accès lexical permet de mieux soutenir les élèves dans leur apprentissage d'un vocabulaire très présent dans leurs manuels scolaires et à la base des contenus disciplinaires enseignés.

Si les résultats de recherche permettent de montrer la contribution de la connaissance des racines latines et grecques au traitement de mots formés de ces racines, il serait possible pour les décideurs en éducation d'orienter les programmes à cet effet. Ainsi, une plus grande place pourrait être accordée à l'enseignement des racines latines et grecques en salle de classe. Pour le moment, la place des racines latines et grecques semble encore limitée, d'autant plus que les cahiers d'exercices ne proposent des activités morphologiques que de manière sporadique (Fejzo, 2015).

De plus, la présente recherche permet de documenter le niveau de connaissance des racines latines et grecques chez les élèves de 3^e cycle du primaire, mais aussi leur niveau de connaissance du vocabulaire disciplinaire. À ce jour, en contexte francophone, une seule recherche permet de renseigner sur le fait que les élèves de ce niveau connaissent peu le sens des racines latines et grecques malgré leur occurrence en français (Whissell-Turner et Fejzo, 2021). Advenant des résultats en faveur de la décomposition des mots formés de racines latines et grecques lors de l'identification des mots écrits, donc en faveur de l'apport de ces morphèmes dans la connaissance du vocabulaire disciplinaire, il serait avenant d'outiller les élèves dès la fin du primaire à mieux se servir de ces morphèmes riches de sens. Effectivement, la connaissance de ces morphèmes pourrait être un atout privilégié pour la poursuite des études secondaires et postsecondaires où la proportion des racines latines et grecques est de plus en plus grande (Bar-Illan et Berman, 2007).

Pour terminer, bien que la présente recherche soit menée en français langue première, les racines latines et grecques sont au cœur de la formation des mots dans diverses langues romanes (p. ex. italien, espagnol) et même dans certaines langues germaniques (p. ex. anglais, allemand) (Dal et Amiot, 2008). De ce fait, il est possible que les résultats de cette recherche fournissent des pistes prometteuses pour l'apprentissage de langues secondes, notamment du français langue seconde, contexte où la morphologie contribue davantage aux connaissances du vocabulaire des élèves du primaire dont le français n'est pas leur langue maternelle comparativement à leurs pairs francophones (Fejzo, 2021).

CHAPITRE 2

CADRE CONCEPTUEL

La présente étude porte sur l'identification des mots écrits composés formés de racines latines ou grecques. Pour pouvoir traiter en profondeur de cette composante de la lecture, plusieurs concepts sont à définir et des modèles théoriques sont présentés. Dans un premier temps, la construction des mots en français est expliquée dans une perspective linguistique permettant de définir les concepts centraux de la recherche (2.1). Dans un deuxième temps, différentes catégorisations des types de vocabulaire sont présentées afin de mieux cerner le type de vocabulaire ciblé dans le cadre de cette étude, le vocabulaire disciplinaire (2.2). Dans un troisième temps, cela mènera à parler de la façon dont la connaissance du vocabulaire se développe (2.3). Dans un quatrième temps, l'identification des mots écrits, étape fondamentale à la lecture de mots, sera définie et des modèles qui tiennent compte de la morphologie dans le traitement des mots seront présentés (2.4). Dans un cinquième temps, des instruments de mesure disponibles pour mesurer l'identification des mots écrits sont décrits afin d'orienter les choix méthodologiques (2.5). Dans un sixième temps, les facteurs influençant le développement de la connaissance du vocabulaire sont rapportés (2.6) avant de recenser, dans un septième temps, les études empiriques qui ont porté sur le traitement morphographique afin de permettre d'identifier les facteurs qui influencent l'identification des mots polymorphémiques (2.7) Le cadre conceptuel se termine en énonçant les questions et les hypothèses de la recherche (2.8).

2.1 La construction des mots en français

Cette section présente divers concepts qui permettent de circonscrire l'objet d'étude. Le premier concept présenté est celui du *mot* (2.1.1) suivis de ceux de *morphologie* et de *morphèmes* (2.1.2). Ces définitions permettent de passer par les définitions des concepts de *préfixes*, de *suffixes*, de *racines*, de *radical* et de *base* afin de sélectionner la définition de *racine latine et grecque* choisie pour cette recherche (2.1.3). Enfin, les procédés de formation des mots en français sont brièvement décrits en portant une attention particulière à la composition, soit le procédé de formation des mots privilégié avec les racines latines et grecques (2.1.4). Cette information aura une incidence, comme il sera démontré dans la section *Méthodologie*, sur le choix des items des différentes mesures (3.3).

2.1.1 Une définition du concept *mot*

La définition du concept *mot*, central à ce projet, « est ambigu[ë] et son usage en linguistique peut prêter à confusion » (Polguère, 2016, p. 53). Par exemple, on peut se demander si *chaton* et *chatons* sont deux mots distincts ou deux formes différentes du même mot. En linguistique, on précisera que *chaton* et *chatons* sont deux mots-formes du même lexème. On définit un *mot-forme* comme un signe linguistique autonome dans la langue doté de cohésion interne, soit qui ne peut pas être séparé d'un autre mot-forme (Polguère, 2016). Le *lexème* est l'unité linguistique regroupant tous les mots-formes qui ne se distinguent que par leurs marques flexionnelles, soit les marques grammaticales (Polguère, 2016). Dans le cadre de cette thèse, dans le but d'alléger le texte, le terme *mot* sera utilisé pour référer aux mots-formes. Ainsi, *chaton* et *chatons* sont considérés comme deux mots.

2.1.2 Une définition des concepts *morphologie* et *morphèmes*

L'étude de la forme des mots, de leur structure, est la *morphologie* (Huot, 2005), du grec *morph* et *log* qui signifient respectivement « forme » et « étude » (Cellard, 2013). Les mots sont composés d'un ou plusieurs *morphèmes*, soit les plus petites unités linguistiques porteuses de sens (Huot, 2005) (p. ex. *irrespectueux* = *in-* + *respect* + *-eux*).

Selon leurs particularités linguistiques, les morphèmes sont de deux types : les affixes et les racines. Les *affixes* sont des morphèmes lexicalement non autonomes qui se placent respectivement avant ou après la racine d'un mot (Huot, 2005; Polguère, 2016) et sont respectivement désignés *préfixes* et *suffixes*. Les préfixes (p. ex. *co-*, *é-*, *dé-*, *re-*) renferment un sens très général (directionnel, associatif, privatif ou négatif) (Huot, 2005). Comme les préfixes, les suffixes sont des unités de sens à interprétation très générale. Ils sont utilisés, notamment, pour déterminer la classe grammaticale d'un mot (Huot, 2005). Il existe deux types de suffixes : flexionnels et dérivationnels. Les *suffixes flexionnels* consistent en les marques grammaticales ajoutées aux racines (p. ex. le *-s* marquant le pluriel, les terminaisons verbales comme le *-ent*, le *-e* du féminin, etc.). Ils peuvent servir à marquer le genre, le nombre, la personne, le temps et le mode (Huot, 2005). Les *suffixes dérivationnels* servent à construire des mots polymorphémiques dérivés (p. ex. *-eux*, *-age*, *-ion*, *-ment*, *-té*) (Huot, 2005).

La *racine* est reconnue pour désigner le sens essentiel du mot (Dubois *et al.*, 2002; Huot, 2005). Par exemple, dans le mot *fillette*, la racine *fille* renvoie directement au construit du mot, dans ce cas-ci à la personne. Le suffixe *-ette*, quant à lui, sert plutôt à qualifier ou à caractériser cette personne (p. ex.

fillette = « une fille qui a la caractéristique d’être petite »). La racine est l’élément central du mot. La linguistique indo-européenne réfère à la racine comme étant d’abord l’identité du mot, mais aussi une unité monosyllabique (Dubois *et al.*, 2002; Huot, 2005). Plus loin, on verra qu’un autre terme – *radical* – est utilisé pour désigner les unités polysyllabiques porteuses du sens principal du mot. En français, la racine est un morphème insécable, donc qui ne peut pas être réduit à une plus petite unité de sens (p. ex. *jeune*). La liste des mots qui contiennent la même racine forme ce qu’on appelle une *famille morphologique* (p. ex. *jeunesse, jeunot, etc.*).

La racine peut être accompagnée d’un allongement, formé d’un ou deux phonèmes, qui permet la formation des mots en français. La racine et son allongement forment ce qu’on appelle le *thème* (Huot, 2005). Par exemple, l’allongement *at*, dépourvu de sens, n’est ni un préfixe ni un suffixe. Il est ajouté à la racine *form* pour construire les mots tels que *formation* (*form + at + -ion*) et *formatif* (*form + at + -if*) (Huot, 2005). Prenons un autre exemple, mais cette fois-ci formé de racines grecques. Dans le mot *hydrographe*, le *o* à la jonction entre les racines *hydr* et *graph* est l’allongement de la racine *hydr*. Plus précisément, il s’agit d’une voyelle thématique⁴. Cet ajout est nécessaire en morphophonologie à la rencontre de deux consonnes (p. ex. *hydr + o + graph[e]*). Dans cet exemple, le *e* est un *e* graphique muet⁵ (Catach, 2016).

Dans la littérature scientifique, deux autres termes sont souvent utilisés interchangeablement au lieu du terme *racine* (Huot, 2005). Il s’agit des termes *radical* et *base*. Selon Huot (2005), le *radical* est un morphème polysyllabique insécable et porteur du sens principal du mot (Huot, 2005). Le terme *radical* serait aussi utilisé pour les morphèmes savants d’origine latine ou grecque (qu’on identifiera plus tard comme *racine latine ou grecque*), qu’ils soient monosyllabiques ou polysyllabiques (p. ex. *hydr, mens, bi, extra, pré, post, penta, etc.*). Bien que ces morphèmes particuliers soient souvent confondus avec les préfixes, puisqu’ils étaient jadis des prépositions et qu’ils sont souvent placés en

⁴ Pour certains auteurs (p. ex. Dal et Amiot, 2008), cette voyelle thématique est aussi appelée *joncteur*. Dans la présente thèse, lorsque nous souhaiterons mettre l’accent sur la structure des mots, la voyelle thématique sera séparée de la racine (p. ex. *hydr-o-graph-ie*). Notons que la voyelle thématique *i* est généralement préférée à la suite des racines latines (p. ex. *agr-i-cult-ure, équ-it-able*) alors que le *o* suit habituellement les racines grecques (p. ex. *hydr-o-graph-ie, andr-o-log-ie*). Comme les voyelles thématiques ne sont pas porteuses de sens, elles contribuent à réduire la transparence du mot et la facilité avec laquelle ce dernier serait décomposé (Dressler, 2006).

⁵ Lorsque nécessaire, le *e* graphique muet sera présenté entre parenthèses afin de le distinguer de la racine à laquelle il s’attache.

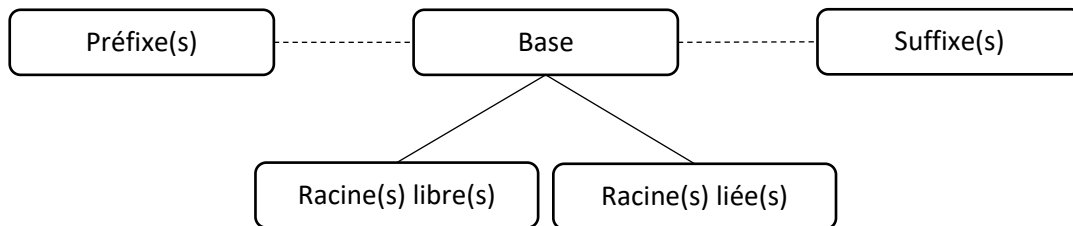
début de mot (p. ex. *hydrographie, mensuel, bipède, extraordinaire, etc.*), ils sont considérés comme des radicaux puisqu'ils renferment un sens concret (Huot, 2005). Ces radicaux savants ont le même rôle que les racines, soit celui de former des mots polymorphémiques, mais seraient plus productifs (Huot, 2005). En contrepartie, toujours en linguistique, le terme *radical* est utilisé légèrement différemment. Effectivement, Dubois et ses collaborateurs (2002) attribuent ce terme aux différentes formes orthographiques d'une racine après manipulations phonologiques (p. ex. la racine verbale *ven* de *venir* compte deux radicaux, soit *ven* et *vien*). Les radicaux sont ici considérés comme les points de départ des mots auxquels sont ajoutés les affixes (Dubois *et al.*, 2002). Enfin, le terme *base* est utilisé pour désigner tout ce qui précède le suffixe final d'un mot. Par exemple, *immobile* serait la base du mot suffixé *immobiliser*, tandis qu'*immobiliser* serait la base du mot *immobilisation*. À d'autres moments, la base permet d'identifier la plus simple dérivation d'un mot autonome lexicalement (p. ex. *mobile*) (Crosson et McKeown, 2016). Afin de distinguer ces deux termes du terme *racine* utilisé dans la présente étude, cette dernière définition sera utilisée pour le terme *base*. Le *radical* sera considéré comme les différentes formes orthographiques d'une même racine (p. ex. *mov, mob, mot*).

En comparaison aux définitions théoriques et linguistiques exposées précédemment, Huot (2005) précise que « d'un point de vue morphologique, une racine se définit comme la plus petite suite de sons pourvus d'un sens, et qui assurent à une unité lexicale son individualité parmi l'ensemble des autres unités du lexique » (p. 51). Dans cette optique, aucune distinction terminologique ne subsisterait pour désigner l'origine populaire ou savante des morphèmes porteurs du sens principal du mot ni pour les distinguer selon le nombre de syllabes. La principale distinction surviendrait quant à l'autonomie lexicale. Une racine peut être libre ou liée. Une *racine libre* est autonome lexicalement, c'est-à-dire qu'elle constitue un mot en soi (p. ex. *eau, grenouille*) (Crosson et McKeown, 2016). La *racine liée*, aussi appelée racine secondaire en linguistique (Dubois *et al.*, 2002), n'est pas autonome dans la langue bien qu'elle soit porteuse d'un sens qui lui est propre (p. ex. j'ai *hexa crayons). Elle doit en tout temps être juxtaposée à une autre racine ou à un affixe pour former un mot (p. ex. *inclus, exclus, hexagone*). C'est cette définition de *racine* qui est adoptée pour désigner le ou les morphèmes porteurs du sens principal

du mot dans le cadre de cette thèse (p. ex. *grenouille* ; *maison* dans *maisonnette* ; *télé*⁶ et *phone* dans *téléphone*).

À la lumière des définitions précédentes, il est possible de fournir une conceptualisation de la structure des mots en français telle que présentée en Figure 2.1. Les mots du français comportent donc tous obligatoirement au moins une racine libre ou liée formant la base du mot et à laquelle peuvent s'ajouter des préfixes ou des suffixes.

Figure 2.1 Conceptualisation de la structure des mots en français



2.1.3 Une définition du concept *racines latines ou grecques*

Les termes principaux relatifs à la morphologie étant définis dans la section précédente, il est maintenant possible de situer le concept central de l'étude, les racines latines ou grecques. Pour arriver à sélectionner une définition de ce concept cultivé par un flou terminologique, la variété des définitions ou appellations présentes dans les écrits scientifiques sera d'abord exposée. Cela permettra de constater qu'il est primordial de définir ce concept afin d'éviter toute confusion, et ce autant d'un point de vue linguistique que didactique. Afin de confirmer le choix définitionnel, l'élagage de certains termes utilisés sera fait en démontrant que quelques terminologies utilisées vont à l'encontre du sens même de ces termes. Cette procédure permettra de circonscrire l'objet d'étude de cette thèse doctorale, mais aussi d'assurer l'utilisation d'un métalangage suffisamment précis en contexte d'enseignement.

⁶ Il arrive que la forme orthographique *télé* soit utilisée comme un mot, ce qui pourrait laisser penser qu'il s'agit d'une racine libre. Cependant, lorsque cette forme orthographique est utilisée seule, elle ne signifie pas 'loin'. En fait, la forme orthographique *télé* en tant que mot est une troncation de *télévision*. Cette troncation permet aussi de former des mots composés (voir section 2.1.4). Par exemple, le mot *téléphage* réfère à une personne qui consomme beaucoup de télévision et non à une personne qui mange de loin.

En contexte francophone, les quelques linguistes s'intéressant aux racines latines et grecques ne s'entendent pas sur la terminologie à utiliser. Effectivement, certains linguistes (p. ex. Bouffartigue, 1983; Bouffartigue et Delrieu, 1981; Cellard, 2013) préfèrent l'utilisation du terme *racine* correspondant au terme *root* utilisé dans plusieurs écrits anglophones (p. ex. Crosson et McKeown, 2016; Crosson et Moore, 2017; Green, 2015). D'autres auteurs (p. ex. Amiot, 2010) scindent ces éléments en deux catégories sur la base de leur rôle syntaxique dans la langue : affixes (p. ex. *hyper*, *hypo*, *archi*, *sub*) et constituants de composés néoclassiques (p. ex. *hydr*, *pyr*, *poly*, *multi*). Aussi, l'utilisation des termes *préfixe* et *suffixe* est observée, notamment dans les ouvrages de référence où cette terminologie est source de confusion et d'inconstance (p. ex. Druide informatique inc., s.d.; Larousse, s.d.). Pour donner un exemple de cet inconstance terminologique, la racine numérale *hexa* est parfois désignée *préfixe* (Druide informatique inc., s.d.), *élément* (Le Robert, 2021), *racine* (Cellard, 2013) ou encore *élément formant* (Analyse et traitement informatique de la langue française [ATILF], s.d.). Ailleurs, il arrive que la racine *hexa* ne soit pas du tout répertoriée (Larousse, s.d.). Le cas de figure pour cette racine n'est pas unique et se répète pour de nombreuses racines latines et grecques.

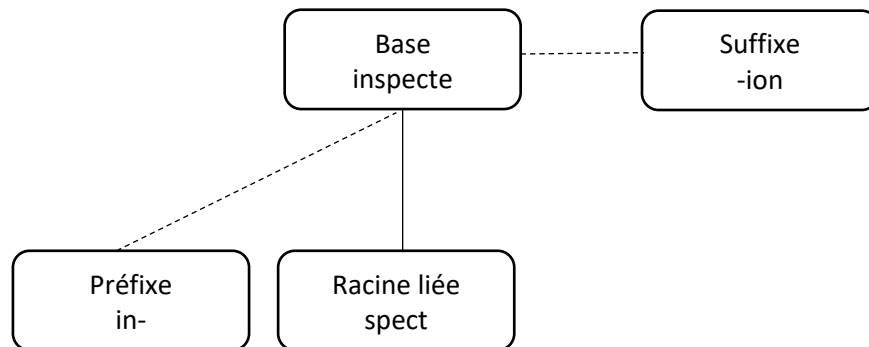
Dans les écrits anglophones, le problème est le même : le concept de racines latines et grecques porte diverses appellations sans toujours être clairement défini. Certains auteurs utilisent le terme *racine* (p. ex. Boyle, 2009; Freeman *et al.*, 2019; Green, 2015; Templeton, 1983), d'autres *mot racine* [ma traduction de *word root*] (p. ex. Manyak *et al.*, 2018), ou encore *partie de mots* [ma traduction de *word part*] (p. ex. Cabral, 2015), *morphème* (p. ex. Wall, 2016) ou *item d'origine latine* [ma traduction de *item of Latinate origin*] (p. ex. Bar-Ilan et Berman, 2007). En linguistique, Henry (2010) utilise *préfixe*, *suffixe* et *racine* pour les morphèmes latins, tandis que les morphèmes grecs sont identifiés comme des *formes de combinaison* [ma traduction de *combining form*]. Dans les travaux de Rasinski et ses collègues (Rasinski *et al.*, 2017; Rasinski *et al.*, 2011; Rasinski *et al.*, 2007, 2008), les auteurs utilisent de façon interchangeable les termes *mot racine*, *racine* ainsi que *partie de mots*. Toutefois, contrairement aux auteurs précédents, leur utilisation du terme *racine* ne renvoie pas au morphème porteur du sens principal du mot, mais plutôt aux plus petites unités porteuses de sens dans le mot. Donc, selon ces auteurs (Rasinski *et al.*, 2008), il existerait trois types de racines : les préfixes, les bases et les suffixes. Leur utilisation du terme *racine* renvoie à la définition traditionnelle d'un morphème, ce qui alimente considérablement le flou terminologique autour de ce concept.

Afin de dissiper le flou terminologique qui plane autour du concept de racines latines et grecques, les appellations erronées seront réfutées en analysant des mots formés de racines latines et grecques à la lumière des définitions présentées précédemment. Commençons par le mot d'origine grecque *hydrophobe* composé des morphèmes *hydr* et *phob*, qui signifient respectivement « eau » et « peur ». En s'en tenant aux définitions précédentes (section 2.1.2), les affixes ont des positions respectives dans le mot, soit avant ou après la racine (Huot, 2005; Polguère, 2016). C'est donc dire qu'un affixe ne peut pas être désigné à la fois préfixe et suffixe. Pour le mot *hydrophobe*, on pourrait croire qu'il est composé d'un préfixe grec, *hydr* ou **hydro* comme dans *hydroélectricité* ou *hydrofuge*, et d'un suffixe grec, *phob(e)* comme dans *agoraphobe* ou *arachnophobe*. En effet, la présence fréquente de ces morphèmes en début et en fin de mot pourrait laisser croire qu'il s'agit d'affixes. Toutefois, les morphèmes *hydr* et *phob* n'occupent pas toujours des places fixes dans les mots (p. ex. *déshydrater*, *phobie*). D'ailleurs, le mot *phobie* est un autre exemple en faveur de la rétention du terme racine pour ce morphème, *phob*, puisqu'il permet de former un mot dérivé suffixé. Effectivement, les lettres *ie*, souvent attribué à tort comme une partie intégrante de plusieurs racines (p. ex. *phobie*, *graphie*), ne sont en fait que l'issue du processus de dérivation par l'adjonction du suffixe nominal *-ie*. De plus, laisser entendre que les morphèmes *hydr* et *phob* seraient respectivement un préfixe et un suffixe est inconcevable puisque le mot *hydrophobe* ne serait formé d'aucune racine, ce qui est impossible puisque la racine est nécessaire à la formation d'un mot. Aussi, chacune de ces racines renferme un sens plein et concret, contrairement aux préfixes et aux suffixes qui renvoient plutôt à des interprétations plus générales (Huot, 2005). Enfin, les morphèmes *hydr* et *phob* sont interdépendants. En effet, retirer un des deux morphèmes entraînerait non seulement une perte de sens, mais entraînerait aussi une erreur lexicale puisqu'aucun de ces morphèmes n'est autonome, étant donné que les racines qui sont originaires du latin ou du grec sont toujours liées (Crosson et McKeown, 2016). Cette caractéristique commune aux affixes pourrait expliquer en partie la confusion qui peut s'installer. Or, comme démontré, il s'agirait d'une erreur terminologique et conceptuelle de désigner ces racines grecques, *hydr* et *phob*, par les termes *préfixe* ou *suffixe*.

Afin de démontrer que les définitions présentées dans les précédentes sections s'appliquent aux mots formés de racines latines et grecques, la structure de certains mots est exposée en exemple. Le mot *inspection* est formé de la base *inspecte*, formée elle-même par le préfixe *in-* et la racine liée d'origine latine *spect*, ainsi que du suffixe nominal *-ion* (Figure 2.2). De la base du mot, il est possible de dresser la *famille morphologique* du mot, soit une liste de mots morphologiquement reliés qui partagent tous la

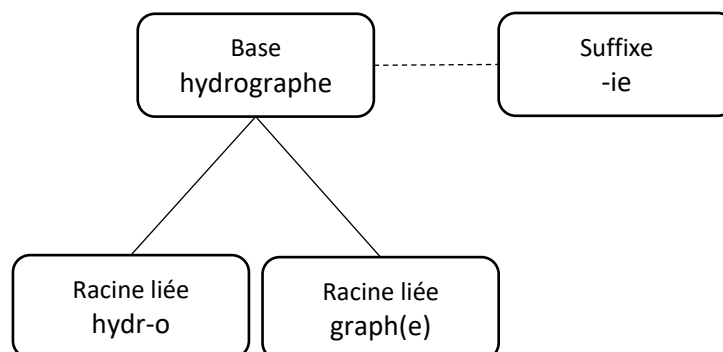
même base. La famille morphologique serait composée des mots *inspecteur* et *réinspection*, par exemple. À partir de la racine liée, il est aussi possible de dresser une liste de mots qui contiennent tous la même racine liée, ce qui constitue la *famille étymologique*. En plus de compter tous les mots de la famille morphologique, la famille étymologique de la racine liée latine *spect* comprendrait aussi des mots tels que *spectacle*, *spectateur* et *prospection*.

Figure 2.2 Conceptualisation de la structure du mot *inspection*



Pour donner un autre exemple, le mot *hydrographie* est formé de la base *hydrographe*, elle-même formée de deux racines grecques liées, *hydr* (accompagnée de sa voyelle thématique *o*) et *graph*, et du suffixe *-ie* (Figure 2.3). La famille morphologique est composée des mots *hydrographique* et *hydrographe*, par exemple. Comme deux racines liées sont répertoriées, il est possible de dresser parallèlement deux familles étymologiques. La famille étymologique de la racine liée grecque *hydr* comprend aussi les mots *hydrater*, *hydravion* et *hydratation* par exemple. La deuxième famille étymologique est formée à partir de la racine liée grecque *graph* : *calligraphie*, *graphique*, *graphiste*, etc.

Figure 2.3 Conceptualisation de la structure du mot *hydrographie*



Ainsi, à la lumière de la démonstration précédente, le terme *racine latine ou grecque* désigne le morphème lié d'origine latine ou grecque porteur du sens principal du mot (Crosson et McKeown, 2016; Huot, 2005).

2.1.4 Les procédés de formation des mots en français

En français, environ 80 % des mots sont polymorphémiques (Rey-Debove, 1984). En anglais, cette proportion serait moindre, estimée à près de 55 % (Casalis *et al.*, 2015). La place de la morphologie est donc plus importante en français qu'en anglais, d'où l'intérêt de s'y pencher. Ainsi, les trois procédés à la base de la formation des mots en français et leurs caractéristiques seront tour à tour présentés, soit la flexion, la dérivation et la composition, en s'attardant davantage à la composition néoclassique, étant donné que ce type de composition caractérise en grande partie les mots formés de racines latines et grecques. La description de ce type de procédé de formation permettra de sélectionner les items pour étudier le traitement morphographique des mots formés de racines latines et grecques chez les élèves du 3^e cycle du primaire. Enfin, des cas qui combinent plus d'un procédé de formation des mots seront exposés. Effectivement, plusieurs mots dans la langue, y compris des mots comprenant des racines latines et grecques, sont formés par la combinaison de plusieurs procédés.

Le premier procédé de formation des mots en français est la flexion. Ce procédé consiste en l'ajout de suffixes flexionnels à une racine (Huot, 2005) dans le but de respecter les règles syntaxiques et sémantiques d'une phrase (Kuo et Anderson, 2006) (p. ex. les chatons miaulent). L'ajout d'un suffixe flexionnel à une racine pour créer un mot fléchi ne modifie pas la classe grammaticale du mot.

La dérivation est le deuxième procédé de formation des mots en français. Il s'agit aussi du procédé le plus productif (Casalis *et al.*, 2015). Lors de la formation d'un mot dérivé, un préfixe ou un suffixe dérivationnel est ajouté à une racine ou à une base (Huot, 2005). La dérivation peut être formée à partir d'une racine libre (p. ex. *maisonnette* = *maison* + *-ette*), d'une racine liée (p. ex. *graphie* = *graph* + *-ie*), ou d'une base (p. ex. *équilibriste* = *équilibre* + *-iste*). Un mot dérivé peut aussi compter de nombreux affixes (p. ex. *inévitablement* = *in-* + *évite* + *-able* + *-ment*).

La dérivation peut aussi être affectée par deux types de transparence, la transparence phonologique et la transparence orthographique (Carlisle et Stone, 2005). Par exemple, le mot *amoureux* est transparent phonologiquement et orthographiquement puisque ni la prononciation ni l'orthographe de la racine

amour ne sont altérées à la suite de l'ajout du suffixe *-eux*. Pour certains mots, seule la transparence phonologique est réduite (p. ex. *matin, matinée*), alors que d'autres mots dérivés sont à la fois phonologiquement et orthographiquement opaques (p. ex. *soin, soigneur*). La transparence phonologique et orthographique influence l'identification des mots écrits. Effectivement, un mot plus opaque phonologiquement est lu plus lentement qu'un mot transparent (Carlisle et Stone, 2005).

Le troisième et dernier procédé de formation des mots en français est la composition⁷. Elle consiste en la jonction de deux ou plusieurs racines libres ou liées (Kuo et Anderson, 2006). Par exemple, les mots *millefeuille* et *psychologue* sont tous deux des mots composés. Les composés sont présents sous différentes formes, soit soudées (p. ex. *téléphone, millepatte*), séparées d'un trait d'union (p. ex. *porte-parole, cerf-volant*), d'un espace (p. ex. *papier journal, pomme de terre*)⁸ ou, dans de plus rares cas, d'une apostrophe (p. ex. *presqu'île*) (Leminen et al., 2018).

En français, il existe trois types de mots composés (Dressler, 2006). Le principal type est celui formé de deux mots autonomes dans la langue, deux racines libres, parfois appelé *composition ordinaire* (Dal et Amiot, 2008) (p. ex. *porte-parole, millepatte*). Le deuxième type de mots composés, en français, est formé par accourcissement, ou troncation d'au moins un mot (Dal et Amiot, 2008). Par exemple, le mot *courriel* résulte de la jonction de deux mots tronqués, *courrier* et *électronique*. Enfin, les composés néoclassiques, parfois appelés *composés savants*, forment le troisième type de mots composés (Dal et Amiot, 2008; Dressler, 2006). La présente étude porte particulièrement sur ce dernier type de mots composés. Les composés néoclassiques sont formés d'au moins une racine liée d'origine latine ou grecque, non autonome lexicalement (Dal et Amiot, 2008; Dressler, 2006). Par exemple, le mot *télétravail* est formé d'une seule racine grecque, *télé*, et de la racine libre *travail* qui, elle, est autonome dans la langue. Le mot *téléphone*, quant à lui, est un mot composé néoclassique formé de deux racines liées grecques, *télé* et *phon(e)*. Les racines latines et grecques peuvent se trouver à gauche du mot composé (p. ex. *télétravail*), à droite (p. ex. *technophobe*) ou encore aux deux positions (p. ex. *téléphone*).

⁷ Comme la présente étude porte sur le traitement des mots tels que définis à la section 2.1.1, les phrases et expressions figées (p. ex. Gross, 1982) ne sont intentionnellement pas abordées.

⁸ Aussi appelés *locution* (Polguère, 2016).

En anglais, certains mots composés sont particuliers dans la mesure où ils sont formés d'au moins un morphème lié non productif, qui ne se retrouve dans aucun autre mot de la langue (Aronoff et Fudeman, 2023; Dressler, 2006). C'est le cas des mots comme *cranberry* et *raspberry*. Les unités *cran* et *rasp* ne sont pas des morphèmes productifs, même si l'étymologie explique leur présence (Aronoff et Fudeman, 2023; Dressler, 2006). Elles ne se présentent dans aucun autre mot. D'ailleurs, certains linguistes voient peu la pertinence de décomposer ce type de mots (Aronoff et Fudeman, 2023). Malgré que ces morphèmes soient peu fréquents, certains chercheurs s'y intéressent. C'est d'ailleurs le cas pour Solomyak et Marantz (2010), par exemple, qui comparent le traitement de racines libres, de racines liées et de racines dites uniques (non productives). Cependant, notre conceptualisation des types de mots composés nous permet d'analyser attentivement les items proposés et de constater que la plupart des racines uniques identifiées par les chercheurs sont des racines liées qui forment d'autres mots en anglais. Cette distinction permettra de mieux comprendre les résultats des recherches antérieures et d'analyser en profondeur leurs listes d'items.

Pour certains auteurs, pour être considéré comme un mot composé en français moderne, quel que soit le type de composition, chaque constituant doit être identifiable et lié au sens du mot entier (p. ex. Huot, 2005). Par exemple, le mot *dorénavant* (*d'or + en avant*) serait exclu bien que son étymologie puisse être intéressante. Cependant, dans d'autres ressources, ce mot peut être considéré comme un mot composé, comme c'est le cas dans le *Trésor de la langue française informatisé* (*dorénavant* TLFi) (Analyse et traitement informatique de la langue française [ATILF], s.d.). Dans le même ordre d'idées, les mots directement empruntés au latin ou au grec ne sont pas toujours considérés comme des composés néoclassiques pour certains auteurs (p. ex. Dal et Amiot, 2008), comme ils n'ont pas été directement construits dans la langue cible à l'aide de racines latines ou grecques (p. ex. *misogyne*, emprunté directement au grec). D'autre part, certains auteurs (p. ex. Huot, 2005) les incluent sur la base de leur interprétation unique et constante formée à l'aide de morphèmes distincts (p. ex. *misogyne* = *mis + o + gyn(e)* = « haine » + « femmes »). Sans alimenter ce débat ou prendre position dans le cadre de cette thèse doctorale, des mots composés dont l'analyse morphologique permet, à ce jour, de fournir des indices sur le sens total ou partiel du mot en entier seront utilisés.

De plus, la composition néoclassique est présente dans plusieurs langues romanes (p. ex. français, italien, espagnol), mais aussi germaniques (p. ex. anglais, allemand) (Dal et Amiot, 2008). D'ailleurs, les racines latines et grecques à la base de la composition néoclassique bénéficient même d'un statut

dit international en référence à leurs grandes similitudes orthographiques et phonologiques d'une langue à l'autre (p. ex. *biologie* en français, *biology* en anglais, *biología* en espagnol, *Biologie* en allemand, *biologia* en italien, etc.) (Dal et Amiot, 2008). Cet aspect est particulièrement intéressant pour la présente recherche. Effectivement, comme peu de recherches sur les racines latines et grecques ont été effectuées en contexte francophone, savoir que des similitudes sont observables entre les langues permettra d'explorer des recherches dans différents contextes linguistiques, dont l'anglais.

Pour continuer, un des morphèmes du mot composé transmet habituellement ses propriétés sémantiques et syntaxiques au mot entier. Il s'agit de la tête du mot composé⁹ (Dal et Amiot, 2008; Dressler, 2006). Dans les langues germaniques, comme l'anglais, la tête occupe toujours, sauf de rares exceptions, la position de droite dans le mot (p. ex. *blackboard*, où *board* constitue la tête du mot puisqu'il s'agit d'un type de tableau dont sa particularité est d'être noir) (Dressler, 2006; Libben, 2006). Dans les langues romanes, particulièrement en français, la tête peut se retrouver tantôt à gauche (p. ex. *oiseau-mouche*), tantôt à droite (p. ex. *beau-père*) (Jarema, 2006; Libben, 2006). Cependant, il arrive que les deux termes assurent cette fonction et qu'il soit impossible de distinguer comme tête un morphème plus que l'autre (p. ex. *sourd-muet*, *aigre-douce*) (Dressler, 2006). De façon générale, en composition ordinaire en français, la tête est placée à gauche (p. ex. *oiseau-mouche*, c'est un type d'oiseau qui rappelle la mouche), alors qu'elle est souvent à droite en composition néoclassique (p. ex. *morphologie*, c'est un type de science qui étudie, *log*, les formes, *morph*) (Dal et Amiot, 2008). En résumé, la composition néoclassique se distingue de la composition ordinaire par le fait qu'elle est composée d'au moins une racine d'origine latine ou grecque non lexicalement autonome dans la langue cible et que, lorsque présente, la tête se situe habituellement à droite plutôt qu'à gauche (Amiot, 2010).

⁹ Les composés peuvent être endocentriques ou exocentriques (Dressler, 2006). Un composé est *endocentrique* lorsque la tête est directement présente dans le mot composé (p. ex. *oiseau-mouche* pour désigner un type d'oiseau). Un composé *exocentrique* renvoie aux mots dont la tête doit être inférée (p. ex. *porte-parole* pour désigner une personne qui porte sa voix pour d'autres). Pour chaque mode (endocentrique et exocentrique), Libben (1998) énumère quatre degrés de transparence sémantique pour les mots morphologiquement composés en s'appuyant sur le sens de chacun des morphèmes. Le premier niveau est le plus transparent. Il contient les mots formés de deux morphèmes qui contribuent directement au sens du mot global (p. ex. *porte-manteau*). Le deuxième niveau considère les mots dont la tête est transparente alors que le second morphème est plus opaque (p. ex. *supermarché*). Au troisième niveau, c'est plutôt la tête qui est opaque alors que le second morphème est plus transparent (p. ex. *téléverser*). Enfin, le dernier degré de transparence des mots composés, soit le niveau le plus opaque, contient des mots où les deux morphèmes sont opaques et ne peuvent servir à restituer le sens global du mot entier (p. ex. *croquemonsieur*).

Pour terminer, un mot polymorphémique peut résulter de plus d'un procédé de formation. On peut, par exemple, ajouter des suffixes flexionnels à des mots dérivés (p. ex. *maisonnettes* = *maison* + *-ette* + *-s*) ou composés (p. ex. *millefeuilles* = *mille* + *feuille* + *-s*). Notamment avec la composition néoclassique (Amiot, 2010; Dressler, 2006), la combinaison de la composition et de la dérivation est fréquente (p. ex. *hydrographie* = *hydr* + *o* + *graph* + *-ie*). Enfin, des mots peuvent aussi avoir été soumis aux trois procédés de formation (p. ex. *hydrographies* = *hydr* + *o* + *graph* + *-ie* + *-s*).

Alors que certains auteurs (p. ex. Ferrand, 2001) ne considèrent que deux procédés principaux de formation des mots, la flexion et la dérivation (qui, elle, contiendrait aussi les mots composés), il est généralement admis que la flexion, la composition et la dérivation sont trois procédés distincts à la base de la formation des mots polymorphémiques (p. ex. Huot, 2005; Kuo et Anderson, 2006). D'autres (p. ex. Dressler, 2006) différencient le procédé à la base de la forme du mot et celui qui permet d'y attribuer un sens. Par exemple, le mot téléphonique (*télé* + *phon* + *-ique*) est un mot dérivé formé à partir du composé néoclassique *téléphone*. Dans ce cas, selon Dressler (2006), bien que la forme du mot *téléphonique* relève partiellement de la composition, son sens relèverait plutôt de la dérivation (*téléphonique* = « relatif au téléphone »). Dans le cadre de cette thèse doctorale, lors du choix des items, tous les procédés au sein de la formation des mots polymorphémiques seront analysés. Cette décision permettra de choisir des mots qui sont uniquement composés, donc sans ajouts d'affixes ni flexionnels ni dérivationnels. Ainsi, ce contrôle permettra d'associer les résultats obtenus au traitement des racines liées en composition plutôt qu'à l'influence des différents procédés de formation des mots.

2.2 Les différentes catégorisations des types de vocabulaire

En linguistique, le terme *vocabulaire* est associé à la liste de tous les mots utilisés dans un texte ou un corpus (Dubois *et al.*, 2002). Le terme *lexique* réfère plutôt à un ensemble d'unités qui n'est pas restreint à un texte ou à un individu, par exemple l'ensemble des mots d'une langue (Dubois *et al.*, 2002; Polguère, 2016). Or, selon Legendre (2005), on peut parler autant de vocabulaire que de lexique en référant à l'ensemble des mots connus d'un individu à un moment donné. En psycholinguistique, l'utilisation du terme *vocabulaire* est plutôt répandue lorsque les chercheurs anglophones réfèrent aux mots (p. ex. Baumann *et al.*, 2012; Baumann *et al.*, 2003; Baumann *et al.*, 2002; Baumann et Graves, 2010; Beck *et al.*, 2013; Hiebert, 2020; Hiebert et Cervetti, 2012; Kirby et Bowers, 2018; Marzano, 2020; McKeown *et al.*, 2012; McKeown *et al.*, 2018). En contexte francophone, on rencontre autant le terme *vocabulaire* (p. ex. Fejzo *et al.*, 2014; Tremblay *et al.*, 2018) que *lexique* (p. ex. Anctil *et al.*, 2018; Drouin, 2007; Tremblay,

2015). Pour éviter un flou terminologique, dans le cadre de cette thèse, le vocabulaire sera considéré comme une composante du lexique. Le terme *vocabulaire* réfèrera donc à la liste de mots composant un texte, aux mots utilisés dans un domaine spécifique ou encore connus d'un individu. L'utilisation du terme *lexique* réfèrera donc à l'ensemble des mots d'une langue, en l'occurrence le français.

Il existe quatre types de vocabulaire catégorisés selon les processus réceptif (compréhension) et expressif (production) ainsi que deux modes d'expression (écrit et oral) (Baumann *et al.*, 2012; Nation, 2001; Pearson *et al.*, 2012; Pikulski et Templeton, 2004; Templeton, 2004). Le vocabulaire réceptif renvoie à la compréhension des mots nécessaires à la réception d'un message oral ou écrit et à l'extraction de son sens. Le vocabulaire expressif, quant à lui, concerne plutôt les mots utilisés pour émettre un message oral ou écrit.

Le type de vocabulaire présent dans les textes écrits (vocabulaire réceptif à l'écrit) a fait l'objet de divers classements par les chercheurs, notamment basés sur la fréquence d'apparition du mot, ou de sa famille morphologique, dans les textes. Deux principaux classements basés sur des travaux anglophones sont brièvement présentés : celui de Beck et ses collègues (2002; 2013) et celui de Baumann et Graves (2010).

Beck et ses collègues (2002; 2013) prennent appui sur la fréquence d'utilisation des mots pour diviser leur classement du vocabulaire en trois tiers. Le premier tiers comprend les mots très fréquents, surtout à l'oral, ce qui correspond au *vocabulaire courant* (p. ex. *bébé, horloge*). Ce type de vocabulaire s'acquiert grâce à l'exposition aux interactions sociales et ne nécessite aucun enseignement explicite pour les locuteurs natifs (Beck *et al.*, 2002). Le deuxième tiers comprend des mots assez fréquents à l'oral, mais utilisés par des locuteurs qui maîtrisent la langue (p. ex. *industriel, coïncidence*). Enfin, le troisième tiers comprend les mots à faible occurrence qui se retrouvent surtout à l'écrit, dans les textes à domaine spécifique (p. ex. *péninsule, isotope*).

Pour leur part, Baumann et Graves (2010) distinguent deux grandes catégories de vocabulaire : le vocabulaire courant et le vocabulaire scolaire [ma traduction de *academic vocabulary*]. Selon ces auteurs, le vocabulaire scolaire peut être à nouveau subdivisé. Effectivement, les textes scolaires peuvent contenir des mots issus du vocabulaire transdisciplinaire [ma traduction de *general academic vocabulary*] ou du vocabulaire disciplinaire [ma traduction de *domain-specific academic vocabulary*]. Le *vocabulaire transdisciplinaire* comprend des mots qui peuvent se retrouver dans différentes disciplines (p. ex. *analyse, représentation, souligner*). Plusieurs de ces mots sont polysémiques, revêtant différents sens

selon la discipline dans laquelle ils apparaissent ou selon le contexte (p. ex. *valeur* en éthique et en mathématiques). En revanche, les mots issus du *vocabulaire disciplinaire* se retrouvent presque exclusivement dans des textes spécifiques à un domaine d'apprentissage et ne renferment qu'un seul sens (p. ex. *hydrographie, photosynthèse, polygone*).

Dans le cadre de cette étude, la définition de vocabulaire disciplinaire de Baumann et Graves (2010) est retenue, soit les mots associés à un domaine spécifique d'apprentissage et référant à un concept précis. La thèse porte sur le vocabulaire disciplinaire puisqu'une grande partie de ces mots sont formés de racines latines (p. ex. *carbone*) et grecques (p. ex. *photosynthèse*) (Fang, 2006; Fazio et Gallagher, 2014; Nagy et Townsend, 2012; Padak *et al.*, 2008). Effectivement, comme le latin, lui-même influencé du grec notamment en mathématiques et en sciences, a occupé une place de choix en éducation environ jusqu'aux années 1960, le vocabulaire disciplinaire reste grandement marqué par cet héritage linguistique (Boyle, 2009). Aussi, le vocabulaire disciplinaire est souvent polymorphémique (Nagy et Townsend, 2012; Zoski *et al.*, 2018), c'est-à-dire composé de plus d'un morphème. En contexte anglophone, selon une analyse des mots utilisés dans les documents ministériels du cours de science au primaire, près de 80 % des mots seraient polymorphémiques (Fazio et Gallagher, 2014).

D'ailleurs, comme pour le vocabulaire transdisciplinaire, on retrouve souvent des mots nominalisés parmi ces mots polymorphémiques, ce qui ajoute à la complexité (Fang, 2006; Nagy et Townsend, 2012; Zoski *et al.*, 2018). La nominalisation est le fait de transformer un adjectif ou un verbe en nom (Fang, 2006; Nagy et Townsend, 2012). Ce nom ainsi créé est souvent abstrait. D'ailleurs, lors d'une tâche de complétion de phrases à l'aide de choix de réponses, l'étude de Nippold et Sun (2008) a permis de conclure que les élèves de 10 à 13 ans performant moins bien avec les noms dérivés (p. ex. *locomotion*) qu'avec les adjectifs dérivés (p. ex. *moléculaire*) tirés de leurs manuels scolaires. Pour les choix méthodologiques de la présente étude, aucun mot dérivé ne sera sélectionné afin de contrôler le procédé de formation des mots : seuls les mots composés seront retenus.

De plus, le vocabulaire disciplinaire compte de nombreux mots abstraits (Nagy et Townsend, 2012). Contrairement au vocabulaire courant utilisé par les élèves (p. ex. *arbre*), le vernaculaire n'est pas suffisant pour exprimer les concepts issus du vocabulaire disciplinaire, notamment les concepts scientifiques (Fazio et Gallagher, 2014). Souvent, aucun synonyme ne peut être utilisé pour remplacer un terme scientifique en conservant le même sens (Fang, 2006). La compréhension de ce type de

vocabulaire est donc intimement liée aux connaissances conceptuelles de la discipline associée (Fazio et Gallagher, 2014). Enfin, le vocabulaire disciplinaire se caractérise par sa faible fréquence (Green, 2008; Nagy et Townsend, 2012). En effet, les élèves, et même les adultes, sont rarement confrontés à ce type de vocabulaire à l'oral qui est surtout présent à l'écrit (Beck *et al.*, 2002; Beck *et al.*, 2013). Le vocabulaire disciplinaire est donc un type de vocabulaire particulier qui se définit par une combinaison de caractéristiques : sa structure morphologique complexe, sa faible fréquence, son caractère abstrait ainsi que la présence de racines latines et grecques.

2.3 Le développement de la connaissance du vocabulaire

Selon l'hypothèse de qualité lexicale (Perfetti, 2007), le développement de la connaissance des mots est essentiel à l'acte de lire. Le lecteur doit développer des représentations lexicales de qualité qui comprennent des représentations précises quant à la forme (orthographique et phonologique), au sens (morphosyntaxique et sémantique) ainsi qu'à l'usage des mots. Des représentations précises et interreliées faciliteraient l'accès lexical (Perfetti, 2007). L'interconnexion entre les dimensions sémantique, phonologique, orthographique et syntaxique permet des représentations lexicales de grande qualité, favorisant ainsi une récupération plus rapide du sens du mot dans le lexique mental lors de la lecture.

Dans le même ordre d'idées, Nation (2001) est d'avis que la connaissance d'un mot correspond au cumul des connaissances issues de divers aspects. Bien que ses écrits portent sur l'apprentissage d'une langue seconde, les aspects avancés s'appliquent aussi aux apprenants d'une langue première. Pour Nation (2001), ces connaissances sont au nombre de trois, soit sur la forme, sur le sens et sur l'usage du mot. Les connaissances sur la forme du mot renvoient à la forme orthographique (p. ex. *hydrographie*) et à la forme phonologique (p. ex. [idrɔgrafi]), mais aussi à la structure morphologique du mot (p. ex. *hydr + graph + -ie*). Les connaissances sur le sens renvoient d'une part à la sémantique, mais aussi aux connaissances conceptuelles que renferme le mot (concepts et référents), ainsi qu'aux associations avec d'autres mots (antonymes, synonymes, champ lexical). Enfin, les connaissances relatives à l'usage du mot concernent les fonctions grammaticales, les collocations ainsi que les contraintes d'utilisation du mot (Nation, 2001). Toujours selon Nation (2001), la majorité des aspects de la connaissance des mots se développe de façon implicite, soit en portant attention aux mots et en s'assurant de les rencontrer à de multiples reprises. Toutefois, les connaissances relatives au sens des mots nécessitent un apprentissage explicite et devraient être la cible des enseignements, tout comme certaines composantes de l'usage des

mots, comme les collocations (p. ex. on peut dire *tomber amoureux*, mais on ne peut pas dire *chuter amoureux*) ou les contraintes d'utilisation (p. ex. Au Québec, le mot *char* est d'usage familier pour désigner une automobile. On l'évitera en contexte scolaire.) (Nation, 2001).

Des chercheurs ont tenté d'établir une certaine progression de la connaissance des mots. Par exemple, Dale (1965) a proposé une gradation de la connaissance des mots en quatre étapes : 1) aucune connaissance, l'individu n'a jamais été en contact avec le mot; 2) aucune connaissance du sens, mais l'individu a vu ou entendu le mot; 3) connaissances peu approfondies, l'individu reconnaît le mot en contexte, et; 4) bonne connaissance du mot, l'individu est capable de l'expliquer à d'autres. Par ailleurs, Beck et ses collaborateurs (1987) proposent une gradation très semblable en ajoutant une étape intermédiaire entre la troisième et la dernière étape de la connaissance des mots de Dale (1965). En effet, les auteurs précisent qu'après l'étape de la reconnaissance du mot dans un contexte bien précis, l'individu a suffisamment de connaissances sur le mot, mais ne l'utilise pas de façon appropriée. Enfin, la dernière étape pour Beck et ses collaborateurs (1987) considère à la fois la richesse de la connaissance du mot, mais aussi la capacité pour l'apprenant à l'utiliser adéquatement de façon décontextualisée. La connaissance des mots ne serait donc pas un processus dichotomique qui permettrait à l'individu de développer soudainement une connaissance approfondie d'un mot qu'il ne connaissait pas du tout avant de l'entendre ou de le lire (Beck *et al.*, 1987; Dale, 1965).

Bien que ces propositions de développement de la connaissance des mots puissent sembler linéaires et hiérarchiques, il faut plutôt concevoir la connaissance d'un mot comme évoluant sur un continuum qui permet une infinité de degrés de connaissance. La connaissance sur le mot deviendra de plus en plus approfondie et précise à chaque rencontre avec le mot. Cet aspect incrémental de la connaissance des mots est d'ailleurs un des cinq aspects évoqués par Nagy et Scott (2000) quant à la complexité de la connaissance des mots. La connaissance des mots revêt aussi un caractère multidimensionnel, comme le suggère également Nation (2001), c'est-à-dire que les connaissances sur le mot peuvent être de types différents (p. ex. forme orale, forme écrite, classe grammaticale, sens, fréquence, etc.). Bien connaître un mot implique davantage que le simple rappel de sa définition. La découverte de ces types de connaissances ne respecte pas un ordre établi et peut se développer différemment d'un individu à l'autre. De plus, chaque type de connaissance évolue à lui seul sur un continuum. Par exemple, la connaissance relative à la forme écrite du mot peut être plus ou moins développée d'un individu à l'autre, ce qui serait indépendant des connaissances d'un autre type, comme la forme orale. Le troisième

aspect de la connaissance d'un mot de Nagy et Scott (2000) concerne la polysémie qui touche un grand nombre de mots, particulièrement les plus fréquents. Une même forme orthographique peut compter différents sens plus ou moins liés ou, selon le contexte, revêtir diverses nuances (p. ex. Diminue le *volume* de la télévision. J'ai emprunté ce *volume* à la bibliothèque. Les élèves calculent le *volume* de ce polyèdre.). Aussi, les connaissances quant aux mots sont interreliées, c'est-à-dire qu'elles se construisent par les liens avec les mots connus et les concepts familiers. Connaitre un mot, c'est en connaître d'autres. Les expériences personnelles favorisent aussi la connaissance d'un mot (Perfetti, 2007; Reichle et Perfetti, 2003). Enfin, cette connaissance est hétérogène et dépend du type de mot. Par exemple, la profondeur de la connaissance du déterminant *la* diffère d'autres mots auxquels certaines caractéristiques ne s'appliquent pas. Des connaissances différentes sont nécessaires quant à la forme orthographique *la*, lorsque celle-ci occupe le rôle de déterminant (p. ex. *la* règle), de pronom (p. ex. regarde-*la*) ou encore de nom (p. ex. jouer un *la* dièse). Bref, la connaissance des mots est un concept complexe dû à ses caractères incrémental, multidimensionnel, polysémique, interrelié et hétérogène (Nagy et Scott, 2000).

Une autre vision de la connaissance du vocabulaire est celle concernant les dimensions du vocabulaire, soit l'étendue, la profondeur, l'organisation et le traitement automatique des mots (Qian, 2002). L'*étendue du vocabulaire* renvoie au nombre de mots connus par un individu, quels que soient les aspects de ces connaissances. D'un individu à l'autre, l'étendue du vocabulaire varie, de restreinte à très vaste. La *profondeur du vocabulaire* concerne la qualité des connaissances sur les divers aspects des mots comme la forme orthographique, le ou les sens, l'usage, etc. Tout comme l'étendue, la profondeur du vocabulaire varie sur un continuum de riche à pauvre pour chaque individu, mais aussi pour chacun des mots, ce qui rejoint le caractère multidimensionnel de la connaissance des mots décrit par Nagy et Scott (2000). L'étendue et la profondeur du vocabulaire sont reconnues comme étant deux concepts extrêmement liés et ayant un impact sur les performances en compréhension en lecture, notamment dans un contexte scolaire (Qian, 2002). L'*organisation du vocabulaire* concerne les relations entre les mots, soit le champ lexical, la polysémie, la synonymie ou encore les familles morphologiques, ainsi que la représentation des mots dans le lexique mental. Enfin, le *traitement automatique des mots* renvoie aux processus cognitifs nécessaires au traitement des mots, que ce soit durant l'identification ou la production, à la fois en modalité orale et écrite (Qian, 2002). Ces quatre dimensions sont étroitement liées les unes aux autres et interagissent entre elles, notamment sur les processus cognitifs mis en œuvre lors du traitement des mots (Qian, 2002). Ainsi, la façon dont les mots sont identifiés et traités

influencerait la connaissance du vocabulaire. Comme la présente étude porte sur le traitement des mots écrits, les modèles d'identification des mots écrits qui tiennent compte de la morphologie seront présentés dans la section suivante.

2.4 L'identification des mots écrits

Cette section vise d'abord à définir le concept *identification des mots écrits* tel qu'il sera utilisé dans le cadre de cette thèse doctorale (2.4.1). Ensuite, les modèles d'identification des mots écrits sont présentés en focalisant sur les modèles qui visent à expliquer l'identification des mots écrits polymorphémiques puisque ce sont ces mots qui sont concernés dans le cadre de cette étude (2.4.2). Puis, cette présentation des différents modèles permet de définir le traitement morphographique dans une perspective cognitive (2.4.3).

2.4.1 Une définition du concept *identification des mots écrits*

Un des concepts centraux de cette étude, l'identification des mots écrits, est largement étudié en psycholinguistique. Plusieurs définitions circulent pour illustrer ce concept. Par exemple, Morais (1999) définit ce concept complexe comme la mise en œuvre des processus permettant « d'identifier chaque mot en tant que forme orthographique ayant une signification et de lui attribuer une prononciation » (p.115). Le point fort de cette définition est la considération de deux des aspects de la connaissance des mots de Nation (2001) : la forme (autant orthographique que phonologique) et le sens. Cependant, cette définition ne permet pas de situer dans le temps l'accès à la forme orthographique et au sens.

La définition de l'identification des mots écrits de Hoover et Tunmer (2020) permet davantage de représenter les étapes d'accès à la forme et au sens du mot. Ces auteurs définissent l'identification des mots écrits comme l'habileté à reconnaître et à identifier rapidement et précisément un mot écrit afin de récupérer son sens dans le lexique mental. Cette définition laisse comprendre que l'identification rapide et précise d'un mot permettra, dans une étape subséquente, d'accéder à son sens. Un autre aspect qui différencie cette définition de la première est la place accordée à la vitesse et à la précision de la lecture du mot. Ces distinctions permettent de mieux opérationnaliser le concept d'identification des mots écrits. En conséquence, c'est cette définition qui est retenue dans le cadre de cette étude.

2.4.2 Les modèles de l'identification des mots écrits

Différents modèles fournissent une schématisation de l'identification des mots écrits sans parler explicitement des mots polymorphémiques. Les modèles à double voie (Coltheart, 1978), qui mèneront aux modèles développementaux qui visent à expliquer l'apprentissage des processus d'identification des mots écrits chez le lecteur débutant (Frith, 1985; Harris et Coltheart, 1986; Marsh *et al.*, 1981), supposent l'existence de deux principales procédures pour identifier les mots écrits : la voie directe qui renvoie à l'accès instantané au mot stocké dans le lexique interne et la voie indirecte qui renvoie à la mise en correspondance des graphèmes et des phonèmes. Quant à eux, les modèles par analogie accordent une importance centrale aux mots déjà connus, notamment en avançant que l'identification des mots écrits passe d'abord par l'identification des mots partageant la même rime (Goswami et Bryant, 1990). Enfin, les modèles connexionnistes (Coltheart *et al.*, 1993; Grainger et Holcomb, 2009; McClelland et Rumelhart, 1981; Paap *et al.*, 1982; Seidenberg et McClelland, 1989b) offrent une réponse universelle à l'identification des mots écrits. On peut donc inférer que les mots polymorphémiques sont considérés dans ce traitement universel. Cependant, comme ces quelques modèles cités n'abordent pas spécifiquement le traitement des mots polymorphémiques, ils ne seront pas détaillés. La présentation des modèles connexionnistes qui en parlent spécifiquement est disponible à la section 2.4.2.4. En somme, le lecteur qui désire des détails supplémentaires sur les modèles d'identification des mots écrits énumérés dans le présent paragraphe pourra se référer à l'Appendice A.

Le modèle développemental à double fondation de Seymour (1999, 2007) est un point de départ intéressant pour comprendre l'identification des mots écrits polymorphémiques chez les lecteurs débutants dans la mesure où il accorde une certaine place à la morphologie en fin d'apprentissage. Selon Seymour (1999, 2007), deux procédures coexistent en début d'apprentissage de la lecture : la procédure logographique (lecture globale) et la procédure alphabétique (lecture graphème par graphème). Par la suite, les procédures orthographique et morphographique s'installent progressivement tout en chevauchant la ou les procédures précédentes. Grâce aux procédures logographique et alphabétique, le lecteur élargit son lexique orthographique, ce qui lui permet d'utiliser principalement la procédure orthographique pour la lecture de mots. Enfin, plus tardivement, le système morphographique s'installe et permet la lecture de mots polymorphémiques par la reconnaissance de morphèmes (Seymour, 1999, 2007). Dans ce modèle (Seymour, 1999, 2007), l'identification des mots écrits est soutenue par la conscience linguistique (phonologique et morphologique) des lecteurs qui prend sa source du langage oral et qui veille à influencer le développement des processus d'identification des mots écrits. La

conscience de la structure de la langue, comme « la conscience morphologique [qui] focalise sur la conscience qu'ont les enfants de la structure morphémique des mots et de leur habileté à réfléchir sur cette structure et à la manipuler »¹⁰ [ma traduction] (Carlisle, 1995, p. 164), a alors une incidence sur le traitement des mots écrits, particulièrement sur les mots polymorphémiques. Bien qu'intéressant pour comprendre le développement des processus d'identification des mots écrits chez les lecteurs débutants, ce modèle à double fondation de Seymour (1999, 2007) n'est pas suffisamment explicite sur la façon dont les mots polymorphémiques sont traités. En effet, on ne sait pas si tous les mots polymorphémiques sont systématiquement identifiés par le système morphographique et, si ce n'est pas le cas, sur quels critères est basée cette identification. Le modèle à double fondation de Seymour (1999, 2007) ne permet pas non plus de savoir dans quelle mesure et de quelle façon a lieu la reconnaissance des morphèmes lors de l'identification des mots écrits polymorphémiques.

Les recherches expérimentales ont mené à diverses hypothèses relatives au traitement morphographique, plus particulièrement à la façon dont les mots polymorphémiques sont stockés en mémoire et traités lors de la lecture. Les principaux modèles qui visent à expliquer la place et le rôle des représentations morphographiques lors de l'identification des mots polymorphémiques sont regroupés en quatre catégories : les modèles de décomposition morphologique, les modèles holistiques, les modèles hybrides et les modèles connexionnistes distribués. Ces modèles se distinguent par plusieurs caractéristiques comme le moment d'activation du traitement morphographique dans le traitement lexical, la façon dont les unités sont stockées dans le lexique mental ou encore la nature même de la décomposition, si elle est considérée. Des modèles de chacune de ces catégories sont successivement présentés afin d'être en mesure de sélectionner ceux qui seraient les plus adaptés pour permettre d'effectuer des choix méthodologiques rigoureux ainsi que pour interpréter les résultats de la présente recherche.

2.4.2.1 Les modèles de décomposition morphologique

Les modèles de décomposition morphologique (p. ex. Taft et Forster, 1975) sont des modèles sublexicaux, c'est-à-dire qu'ils prévoient un traitement des unités sublexicales, soit plus petites que le mot, avant l'accès lexical. Comme leur nom l'indique, les modèles de décomposition morphologique

¹⁰ «Morphological awareness focuses on children's conscious awareness of the morphemic structure of words and their ability to reflect on and manipulate that structure. »

prévoient que les mots affixés passent par une étape de décomposition morphologique, soit de séparer la racine du mot de ses affixes, avant l'accès au lexique mental. Les modèles de Taft et Forster (1975) et de Taft (1994), présentés ci-après, s'appuient sur les données d'une tâche de décision lexicale reposant sur le traitement de mots polymorphémiques tirés du vocabulaire anglais et formés de racines liées (Taft et Forster, 1975). Effectivement, les auteurs (Taft et Forster, 1975; Taft, 1994) ont observé que les racines liées (p. ex. *vive* dans *revive*), bien que non autonomes lexicalement, demandaient un temps de réponse plus long que les pseudoracines liées (p. ex. **lish* dans *relish*) ou les parties finales des mots ne correspondant ni à des racines liées ni à des pseudoracines (p. ex. **nace* dans *menace*). La réponse à la tâche de décision lexicale était aussi influencée par la fréquence de la racine liée d'un mot préfixé, même lorsque la fréquence du mot entier était contrôlée (Taft, 1975). Ces différences significatives dans le temps de réaction soutiennent la représentation lexicale d'au moins certaines racines liées dans le lexique mental (Taft, 1994; Taft et Forster, 1975). De ce fait, la description des modèles suivants s'applique autant pour les mots polymorphémiques formés de racines libres que de racines liées.

Le modèle de décomposition obligatoire prélexicale de Taft et Forster (1975) s'appuie sur la décomposition automatique des mots en reconnaissant automatiquement les préfixes et en les détachant du reste du mot avant l'accès au lexique mental où seules les racines sont stockées. Selon ce modèle (Taft et Forster, 1975), le mot *refaire* passe par une étape de décomposition où le préfixe *re-* est automatiquement retiré avant que la racine *faire* soit traitée. Les mots *faire*, *refaire* et *défaire* ont donc tous la même entrée lexicale dans le lexique mental (Ferrand et Ayora, 2009). Aussi, selon ce modèle (Taft et Forster, 1975), les mots pseudoaffixés, c'est-à-dire qui contiennent une suite orthographique correspondant à un affixe (p. ex. *re* dans *recette* ou *religion*), sont traités comme des mots réellement affixés. Cette décomposition obligatoire des mots pseudoaffixés entraîne alors la recherche de la pseudoracine, par exemple **ligion* dans *religion*, alors que celle-ci n'est pas présente dans le lexique mental. L'étape de vérification n'aboutissant pas, une réanalyse du mot s'impose en réintroduisant le pseudoaffixe préalablement retiré afin d'analyser le mot en entiereté, ici *religion*. Selon ce point de vue, le traitement des mots pseudoaffixés est toujours plus lent que celui des mots réellement affixés. Or, la décomposition morphologique, bien qu'elle demeure une stratégie existante, ne semble pas être une règle universelle, mais plutôt s'activerait pour quelques exceptions seulement (Rubin *et al.*, 1979). En plus, compte tenu de la présence importante de mots pseudoaffixés en français (Pavard, 1983), comme en anglais (Schreuder et Baayen, 1994), des auteurs avancent que les modèles de décomposition obligatoire seraient inefficients (Grainger et Beyersmann, 2017; Schreuder et Baayen, 1994).

Effectivement, un modèle de décomposition obligatoire qui reconnaît instantanément les préfixes pour les détacher du reste du mot tel que celui de Taft et Forster (1975) est de 3 à 8 fois moins performant qu'un modèle holistique où les mots qui partagent le même préfixe sont stockés et activés ensemble (Schreuder et Baayen, 1994). Aussi, le modèle de décomposition prélexicale obligatoire permet difficilement d'expliquer le traitement des mots composés formés de deux racines libres (Sandra, 1991). En plus du taux d'erreur important et des coûts de traitement cognitif élevés engendrés par ce modèle, une autre limite soulevée concerne le type d'unités stockées dans le lexique mental. Bien que seules les racines *y* sont stockées, la reconnaissance automatique des affixes menant à la décomposition obligatoire suggère la présence d'un stock de préfixes (et probablement de suffixes) indépendant du lexique mental (Ferrand, 2001). Toutefois, cette avancée ou hypothèse n'a jamais été démontrée (Taft, 1994). D'ailleurs, Taft (1994) dénonce la rigidité de son modèle en soulevant l'ambiguïté générée par la présence de ce qu'il nomme *les formes combinatoires initiales*, soit les racines latines et grecques comme *poly* ou *bio*. Effectivement, le modèle de Taft et Forster (1975) ne permet pas de savoir de quelle façon ces morphèmes liés sont traités. On ne sait pas s'ils sont reconnus automatiquement comme les préfixes puis détachés du reste du mot, ce qui nécessiterait un stock de préfixes et de racines liées très étendu, ou s'ils sont plutôt traités comme une des racines d'un mot composé. Le modèle à activation interactive (Taft, 1994) souhaite remédier à cette limite.

Le modèle à activation interactive de la morphologie prélexicale (Taft, 1994) prévoit trois niveaux de représentations après la représentation orthographique visuelle du mot : morphologique, lexical et conceptuel. Dans ce modèle revisité, la décomposition morphologique prélexicale n'est plus obligatoire. De ce fait, le stock d'affixes, aspect critiqué dans le modèle précédent (Taft et Forster, 1975), n'est plus nécessaire. Au niveau morphologique, les mots polymorphémiques sont traités sur la base des affixes et des racines liées. Les formes globales des mots ainsi que les racines libres sont représentées au niveau lexical. Enfin, l'accès aux informations sémantiques des mots se produit au niveau conceptuel du modèle à activation interactive de la morphologie prélexicale (Taft, 1994). À la suite de l'identification d'un mot polymorphémique, chaque morphème lié est traité indépendamment au niveau morphologique et active les mots reliés morphologiquement au niveau lexical avant de permettre l'activation sémantique au niveau conceptuel (Taft, 1994). Le morphème lié le moins fréquent est celui qui permet l'activation la plus rapide aux niveaux subséquents. Par exemple, dans le mot *inclus*, un mot polymorphémique formé de deux morphèmes liés (le préfixe *in-* et la racine liée *clus*), le morphème *in-* active un très grand nombre de mots au niveau lexical (p. ex. *inclus*, mais aussi *ingérer*, *installer*, *incorporer*, *injecter*,

incruster, etc.), ce qui est peu efficace, puisque c'est très long d'activer autant de mots. C'est donc la racine *clus* qui joue un rôle plus important dans l'activation lexicale du mot *inclus* en activant un moins grand nombre de mots (p. ex. *inclus*, *exclus*, *exclusif*, *inclusif*, *exclusivement*, etc.). Le mot activé le plus souvent au niveau lexical (ici *inclus* a été activé une fois pour chaque morphème, donc deux activations) mène à une plus grande activation du sens du mot au niveau conceptuel. Dans l'optique où le mot polymorphémique est composé d'une racine libre (p. ex. *faire* dans *refaire*), seul le préfixe *re-* est représenté au niveau morphologique afin d'éviter la redondance des éléments. La racine libre (p. ex. *faire*) est directement activée au niveau lexical, tout comme le mot dans sa forme globale (p. ex. *refaire*). Il va sans dire que, pour assurer le fonctionnement adéquat de ce modèle, tous les mots dans leur forme globale, mais aussi tous les morphèmes libres et liés, doivent être représentés dans le lexique mental (Ferrand, 2001), ce qui impliquerait une redondance quant au stockage des informations.

En résumé, les modèles de décomposition morphologique ont conduit au constat que l'accès au sens du mot entier est nécessaire dans plusieurs cas. Effectivement, l'accès uniquement au sens des morphèmes peut s'avérer trompeur (p. ex. *redouter* ne signifie pas « douter à nouveau ») (Schreuder et Baayen, 1994) et la présence de pseudoaffixes peut mener à des hypothèses erronées (p. ex. *recette* ne signifie pas « cette à nouveau ») (Grainger et Beyersmann, 2017; Pavard, 1983; Schreuder et Baayen, 1994). Ces principales limites ont permis de tracer la voie aux modèles holistiques.

2.4.2.2 Les modèles holistiques

Les modèles holistiques soutiennent que tous les mots, qu'ils soient monomorphémiques ou polymorphémiques, sont présents de façon globale dans le lexique mental (p. ex. Butterworth, 1983; Manelis et Tharp, 1977). De ce fait, même les mots partageant une même racine (p. ex. *chat*, *chatte*, *chaton*, *chatons*, etc.), qu'ils soient des mots de base, des mots fléchis ou encore des mots dérivés, sont stockés et traités indépendamment les uns des autres. Au sein des modèles holistiques, il n'y a aucune représentation morphologique, ce qui implique un stockage important dans le lexique mental (Diependaele *et al.*, 2012). Dans cet ordre d'idées, Manelis et Tharp (1977) soutiennent l'hypothèse que les mots suffixés sont traités globalement plutôt que décomposés. Ils justifient leurs résultats contradictoires à ceux de Taft et Forster (1975) qui soutiennent plutôt la décomposition morphologique des mots préfixés en émettant l'hypothèse que seuls les mots préfixés pourraient activer la décomposition morphologique due à un traitement lexical des mots de gauche à droite en lecture. Cependant, des résultats de recherche ont ensuite démontré que les suffixes sont aussi considérés lors

du traitement morphologique des mots polymorphémiques autant en flexion (p. ex. Caramazza *et al.*, 1988; Caramazza *et al.*, 1985; Laudanna et Burani, 1985) qu'en dérivation (p. ex. Burani et Caramazza, 1987; Colé *et al.*, 1986; Colé *et al.*, 1989), venant invalider cette hypothèse, du moins sa généralisation.

D'autres résultats remettent en cause les modèles holistiques. Effectivement, il a été démontré que les pseudomots, soit des non-mots formés de morphèmes existants, sont traités plus rapidement que les non-mots, et ce, en tout temps (Taft et Forster, 1975) ou dans certains contextes seulement (Rubin *et al.*, 1979). Ces résultats suggèrent la prise en compte des morphèmes lors du traitement des mots ou des pseudomots polymorphémiques avant l'accès aux représentations lexicales. De plus, le fait qu'il soit possible de créer et d'accéder au sens de néologismes formés de morphèmes existants et qui respectent les règles de formation des mots dans la langue cible (p. ex. **infleurissable*) s'avère une limite des modèles holistiques (Butterworth, 1983). Si on part du principe que tous les mots sont indépendamment représentés dans le lexique mental et que cette représentation est lexicale plutôt que morphologique, la compréhension (ou la production) de tels mots serait difficile, voire impossible. Butterworth (1983) tente de contourner cette limite en suggérant que les locuteurs s'appuient sur des mots semblables, de façon analogique, pour de telles créations lexicales. Cette hypothèse a été validée dans différentes langues comme l'anglais (p. ex. Mattiello, 2016) et le néerlandais (p. ex. Krott *et al.*, 2001).

Bien que les modèles holistiques ne considèrent pas la dimension morphologique dans les représentations lexicales, ils ne peuvent nier que les mots morphologiquement reliés partagent des liens lexicaux, ce qui pourrait mener à une reconnaissance plus rapide des mots morphologiquement reliés lors des tâches d'amorçage. Même si tous les mots morphologiquement reliés restaient indépendamment représentés, l'hypothèse voulant qu'ils soient regroupés dans le lexique mental est émise, sans précision à savoir si le mot de base est à la tête de chacune des catégories (Butterworth, 1983). Cette hypothèse inspire le modèle de la morphologie supralexicale.

Une extension des modèles holistiques (p. ex. Butterworth, 1983; Manelis et Tharp, 1977) et du modèle connexionniste d'identification des mots écrits de McClelland et Rumelhart (1981) a été proposée pour expliquer les effets d'amorçage morphologique. En effet, le modèle de la morphologie supralexicale (Giraud, 2005; Giraud et Grainger, 2000, 2001, 2003; Grainger *et al.*, 1991) est le seul modèle à activation interactive qui situe les représentations morphologiques à l'interface entre les représentations lexicales orthographiques et les représentations sémantiques. Il prévoit donc l'accès aux représentations

morphologiques après l'accès lexical plutôt qu'avant l'accès lexical comme le prévoient les modèles à décomposition morphologique (p. ex. Taft, 1994; Taft et Forster, 1975). Dans le modèle de la morphologie supralexicale (Giraud, 2005; Giraud et Grainger, 2000, 2001, 2003; Grainger *et al.*, 1991), les représentations morphologiques sont organisées en familles morphologiques dans le lexique mental. Des connexions facilitatrices bidirectionnelles relient les représentations lexicales et morphologiques. De ce fait, la lecture d'un mot, qu'il soit monomorphémique (p. ex. *faire*) ou polymorphémique (p. ex. *défaire*), permet l'activation au niveau morphologique de la racine (p. ex. *faire*) et est renvoyée du niveau morphologique au niveau lexical à tous les autres membres de la famille morphologique (p. ex. *refaire*, *malfaire*, *savoir-faire*, etc.). Ces connexions bidirectionnelles expliquent les effets d'amorçage morphologique lors de la lecture de mots morphologiquement reliés.

Des résultats de recherche en amorçage masqué en français ont permis de soutenir le modèle supralexicale plutôt que les modèles sublexicaux. Effectivement, l'effet d'amorçage lors de la lecture d'un mot suffixé s'avère équivalent que l'amorce soit un mot monomorphémique ou polymorphémique (Giraud et Grainger, 2001). Cette hypothèse est incompatible avec les modèles sublexicaux qui prédisent un effet d'amorçage plus grand pour les mots monomorphémiques que pour les mots polymorphémiques puisqu'aucune décomposition ne serait nécessaire. De plus, aucun effet d'amorçage n'a été observé lorsque les amorces étaient des pseudoracines (Giraud et Grainger, 2001), contrairement aux prédictions des modèles sublexicaux qui prévoient une décomposition autant pour les mots pseudoaffixés que pour ceux réellement affixés. Aussi, les modèles supralexicaux prévoient un effet d'amorçage facilité par la fréquence du mot entier dans la langue alors que les modèles sublexicaux prévoient un effet de fréquence cumulée, soit la somme des fréquences des mots reliés morphologiquement. Or, les résultats des études empiriques n'ont pas démontré d'effet de fréquence cumulée dans un paradigme d'amorçage masqué, alors que la fréquence des mots entiers a démontré un effet facilitateur (Giraud et Grainger, 2000), un autre argument en défaveur des modèles sublexicaux. Toutefois, plusieurs des arguments avancés par les modèles supralexicaux peuvent aussi être expliqués par les modèles hybrides tels que l'effet de fréquence du mot entier ainsi que l'influence des racines et des mots pseudoaffixés (Giraud, 2005).

2.4.2.3 Les modèles hybrides

À l'interface entre les modèles de décomposition morphologique obligatoire et les modèles holistiques, les modèles hybrides reposent sur le traitement des mots écrits via deux voies, une voie de traitement

du mot en entier et une voie de décomposition morphologique (p. ex. Caramazza *et al.*, 1985; Chialant et Caramazza, 1995; Frauenfelder et Schreuder, 1992; Schreuder et Baayen, 1995).

Le modèle d'activation parallèle lexicomorphologique (MAPL) [ma traduction de *Augmented Addressed Morphology* (Caramazza *et al.*, 1985)] prévoit deux niveaux de traitement entre l'identification visuelle du mot et sa représentation sémantique : les unités d'accès et les représentations lexicales. Les unités d'accès renvoient à la façon dont le mot est accédé, soit de façon globale (ou voie globale) ou par ses unités sublexicales, les morphèmes (ou voie de décomposition morphologique). Les représentations lexicales orthographiques servent à l'accès lexical du mot et à la récupération de son sens. Ces représentations prévoient un stockage décomposé des mots en morphèmes dans le lexique mental. Lorsqu'un mot monomorphémique est rencontré (p. ex. *faire*), la voie globale est empruntée. Le mot est donc traité en entier et l'accès à son sens l'est tout autant. Lors de la lecture d'un mot polymorphémique (p. ex. *défaire*), les deux voies sont simultanément activées, la voie globale et la voie de décomposition morphologique. La voie d'accès la plus rapide mène aux représentations lexicales à l'instar d'une course. Lorsque disponible, l'activation de la forme globale du mot reste toujours la plus rapide, la plus efficace. De façon générale, les mots connus, bien qu'ils soient polymorphémiques (p. ex. *défaire*) ou irréguliers (p. ex. *put*, verbe *pouvoir* conjugué au passé simple à la troisième personne du singulier), sont accédés plus rapidement par leur forme globale. La fréquence du mot prédit l'activation de cette voie (Caramazza *et al.*, 1988). Les mots nouveaux polymorphémiques ou peu rencontrés, mais tout de même transparents, sont accédés grâce à la décomposition morphologique, notamment par l'accès à la racine. Cette voie de décomposition morphologique est influencée par divers facteurs comme la fréquence du mot entier et de chacun des morphèmes, la fréquence cumulée des mots de la famille morphologique ou encore la transparence sémantique (p. ex. Caramazza *et al.*, 1988; Chialant et Caramazza, 1995). De ce fait, le mot polymorphémique *défaire* active simultanément les deux voies à la suite de la rencontre visuelle avec le mot. Compte tenu de sa grande fréquence, il est accédé par sa forme globale et permet l'accès au sens général du mot, puis les morphèmes *dé-* et *faire* peuvent être récupérés dans les représentations lexicales. D'autre part, le mot *déflouraison* emprunte aussi les deux voies lors de l'identification du mot. Cependant, tandis que le mot est plus rare et probablement méconnu du lecteur, la récupération du sens global du mot est probablement impossible. Comme les affixes (*dé-* et *-aison*) ainsi que la racine du mot (*fleur*) sont plus fréquents que le mot en soi, donc plus susceptibles d'être connus, et que le mot est entièrement décomposable en morphèmes, la voie de décomposition morphologique mène plus rapidement aux représentations lexicales (Chialant et

Caramazza, 1995). Au sein de cette voie, le mot active d'abord les morphèmes constituants ainsi que leurs sens respectifs pour permettre ensuite une représentation globale du mot. Dans ce modèle, sont donc stockés indépendamment les morphèmes (racines et affixes) et les formes irrégulières (p. ex. *put*) (Ferrand et Ayora, 2009). Il est à noter que si le MAPL a initialement été développé au regard de la morphologie flexionnelle (p. ex. Caramazza *et al.*, 1988; Laudanna et Burani, 1985), des résultats qui soutiennent ce modèle hybride ont également été obtenus avec des items dérivés (p. ex. Burani et Caramazza, 1987; Laudanna *et al.*, 1994).

Très similaire au MAPL (Caramazza *et al.*, 1985), le modèle de compétition morphologique (MCM) [ma traduction de *Morphological Race Model*] proposé par Frauenfelder et Schreuder (1992) voit toujours le traitement des mots écrits sous forme de course entre deux voies. Dans ce modèle, la fréquence des mots est le facteur déterminant pour le traitement d'un mot par la voie globale. Pour la voie de décomposition morphologique, d'autres facteurs sont à considérer comme la transparence phonologique de la racine ainsi que la transparence sémantique du mot. Effectivement, plus le mot est transparent phonologiquement, c'est-à-dire que la racine ne subit aucune modification phonologique (et orthographique) à la suite de l'ajout d'un affixe (p. ex. *rapide*, *rapidement*), par opposition aux mots plus opaques (p. ex. *musique*, *musicien*), plus il est traité rapidement par la voie de décomposition morphologique. La *transparence sémantique* réfère, quant à elle, à la probabilité de déduire le sens d'un mot en s'appuyant uniquement sur le sens des morphèmes. Encore une fois, plus le mot est transparent sémantiquement (p. ex. *refaire* = *re-* + *faire* = « faire à nouveau »), plus il est traité rapidement par rapport aux mots plus opaques (p. ex. *inclus* = *in-* + *clus* = « compris à l'intérieur »). Enfin, la principale différence entre le MAPL (Caramazza *et al.*, 1985) et le MCM (Frauenfelder et Schreuder, 1992) réside dans le fait que ce dernier prévoit un autre facteur relatif au traitement des mots polymorphémiques, soit le niveau d'activation de la racine et des affixes. Plus le niveau d'activation est élevé pour un morphème donné, plus le traitement morphologique est efficace. Ce facteur est à concevoir à la manière d'un système de pondération octroyé aux racines et aux affixes à la suite d'une analyse de décomposition morphologique réussie. Lorsqu'un mot polymorphémique est traité avec succès par la voie de décomposition morphologique (p. ex. *redistribution*), le niveau d'activation de chacun de ses morphèmes augmente (p. ex. *re-*, *distribue* et *-tion*). De ce fait, la lecture d'un autre mot contenant des morphèmes communs (p. ex. *relocalisation*) est plus rapide dû à leur niveau d'activation plus élevé. Il est à noter que le niveau d'activation n'augmente que si le mot est traité par la voie morphologique. Si le mot est trop fréquent ou si le lecteur y a été exposé à de nombreuses reprises et que le mot est traité

par la voie globale (p. ex. *redire*), le niveau d'activation n'est augmenté pour aucun morphème. Enfin, le niveau d'activation des morphèmes peut aussi diminuer si les morphèmes sont rarement activés. Donc, une activation récurrente de certains morphèmes dans un court laps de temps ne garantit pas une activation rapide à long terme. En résumé, selon le modèle MCM (Frauenfelder et Schreuder, 1992), la voie globale est la plus rapide pour les mots de haute fréquence, mais aussi pour les mots opaques ou monomorphémiques pour lesquels la décomposition morphologique échoue. Les mots de moyenne à basse fréquence ainsi que les nouveaux mots polymorphémiques sont, quant à eux, traités par la voie de décomposition morphologique en fonction de facteurs de transparence (phonologique et sémantique) et du niveau d'activation de chacun des morphèmes (Frauenfelder et Schreuder, 1992).

Schreuder et Baayen (1995) proposent aussi un modèle à deux voies où les mots sont traités par deux accès soit l'accès direct et l'accès par les constituants morphémiques. Contrairement aux modèles précédents (Caramazza *et al.*, 1985; Frauenfelder et Schreuder, 1992), les deux voies ne compétitionnent pas pour accéder au sens du mot. Plutôt, elles interagissent pour converger vers les représentations sémantiques et syntaxiques, en passant d'abord par des nœuds conceptuels [ma traduction de *concept nodes*]. Les nœuds conceptuels permettent d'activer diverses informations sémantiques et syntaxiques pour mener à la représentation juste du mot par la combinaison de ces informations. Ce modèle prévoit aussi des rétroactions des représentations sémantiques et syntaxiques vers les nœuds conceptuels, et des nœuds conceptuels vers les voies d'accès. C'est d'ailleurs ce phénomène de rétroaction qui permet, après plusieurs rencontres avec le mot, qu'un mot polymorphémique soit éventuellement traité à partir de sa forme globale plutôt qu'à partir de sa forme décomposée (Schreuder et Baayen, 1995).

En résumé, les trois modèles hybrides principaux développés (MAPL, Caramazza *et al.*, 1985; MCM, Frauenfelder et Schreuder, 1992; Schreuder et Baayen, 1995) prévoient le traitement des mots écrits selon deux voies parallèles, une voie globale et une voie de décomposition morphologique. La voie globale est davantage influencée par la fréquence du mot entier tandis que la voie de décomposition morphologique est influencée par une variété de facteurs comme la fréquence du mot entier, la fréquence cumulée des mots de la même famille morphologique ou encore la transparence phonologique et sémantique.

Des limites peuvent toutefois être soulevées pour ces modèles hybrides. Effectivement, la décomposition morphologique est réservée aux mots transparents et réguliers alors que tous les autres

mots sont traités plus efficacement par la voie globale. Or, la langue française compte de nombreuses ambiguïtés et exceptions qui dévient des règles de formation régulières des mots, et ce, pour plusieurs raisons dont l'évolution diachronique de la langue (p. ex. *chevalet* n'est plus utilisé désormais pour désigner un petit cheval, bien que le mot présente une structure morphologique régulière) ou les emprunts lexicaux (p. ex. *stimulus*, au pluriel, s'écrivait *stimuli* avant la rectification orthographique : on peut toujours le rencontrer sous cette forme). Une autre problématique relève du statut des racines liées, notamment lorsque leur sens est moins transparent (p. ex. *clus* dans *inclus*). On pourrait aussi penser que les mots composés relativement opaques (p. ex. *gratte-ciel*, *croquemonsieur*, *pot-pourri*) pourraient être problématiques pour les modèles hybrides. Lorsqu'ils seraient rencontrés pour la première fois, seraient-ils traités de façon globale ou décomposée ? Les modèles connexionnistes distribués tentent de remédier à ces irrégularités en prévoyant le traitement de tous les mots dans un même réseau.

2.4.2.4 Les modèles connexionnistes distribués

Les modèles connexionnistes distribués qui abordent la morphologie ont été développés en s'appuyant d'abord sur la morphologie dérivationnelle en anglais (Gonnerman et Plaut, 2000; Gonnerman *et al.*, 2007; Plaut et Gonnerman, 2000; Seidenberg et Gonnerman, 2000). Inspirés des modèles connexionnistes de l'identification des mots (p. ex. McClelland et Rumelhart, 1981; Seidenberg et McClelland, 1989a), les modèles connexionnistes distribués ne prévoient aucun stock de mots ni de morphèmes dans le lexique mental. Dans ces modèles connexionnistes non différenciés, les mots monomorphémiques et polymorphémiques, réguliers ou irréguliers, sont tous traités de la même façon (Diependaele *et al.*, 2012; Ferrand et Ayora, 2009; Gonnerman *et al.*, 2007). L'identification d'un mot est le résultat de l'activation de ses différentes représentations. Ces convergences sont graduées, présentées comme sur un continuum. À l'une des extrémités du continuum, deux mots ne partagent aucune similarité quant à la forme (phonologique et orthographique) et au sens, tandis qu'à l'autre extrémité, des similarités parfaitement identiques à chacune des trois dimensions mènent au même stimulus. Pour donner un exemple, la lecture du mot *redire* converge fortement avec des mots tels que *dire* et *médire*, puisqu'ils partagent des représentations des trois types : orthographiques, phonologiques et sémantiques. Toutefois, *redire* convergerait plus faiblement avec un mot tel que *reporter*, malgré la proximité des représentations orthographiques, phonologiques et sémantiques attribuables au préfixe *re-*. Avec le mot *chaton*, par exemple, les convergences avec les représentations orthographiques, phonologiques et sémantiques du mot *redire* sont inexistantes. En effet, les mots *chaton* et *redire* ne

partagent aucune similarité quant à leur forme ou à leur sens, donc ne sont aucunement liés morphologiquement.

Selon ces modèles, les résultats des mesures comportementales permettant de renseigner sur le traitement des mots (p. ex. paradigme d'amorçage) varient selon le degré de transparence quant au sens et à la forme des mots. C'est d'ailleurs ce qui est ressorti de quelques études ayant confirmé un effet d'amorçage gradué selon la similarité sémantique des mots, avec similarités orthographiques ou phonologiques selon le paradigme (Gonnerman et Plaut, 2000; Gonnerman *et al.*, 2007). De façon générale, les paires de mots qui partagent des relations sémantiques fortes produisent un grand effet d'amorçage, celles qui sont moins reliées produisent un effet plus petit, tandis que les paires qui ne sont pas du tout reliées sémantiquement ne produisent aucun effet facilitateur, même s'ils sont reliés phonologiquement (Gonnerman *et al.*, 2007).

En résumé, contrairement aux modèles de décomposition morphologique et aux modèles hybrides, les modèles connexionnistes distribués ne prévoient pas de stockage des morphèmes dans le lexique mental (Seidenberg et Gonnerman, 2000). Ces unités sont plutôt représentées statistiquement par la convergence entre les dimensions phonologiques, orthographiques et sémantiques des mots. De ce fait, aucun module indépendant n'est prévu au sein des modèles pour les représentations morphologiques. Ces modèles permettent aussi d'expliquer le traitement de tous les mots, qu'ils soient réguliers ou irréguliers (Seidenberg et Gonnerman, 2000). Il s'agit d'une option compatible avec les résultats actuels des recherches sur le traitement morphographique (Diependaele *et al.*, 2012).

2.4.2.5 Synthèse des modèles de l'identification des mots écrits

En somme, quatre catégories de modèles tentent d'expliquer les processus de traitement lors de l'identification des mots polymorphémiques écrits. D'abord, les modèles de décomposition morphologique ont été présentés (Taft, 1994; Taft et Forster, 1975), puis les modèles holistiques (Butterworth, 1983; Manelis et Tharp, 1977) qui comprennent aussi le modèle de la morphologie supralexicale (Giraud, 2005; Giraud et Grainger, 2000, 2001, 2003; Grainger *et al.*, 1991) ont été décrits. Par souci d'efficacité, pour éviter la redondance du stockage des unités dans le lexique mental ainsi que pour rendre compte de la complexité morphologique de la langue et de la nécessité d'accéder au sens du mot entier dans différents cas de figure, les modèles holistiques et les modèles de décomposition morphologique obligatoire ont été montrés peu probables (p. ex. : Colé *et al.*, 1986;

Diependaele *et al.*, 2012; Grainger et Beyersmann, 2017; Rubin *et al.*, 1979; Schreuder et Baayen, 1994; Taft et Forster, 1975). Pour continuer, ont été exposés les modèles hybrides (Caramazza *et al.*, 1985; Chialant et Caramazza, 1995; Frauenfelder et Schreuder, 1992; Schreuder et Baayen, 1995) et les modèles connexionnistes distribués (Gonnerman et Plaut, 2000; Gonnerman *et al.*, 2007; Plaut et Gonnerman, 2000; Seidenberg et Gonnerman, 2000). À ce jour, ces deux dernières grandes familles de modèles demeurent viables pour expliquer les processus d'identification des mots polymorphémiques. Toutefois, les théoriciens ne se sont pas penchés sur le cas des mots polymorphémiques formés de racines liées en particulier ni sur celui des mots composés. Cependant, on peut émettre des hypothèses sur le traitement de ces mots à partir des modèles hybrides et des modèles connexionnistes distribués. Les modèles hybrides pourraient expliquer le traitement morphographique de mots formés de racines latines et grecques en considérant différents facteurs : la fréquence des affixes et de la racine, la fréquence du mot entier et la transparence sémantique. De plus, comme les mots formés des mêmes racines latines ou grecques partagent habituellement des liens forts quant au sens et à la forme (p. ex. *hydratation, hydrologie, hydravion, etc.*), les modèles connexionnistes distribués pourraient aussi être particulièrement éclairants, d'autant plus qu'ils permettraient de concevoir le traitement de tous les mots dans un même système, même ceux plus opaques ou irréguliers. Les résultats de la présente étude permettront d'alimenter l'un ou l'autre de ces modèles, ou les deux, par rapport au traitement morphographique des mots composés formés de racines latines et grecques.

2.4.3 Le traitement morphographique

Par la présentation des modèles précédents, on constate que l'habileté à identifier les mots écrits peut être divisée en différents processus. Cette thèse porte particulièrement sur le traitement morphographique responsable de l'identification des mots écrits polymorphémiques. Afin d'opérationnaliser le concept de traitement morphographique, les définitions disponibles dans les écrits scientifiques ont été recensées. En psycholinguistique, le terme *traitement morphologique* [ma traduction de *morphological processing*] est parfois lié aux *connaissances morphologiques explicites* où l'individu utilise consciemment ses connaissances (Deacon *et al.*, 2008) et d'autres fois aux *connaissances morphologiques implicites*, inconscientes à l'individu (Nagy *et al.*, 2013). Effectivement, les connaissances morphologiques varieraient sur un continuum d'implicite à explicite (Deacon *et al.*, 2008). Il est donc incertain si l'individu utilise consciemment ses connaissances sur la structure des mots pour traiter, autant à l'oral qu'à l'écrit, les mots polymorphémiques qu'il rencontre. En effet, le traitement morphologique est rarement défini explicitement dans les recherches, même s'il s'agit du concept

central, que ce soit dans les études empiriques (p. ex. : Beyersmann et Grainger, 2018; Beyersmann *et al.*, 2014; Beyersmann *et al.*, 2018; Beyersmann *et al.*, 2016; Davis *et al.*, 2019; Kaczer *et al.*, 2015; Longtin et Meunier, 2005; Longtin *et al.*, 2003) ou théoriques, comme des revues de littérature (p. ex. : Diependaele *et al.*, 2012; Leminen *et al.*, 2018). Pour illustrer ce concept, les auteurs s'appuient davantage sur des modèles de traitement des mots polymorphémiques tels que ceux qui ont été présentés à la section 2.4.2 (pour une revue des modèles, voir Diependaele *et al.*, 2012). Cependant, de façon générale, le traitement morphologique réfère aux mécanismes et processus cognitifs mis en œuvre lors de l'identification de mots polymorphémiques (Amenta et Crepaldi, 2012). Le terme *traitement morphographique* est utilisé dans la présente étude qui porte sur l'identification des mots en modalité écrite et non orale, puisque la présence de la racine *graph* renvoie directement à l'écrit plutôt qu'au terme *traitement morphologique* qui est plus englobant.

2.5 Les mesures de l'identification des mots écrits

Comme les processus cognitifs impliqués en lecture ne sont pas tous observables directement par des tests psycholinguistiques comportementaux, une combinaison de mesures comportementales et psychophysiologiques permet de les inférer. D'abord, la tâche de décision lexicale, communément utilisée pour mesurer les processus d'identification des mots écrits, est présentée (2.5.1). Cette tâche sert d'abord à recueillir des données comportementales d'identification des mots écrits, mais aussi comme prétexte pour la prise de mesures psychophysiologiques. Cette section se termine avec la présentation des principales mesures psychophysiologiques utilisées lors de l'identification des mots écrits et les marqueurs pertinents pour analyser et interpréter les résultats (2.5.2).

2.5.1 La tâche de décision lexicale

La chronométrie mentale est une technique qui permet de mesurer les temps de réaction d'un participant à la suite de la présentation de différents stimulus afin d'inférer sur les processus cognitifs sous-jacents au traitement de cesdits stimulus (American Psychological Association, s.d.). Parmi les tâches les plus fréquentes en psycholinguistique, on retrouve la tâche de décision lexicale. Elle est particulièrement présente dans les recherches combinant les mesures comportementales et psychophysiologiques d'identification des mots écrits. Les participants lisent une série de mots et de pseudomots (ou de non-mots). Pour chaque item, ils doivent indiquer si la suite orthographique apparaissant à l'écran est un mot dans la langue cible. La réponse comportementale est habituellement fournie par pression d'un bouton correspondant à la réponse donnée. Les données comportementales

recueillies sont la précision et le temps de réaction des participants à la tâche de décision lexicale. La *précision* réfère au nombre de bonnes et de mauvaises réponses à la tâche de décision lexicale. Le *temps de réaction* n'est mesurée qu'au regard des réponses correctes. Elle renvoie au nombre de millisecondes écoulées entre la présentation du stimulus et le moment où le participant appuie sur la touche pour fournir une réponse. Il s'agit d'un indice de la vitesse de traitement d'identification des mots écrits.

La tâche de décision lexicale a l'avantage de s'appliquer facilement en contexte expérimental (Ferrand, 2001). Cette tâche permet aussi d'assurer un contrôle rigoureux dans les items proposés en manipulant différentes variables linguistiques (Ferrand, 2001). L'inconvénient concerne la dichotomie de la réponse comportementale, oui/non, qui affecte les résultats comportementaux, soit le temps de réaction et la précision des réponses (Ferrand, 2001). Effectivement, auprès des enfants, la précision peut se voir réduite comme il est possible que le mot présenté soit simplement un mot inconnu des jeunes participants (Henderson et Chard, 1980).

La tâche de décision lexicale peut aussi être combinée au paradigme d'amorçage, une manipulation expérimentale des relations entretenues entre les variables indépendantes. L'*amorçage* est une technique qui prévoit de présenter un item (la cible) précédé d'un autre item (l'amorce) afin de vérifier si la présentation de l'amorce favorise l'activation de la cible. L'amorçage peut se faire avec ou sans masquage. Lors de l'*amorçage non masqué*, le participant voit consciemment l'amorce lors de la tâche. Lors de l'*amorçage masqué*, l'amorce est habituellement affichée moins longtemps, soit pendant 17 à 67 ms (Ferrand, 2001), mais elle est surtout précédée d'un masquage. Cette suite de caractères de type ##### restreint la visibilité de l'amorce de sorte que les participants ne sont habituellement pas conscients d'avoir vu l'amorce, mais celle-ci est traitée automatiquement. L'amorçage masquée est d'ailleurs une technique qui permet d'avoir accès aux processus automatisés et inconscients (Milin *et al.*, 2017). On calcule l'effet d'amorçage par la différence entre les paires amorce/cible reliées et non reliées selon la variable manipulée (p. ex. lien morphologique ou sémantique entre l'amorce et la cible) quant aux mesures de temps de réaction et de précision d'identification de la cible lors de la tâche de décision lexicale (Milin *et al.*, 2017).

Dans le cadre de cette étude, les données comportementales relatives au temps de réaction des participants et à la précision des réponses lors de la tâche de décision lexicale sont recueillies. La technique d'amorçage n'est pas utilisée puisqu'elle n'est pas nécessaire pour répondre aux objectifs de

recherche. Cependant, la compréhension de cette technique sera utile pour interpréter les résultats des recherches antérieures ayant utilisé le paradigme d'amorçage.

2.5.2 Les mesures psychophysologiques

Des données psychophysologiques combinées à des données comportementales permettent d'inférer avec plus de certitude certains processus cognitifs du lecteur. Les mesures comportementales ne permettent de recueillir que deux types de données (temps de réaction et réponse) à un seul moment, soit à la fin de la tâche. Quant à elles, les mesures psychophysologiques permettent de recueillir des données en amont à la réponse du participant, soit à partir du moment où le stimulus est présenté, et peuvent expliquer les réponses comportementales obtenus ou fournir des hypothèses plausibles. Ainsi, les mesures comportementales et psychophysologiques sont complémentaires. Pour renseigner sur les processus cognitifs sous-jacents à des tâches comportementales, comme la tâche de décision lexicale, les études en neuroimagerie utilisent différents instruments de mesure.

L'électroencéphalographie (EEG) mesure l'activité électrique du cerveau qui rayonne à la surface du scalp, plus précisément celle engendrée par l'activation synchrone de grands réseaux de neurones pyramidaux du cortex cérébral (Luck, 2014). L'EEG est reconnue pour son excellente résolution temporelle grâce à son enregistrement en continu permettant de renseigner sur les processus cognitifs à la milliseconde près (Luck, 2014). Cependant, elle fournit une résolution spatiale imprécise, puisque l'origine du signal électrique détectée à une électrode donnée sur le scalp ne provient pas nécessairement de la zone couverte par l'électrode (Luck, 2014). Le signal électrique est plutôt diffus et peut provenir de sources infinies selon l'orientation et la localisation des neurones impliqués (Luck, 2014). La magnétoencéphalographie (MEG) est similaire à l'EEG de sorte qu'elle permet de mesurer les champs magnétiques induits par l'activité électrique cérébrale (Luck, 2014). Elle capte le signal magnétique perpendiculaire à celui électrique capté par l'EEG. Toute comme l'EEG, la MEG fournit une résolution temporelle précise et très fine de l'activité neuronale. Bien que sa résolution spatiale soit habituellement légèrement meilleure que l'EEG, elle reste tout de même imprécise (Luck, 2014). Aussi largement utilisée en neuroimagerie, l'imagerie par résonance magnétique fonctionnelle (IRMf) est une mesure hémodynamique qui repose sur le signal BOLD (*blood oxygen level-dependant*). Ce signal retranscrit les variations de concentration d'oxygène dans le sang, plus précisément celles de *désoxyhémoglobine*, les globules rouges dépourvus d'oxygène (Bandettini et Ungerleider, 2001). L'activation cérébrale dans une région précise nécessite une augmentation du flux sanguin permettant

une concentration plus élevée d'*oxyhémoglobine* dans le sang, soit des globules rouges porteurs d'oxygène. Par conséquent, cela entraîne une diminution du taux de *désoxyhémoglobine*, provoquant une augmentation du signal BOLD. Lorsque les neurones cessent de s'activer, le débit sanguin diminue pour rétablir la concentration d'*oxyhémoglobine* et de *désoxyhémoglobine*, réduisant ainsi le signal BOLD (Bandettini et Ungerleider, 2001). Cet instrument de mesure offre une bien meilleure résolution spatiale que temporelle (Luck, 2014). Effectivement, les changements dans le flux se produisent plus lentement que ceux électriques, c'est pourquoi cet instrument n'est pas souhaitable pour mesurer les processus cognitifs en temps réel (Luck, 2014). Dans le cadre de cette étude, l'électroencéphalographie sera utilisée pour avoir une résolution temporelle du traitement morphographique la plus précise possible. Néanmoins, dans la recension des écrits, les études réalisées grâce à la MEG demeurent intéressantes puisqu'elles fournissent des renseignements éclairants pour interpréter les résultats de recherche. Cependant, comme le signal BOLD de l'IRMf reflète des aspects différents de l'activité cérébrale comparativement à l'EEG et à la MEG (Luck, 2014), les études utilisant cette mesure ne sont pas retenues.

L'indexation de l'EEG sur le moment d'apparition d'un stimulus permet d'en extraire les potentiels évoqués, soit la modification de l'activité électrique moyenne en réponse à un type de stimulus répété de nombreuses fois (Luck, 2014). Il s'agit d'une signature psychophysologique liée au traitement cognitif d'un type d'évènement particulier (Luck, 2014). On réfère aux marqueurs de potentiels évoqués en termes d'amplitude, de polarité, de latence et de topographie. L'*amplitude* correspond au voltage observé à un temps donné (Luck, 2014). Une grande amplitude correspond à un voltage élevé, que ce soit dans l'axe négatif ou positif. Ces axes réfèrent à la *polarité* : certains marqueurs sont de polarité négative (p. ex. N400) alors que d'autres sont de polarité positive (p. ex. P600). La polarité des marqueurs est utilisée à des fins d'identification. Il est habituellement impossible de confirmer des hypothèses sur la base de la polarité (Luck, 2014). Enfin, la *topographie* d'un marqueur de potentiel évoqué renvoie à l'endroit, sur le scalp, où la mesure d'une composante de potentiel évoqué est habituellement la plus importante (Luck, 2014). Enfin, dans la nomenclature d'un marqueur de potentiel évoqué, le nombre suivant la polarité renseigne sur la *latence*, soit le moment approximatif d'apparition du marqueur après la présentation du stimulus. Par exemple, la signature cognitive N400 a un pic environ 400 ms après la présentation d'un stimulus.

De nombreux marqueurs sont impliqués dans l'identification des mots écrits et ils représentent chacun des processus cognitifs sous-jacents ou complémentaires. Par souci de clarté, les marqueurs seront présentés chronologiquement, à partir d'un temps 0, soit la présentation d'un stimulus visuel (Figure 2.4).

Le premier marqueur identifié lors d'une tâche de lecture de mots ou de pseudomots est habituellement le P100. Il s'agit d'un marqueur qui survient très tôt dans le décours temporel et qui est sensible aux caractéristiques visuelles (Luck, 2014). Son pic survient dès 100 ms poststimulus. Malgré cette précocité, il s'agit aussi d'un marqueur qui peut être sensible aux informations sémantiques et morphologiques (Davis *et al.*, 2019).

Par la suite, le marqueur N170 a un pic à environ 170 ms après la présentation d'un stimulus. Il est d'abord connu pour son association à la reconnaissance des visages (Luck, 2014). Durant l'identification des mots écrits, il témoignerait de la reconnaissance instantanée des mots (González *et al.*, 2016; Yoncheva *et al.*, 2010). La fréquence des mots serait d'ailleurs un facteur influençant son amplitude, notamment lorsque le stimulus est un mot composé (Davis *et al.*, 2019).

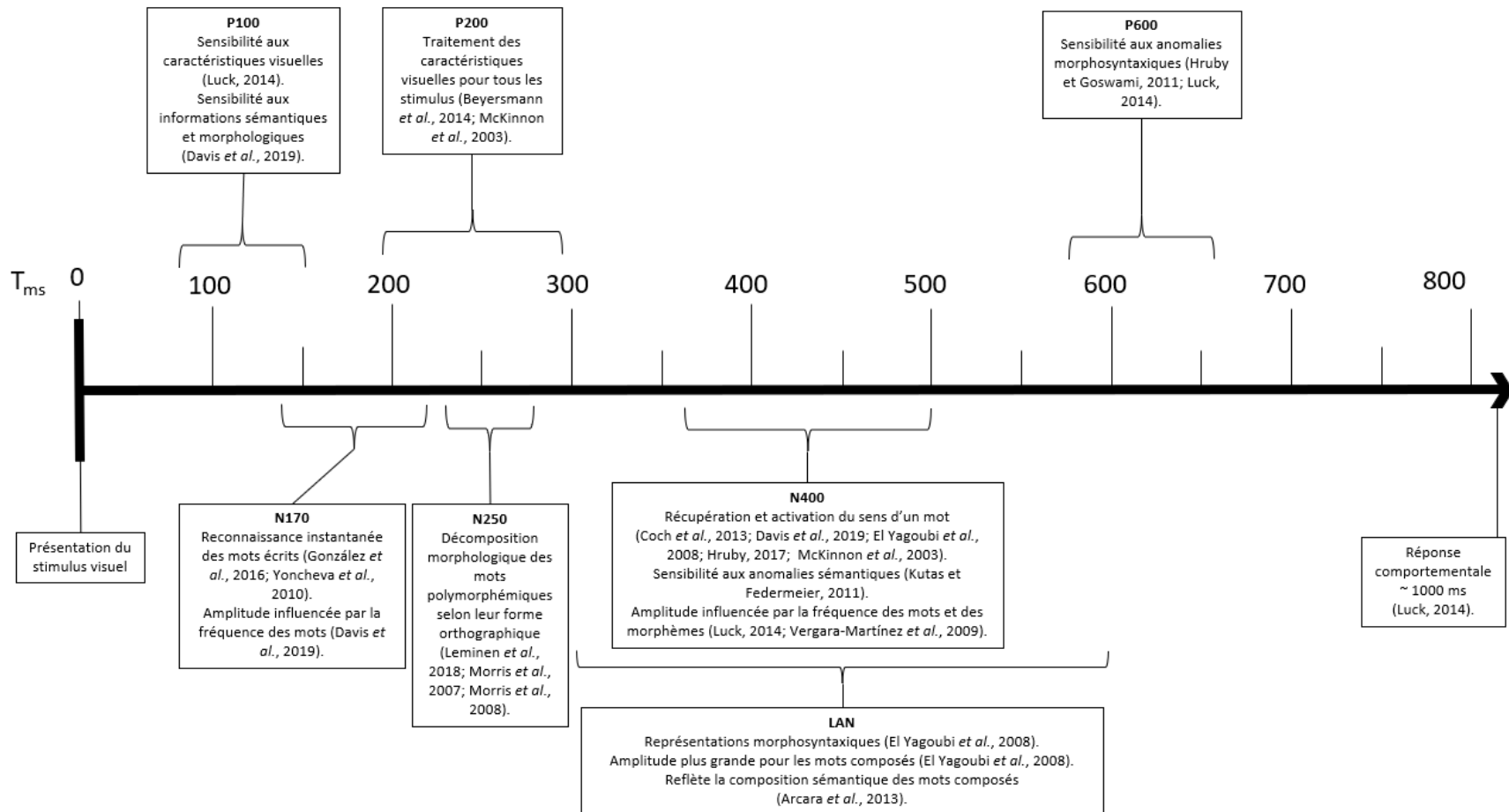
Vers 200 ms, le pic du marqueur P200 marque le traitement des caractéristiques visuelles, notamment lors de la lecture de mots (Beyersmann *et al.*, 2014). Il survient pour tous les stimulus visuels, que les mots ou pseudomots soient formés de racines libres ou liées (McKinnon *et al.*, 2003). Les familles des marqueurs N1 et P2 sont parfois même regroupées dans une même catégorie pour identifier le traitement des caractéristiques visuelles : N1-P2 (El Yagoubi *et al.*, 2008; Vergara-Martínez *et al.*, 2009).

Le potentiel évoqué négatif qui a un pic environ 250 ms après la présentation d'un stimulus est le N250. Il peut commencer dès 175 ms après la présentation du stimulus et terminer vers 300 ms (Holcomb et Grainger, 2006). Ce marqueur est sensible au chevauchement orthographique entre une amorce et sa cible (Holcomb et Grainger, 2006). Sensible à la transparence sémantique des mots et à leur structure morphologique, il témoigne aussi du traitement prélexical des mots polymorphémiques par leur décomposition rapide basée uniquement sur leur forme orthographique (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008). La topographie de ce marqueur est plutôt généralisée, mais les effets les plus grands sont habituellement observables en région frontale (Holcomb et Grainger, 2006).

Le marqueur N400 a un pic maximal environ 400 ms après la présentation d'un stimulus. Sa négativité est plus grande dans la région centropariétale (Kutas et Federmeier, 2011). Le N400 varie entre 300 ms

et 500 ms pour la plupart des auteurs (Davis *et al.*, 2019; El Yagoubi *et al.*, 2008; McKinnon *et al.*, 2003), pouvant même aller jusqu'à 600 ms (Coch *et al.*, 2013). D'autres auteurs identifient le N400 en deux temps : le N400 précoce de 250 ms à 450 ms et le N400 tardif de 500 ms à 650 ms (Beyersmann *et al.*, 2014). Dans tous les cas, le N400 informe habituellement sur la récupération et l'activation du sens d'un mot. Il est associé à l'identification sémantique des mots (Hruby, 2017). Il est aussi sensible aux anomalies sémantiques (Kutas et Federmeier, 2011). De plus, sa négativité est habituellement plus grande lors de la lecture de mots (ou de morphèmes) rares (Luck, 2014; Vergara-Martínez *et al.*, 2009) ou de non-mots et de pseudomots (Coch *et al.*, 2013; McKinnon *et al.*, 2003). Certains auteurs suggèrent que pour les mots composés, il est possible que chacune des racines entraîne un N400 (Koester *et al.*, 2007; Vergara-Martínez *et al.*, 2009). Il n'est pas encore clair si, lors de la lecture de mots polymorphémiques, le N400 est associé aux processus de décomposition morphologique ou plutôt aux processus de composition sémantique. Des résultats menés avec des mots polymorphémiques formés de racines libres et liées soutiennent l'hypothèse que le N400 serait un indice de composition sémantique qui surviendrait inévitablement après la décomposition morphologique (Coch *et al.*, 2013).

Figure 2.4 Décours temporel des principaux marqueurs de potentiels évoqués d'identification des mots écrits



Chevauchant le N400, le marqueur antérieur négatif gauche (ANG) [ma traduction de *left anterior negativity* (LAN)] est associé aux représentations morphosyntaxiques (El Yagoubi *et al.*, 2008). On peut l'observer entre 300 ms et 600 ms (Arcara *et al.*, 2013; El Yagoubi *et al.*, 2008). Sa négativité est plus grande pour les mots composés que pour ceux non composés, suggérant la formation de représentations morphosyntaxiques entre les constituants (El Yagoubi *et al.*, 2008). Il est aussi possible que ce marqueur reflète la composition sémantique lors du traitement des mots polymorphémiques, particulièrement celui des mots composés (Arcara *et al.*, 2013).

Enfin, le P600, le dernier marqueur présenté en lien avec l'identification des mots écrits, est sensible aux anomalies morphosyntaxiques (Hruby et Goswami, 2011; Luck, 2014). Le pic maximal du P600 survient environ 600 ms après la présentation d'un stimulus. Ce marqueur est associé à une vérification syntaxique lors de la lecture de phrases (Hruby, 2017).

Les marqueurs magnétiques évoqués associés à la MEG reflètent sensiblement les mêmes réponses neuronales que ceux des marqueurs de potentiel évoqués (Lopes da Silva, 2013), à l'exception du P600 qui n'a aucun marqueur magnétique évoqué équivalent. Tout comme les marqueurs de potentiels évoqués, les marqueurs magnétiques évoqués sont caractérisés par leur amplitude et leur topographie, mais pas par leur polarité. On précèdera toutefois la temporalité du marqueur par la lettre M (p. ex. M170).

Trois principaux marqueurs magnétiques évoqués renseignent sur les processus cognitifs d'identification des mots écrits, soit le M170, le M250 et le M350. Le M170 est un marqueur qui survient habituellement vers 170 ms poststimulus. Il semble indiquer un traitement visuel de la forme d'un pseudomot ou d'un mot, qu'il soit monomorphémique ou polymorphémique (Fiorentino et Poeppel, 2007). Le M250 est aussi un marqueur qui survient à la suite de la présentation d'un pseudomot ou d'un mot. Ce marqueur plus tardif, vers 250 ms, est associé au traitement phonologique ou orthophonologique du stimulus écrit (Fiorentino et Poeppel, 2007). Enfin, le M350 survient dans une fenêtre temporelle oscillant de 300 ms à 400 ms. Ce marqueur est impliqué dans l'accès sémantique (Fiorentino et Poeppel, 2007).

Dans le cadre de cette étude, les marqueurs de potentiels évoqués N250, associé à la décomposition morpho-orthographique des mots polymorphémiques (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008), et N400, témoignant de la récupération et de l'activation sémantique du mot (Coch *et al.*,

2013; Davis *et al.*, 2019; El Yagoubi *et al.*, 2008; Hruby, 2017; McKinnon *et al.*, 2003), sont d'un intérêt particulier.

2.6 Les facteurs qui influencent le développement du vocabulaire

Pour mieux comprendre le développement du vocabulaire, les chercheurs ont tenté de vérifier si la connaissance des mots pouvait être expliquée par des facteurs linguistiques. En contexte francophone, une étude théorique a permis d'identifier la fréquence des mots comme la variable la plus prédictive de la complexité du vocabulaire, et ce pour tous les niveaux scolaires confondus (Gala *et al.*, 2014). D'autres variables comme la longueur du mot et la polysémie sont aussi ressorties significatives (Gala *et al.*, 2014). Or, cette étude n'a pas porté exclusivement sur les mots du vocabulaire disciplinaire.

En contexte anglophone, des résultats de recherches empiriques ont permis de constater que certaines variables affectent la connaissance du vocabulaire disciplinaire : la polysémie, la longueur du mot, la fréquence et le taux d'abstraction (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019). Selon Nagy *et al.* (1987), seul le taux d'abstraction du mot (nommé *complexité du concept* par les auteurs) serait une variable linguistique suffisante pour nuire à la compréhension en lecture. En somme, la complexité du vocabulaire tiré des manuels scolaires autant en contexte francophone qu'anglophone semble s'accroître plus le mot est rare, long, abstrait et polysémique (Cervetti *et al.*, 2015; Gala *et al.*, 2014; Hiebert *et al.*, 2019; Nagy *et al.*, 1987).

Dans les recherches citées précédemment, la proportion exacte de mots formés de racines latines et grecques est inconnue. Cependant, il serait peu probable que la connaissance de ces mots puisse être affectée par la polysémie. En de rares occasions, les racines latines et grecques peuvent être homonymes, c'est-à-dire partager une même graphie, mais référer à deux sens distincts. C'est le cas notamment de la racine latine *équi*, du latin *equus*, qui, dans un premier temps, renvoie au sens de « cheval » comme dans *équitation* et *équestre* (Cellard, 2013; Rey et Verdier, 2017). Dans un deuxième temps, on retrouve une autre racine dont la forme orthographique est *équi* mais qui a été formée à partir du latin *aequus* qui signifie « égal » (Cellard, 2013). Cette racine sert alors à la formation des mots comme *équitable* ou *équation*. Bien que ces racines partagent leur forme orthographique (et phonologique), elles constituent deux racines distinctes, toutes deux associées à un seul sens. Le lecteur peut s'appuyer sur le contexte de la phrase et la discipline afin de départager le sens de la racine. En mathématiques, par exemple, l'utilisation de la racine *équi* signifie presque toujours « égal ».

À partir des recherches recensées, il semble que la fréquence du mot soit un facteur déterminant qui influence la connaissance du vocabulaire, et ce autant en contexte anglophone que francophone. Or, au sein des mots polymorphémiques, la fréquence des racines serait un meilleur indicateur (Reichle et Perfetti, 2003). Effectivement, il serait plus facile d'accéder au sens d'un mot lorsque la racine est fréquente (p. ex. *défaire*) plutôt que rare (p. ex. *histrionique*). La connaissance de la racine serait nécessaire pour accéder au sens des mots qui ont une faible fréquence (Bowers et Kirby, 2010). D'ailleurs, certains modèles d'identification des mots recensés précédemment (voir section 2.4.2) considèrent les propriétés de la racine du mot pour expliquer le traitement des mots polymorphémiques (p. ex. Caramazza *et al.*, 1988). Cependant, aucune des recherches présentées plus haut n'a contrôlé la fréquence de la racine.

Kieffer et Lesaux (2012) ont tenté de concevoir un modèle de représentation du vocabulaire qui s'appliquerait aux élèves de 6^e année du primaire, autant en anglais langue première qu'en anglais langue seconde. Leurs résultats ont montré qu'un modèle à trois dimensions permettait d'expliquer le mieux les résultats des élèves. Pour leur part, leur mesure de connaissance du vocabulaire était composée de l'étendue du vocabulaire, de la sensibilité contextuelle servant à l'analyse contextuelle des mots inconnus, et de la conscience morphologique. Les chercheurs n'ont pas réussi à établir de distinction entre l'étendue et la profondeur du vocabulaire, suggérant que ces concepts ne captureraient pas de différences individuelles entre les élèves, ou encore qu'il serait difficile de les mesurer séparément (Kieffer et Lesaux, 2012). Or, leurs résultats permettent de conclure que la connaissance des mots (étendue du vocabulaire) diffère des connaissances sur les mots (sensibilité contextuelle et conscience morphologique). De plus, les connaissances morphologiques représentaient un aspect central de la connaissance des mots. On pourrait donc penser que la connaissance de racines latines et grecques favoriserait la connaissance de mots du vocabulaire disciplinaire formés de ces racines.

Comme la connaissance des racines latines et grecques est particulièrement visée par l'étude et fait partie d'un des facteurs de la connaissance du mot (Nation, 2001), il est d'intérêt de se questionner si cette connaissance est aussi influencée par d'autres facteurs propres aux racines latines et grecques. Peu de recherches ont porté sur la complexité des racines latines et grecques. Néanmoins, une récente étude francophone a évalué la relation entre la connaissance des racines latines et grecques chez des élèves de 6^e année du primaire, l'origine des racines, la fréquence des mots et la fréquence des racines (Whissell-Turner *et al.*, 2022). Les résultats ont indiqué que ces facteurs n'influençaient pas la performance des

élèves. Néanmoins, les élèves étaient davantage portés à commettre une erreur en associant le sens d'une racine latine (p. ex. *dict* dans *dictée*) à un autre mot appartenant au champ lexical du mot cible (p. ex. *grammaire*). On pourrait penser que les élèves ont des représentations moins précises du sens des racines latines. Cependant, ces résultats doivent être interprétés comme un point de départ vers d'autres pistes de recherche plutôt que comme des résultats décisifs étant donné les limites soulevées par les auteurs. Pour n'en nommer que quelques-unes, la taille d'échantillon était très petite ($n = 34$) et ne permettait pas de généraliser les résultats obtenus. De plus, il est à noter que la fréquence de la racine a été mesurée post hoc. Comme cette variable n'a pas été contrôlée au moment de la conception du test de connaissance des racines latines et grecques, il s'avère qu'il n'y avait pas une représentativité équivalente de racines rares et fréquentes (Whissell-Turner *et al.*, 2022). Enfin, il est aussi possible que l'origine des racines et leur fréquence n'exercent pas d'influence au niveau comportemental (type de réponse obtenu), mais puissent influencer des niveaux cognitifs sous-jacents au traitement lexical (processus cognitifs). Effectivement, il est possible que ces facteurs n'aient pas influencé la performance comportementale des élèves lorsqu'ils ont eu à sélectionner le sens de la racine, mais que ces facteurs influencent de façon inconsciente le traitement morphographique lors de l'identification de ces mots à l'écrit. La prise de mesures continues en temps réel grâce à l'électroencéphalographie pourrait renseigner à cet effet.

D'autres pistes théoriques pourraient être explorées si on souhaite mieux comprendre les facteurs qui pourraient influencer la connaissance des racines latines et grecques. Effectivement, il a été suggéré que la productivité des racines ainsi que la transparence sémantique pourraient prédire leur connaissance (Templeton, 1983). Une racine à la base de la formation de nombreux mots et dont le sens se rapproche du sens actuel du mot entier (p. ex. *télé* qui signifie « loin » dans *télétravail*) serait mieux connue qu'une racine rare et dont le sens est opaque (p. ex. *circon* qui signifie « autour » dans *circonflexe*). À cet effet, les racines grecques seraient généralement plus transparentes que les racines latines (Templeton, 1983). Ces hypothèses théoriques demeurent à être vérifiées empiriquement.

En résumé, de nombreux facteurs sont susceptibles d'influencer la connaissance des mots chez les élèves du primaire, particulièrement les mots issus du vocabulaire disciplinaire et formés de racines latines et grecques. Parmi les facteurs individuels, l'étendue du vocabulaire, l'habileté à utiliser la stratégie d'analyse contextuelle et la conscience morphologique sont susceptibles d'influencer la connaissance des mots chez les élèves de 6^e année du primaire (Kieffer et Lesaux, 2012). Parmi les facteurs

linguistiques soulevés, la nominalisation des mots du vocabulaire disciplinaire sera contrôlée en ne sélectionnant aucun mot dérivé, ce qui ajouterait à la complexité de la tâche (Fang, 2006; Nagy et Townsend, 2012; Zoski *et al.*, 2018). Seuls des mots composés seront sélectionnés. Pour le choix des racines à la base de la formation de ces mots, la fréquence de la racine (Reichle et Perfetti, 2003; Templeton, 1983) et sa transparence sémantique (Templeton, 1983) devront faire l'objet d'une réflexion puisque ces facteurs pourraient influencer la connaissance des mots retenus. Dans le choix des items, il conviendra aussi de porter une attention particulière aux facteurs soulevés qui pourraient exercer une influence sur la connaissance du mot comme le niveau d'abstraction (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019; Nagy et Townsend, 2012), la longueur (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019) et la fréquence du mot (Gala *et al.*, 2014; Green, 2008; Nagy et Townsend, 2012). Particulièrement, la fréquence du mot est ressortie comme le facteur le plus influent sur la complexité du vocabulaire en français (Gala *et al.*, 2014).

2.7 Les facteurs qui influencent l'identification des mots écrits polymorphémiques

Dans la section précédente, la liste des facteurs linguistiques susceptibles d'influencer la connaissance des mots du vocabulaire disciplinaire et des racines latines et grecques a été dressée. Parmi ceux-ci, la structure morphologique du mot (Fang, 2006; Nagy et Townsend, 2012; Zoski *et al.*, 2018), son niveau d'abstraction (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019; Nagy *et al.*, 1987; Nagy et Townsend, 2012), sa longueur (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019) et sa fréquence (Gala *et al.*, 2014; Green, 2008; Nagy et Townsend, 2012) ajoutent à la complexité du mot. Aussi, la fréquence de la racine (Reichle et Perfetti, 2003; Templeton, 1983) et sa transparence sémantique (Templeton, 1983) peuvent influencer la connaissance des mots. Les facteurs individuels tels que l'étendue du vocabulaire, l'habileté à utiliser la stratégie d'analyse contextuelle et la conscience morphologique sont aussi corrélés à la connaissance des mots (Kieffer et Lesaux, 2012). Il convient toutefois de vérifier si, de façon plus précise, ces facteurs linguistiques et individuels pourraient exercer une influence sur les habiletés d'identification des mots polymorphémiques chez le lecteur, particulièrement sur le temps de réaction des participants et sur la précision des réponses à la tâche de décision lexicale. Donc, des recherches empiriques qui ont veillé à vérifier la contribution de facteurs linguistiques, mais aussi individuels, sur l'identification des mots écrits polymorphémiques par la prise de mesures comportementales ou psychophysologiques seront présentées. Ainsi, l'identification de facteurs linguistiques et individuels exerçant une influence sur le traitement morphographique permettra de faire des choix méthodologiques de façon à mieux contrôler et circonscrire les items de la tâche expérimentale. Il convient de préciser que seules les recherches qui portent sur le traitement morphographique, donc lors de l'identification des mots écrits, sont recensées

puisque'il s'agit de l'objet de la présente étude. Ainsi, les recherches sur le traitement morphologique en modalité auditive (p. ex. Koester *et al.*, 2007; MacGregor et Shtyrov, 2013) sont exclues de cette recension puisque l'interprétation des résultats pourrait être différente de celle en modalité visuelle. En modalité auditive, chacun des morphèmes d'un mot composé est présenté en série, soit un à la suite de l'autre. En modalité visuelle, les deux morphèmes sont présentés au même moment. Il est possible que la modalité visuelle favorise l'activation de processus combinatoires.

À notre connaissance, aucune recherche ne permet de mesurer le traitement morphographique des mots composés formés de racines latines et grecques sur la base de données comportementales et psychophysiques. En conséquence, les recherches qui portent sur le traitement morphographique des mots composés, peu importe le type de racine impliqué, sont recensées. Il convient toutefois de préciser que la composition est le procédé de formation des mots le moins étudié en ce qui concerne le traitement morphologique (Leminen *et al.*, 2018). Le peu de recherches dans le domaine et la disparité de résultats invitent à s'y attarder davantage.

La présentation des études recensées se fera selon la variable indépendante manipulée par les chercheurs, puis tour à tour pour chacune des études. D'abord, les résultats des études qui portent sur l'effet de la structure morphologique des items sur le traitement des mots écrits (2.7.1), c'est-à-dire les études qui ont portées sur le traitement des mots polymorphémiques et monomorphémiques, seront présentées. Puis, les études qui ont porté sur les effets de la fréquence des morphèmes (2.7.2), de la transparence du mot et des morphèmes (2.7.3) ainsi que de la position de la tête dans un mot composé (2.7.4) seront présentées successivement. Pour continuer, puisqu'un manque subsiste en composition, les études portant sur les effets du type de racine dans les mots dérivés feront l'objet d'une autre section (2.7.5). Une synthèse de ces différents facteurs qui influencent l'identification des mots écrits polymorphémiques permettra de clôturer cette partie (2.7.6)

2.7.1 Les études portant sur les effets de la structure morphologique des items dans le traitement des mots composés

Afin de vérifier si l'identification des mots polymorphémiques se distingue de celle des mots monomorphémiques, des recherches ont comparé leur traitement sur la base d'hypothèses comportementales. En effet, à l'aide d'une série de trois tâches d'amorçage masqué, Duñabeitia et ses collègues (2009) confirment l'accès précoce à chacun des morphèmes d'un mot composé bimorphémique transparent en Basque. Les participants pour chaque tâche étaient respectivement 30,

34 et 33 étudiants universitaires basques natifs¹¹. Un effet d'amorçage morphologique a été observé indépendamment de la position du morphème commun entre l'amorce et la cible. Pour donner un exemple en français, c'est donc dire que l'amorce *plateforme* faciliterait l'activation de la cible *platebande* au même titre que l'amorce *contrebande*. Les participants répondaient plus rapidement lorsque les cibles partageaient un morphème commun avec l'amorce que lorsque l'amorce n'était pas reliée morphologiquement. Concernant la précision en lecture, une tendance s'observait pour les mots partageant le deuxième morphème avec leur amorce, ceux-ci étaient mieux reconnus que les mots partageant le premier morphème. Comme une expérimentation subséquente n'a pas permis de montrer un effet d'amorçage orthographique, c'est-à-dire une facilitation lors de l'identification des mots lorsque l'amorce et la cible partagent une suite de lettres non associée à un morphème (p. ex. *salamandre*, *salade*), les auteurs concluent que les effets d'amorçage de la première expérimentation sont purement morphologiques plutôt que morpho-orthographiques. Les résultats de cette recherche suggèrent l'accès précoce aux morphèmes lors de l'identification des mots composés.

En contexte anglophone, l'étude de Fiorentino et ses collaborateurs (2014) a porté sur l'effet de la structure morphologique et de la lexicalité sur le traitement des mots composés dans quatre fenêtres temporelles (0 ms à 275 ms, 275 ms à 400 ms, 400 ms à 700 ms et 800 ms à 900 ms) auprès de 23 adultes monolingues anglophones (16 femmes et 7 hommes). Les participants étaient âgés de 18 à 23 ans ($M = 20$ ans). Ils étaient tous droitiers et avaient une vision normale ou corrigée à la normale. Pour répondre à l'objectif de recherche, quatre conditions expérimentales ont été développées : 1) des mots composés formés de racines libres, 2) des mots monomorphémiques, 3) des pseudomots composés (c'est-à-dire inexistantes en anglais, mais formés de deux morphèmes réels, ici des racines libres), ainsi que, 4) des non-mots (des mots formés de syllabes qui ne constituent pas des morphèmes dans la langue). Les temps de réponse obtenus auprès des participants à la suite d'une tâche de décision lexicale en lecture de mots ont permis de constater que le traitement des mots composés était plus rapide que celui des mots monomorphémiques. Les temps de réaction les plus lents étaient associés aux pseudomots composés, réservant le temps de réaction des non-mots à l'avant-dernière position. Les résultats des données psychophysologiques ont permis de confirmer l'effet de la structure morphologique et de la lexicalité au-delà des résultats comportementaux. Dans la fenêtre temporelle de 0 ms à 275 ms, aucun effet de la structure morphologique ni de la lexicalité n'a été observé. Les effets les plus précoces ont été

¹¹ Aucune information supplémentaire n'est fournie sur les participants.

décelés dans la fenêtre temporelle de 275 ms à 400 ms. Les mots inexistants (pseudomots composés et non-mots) étaient accompagnés d'une plus grande négativité que les mots existants (composés et monomorphémiques), ce qui confirme un effet de lexicalité. Concernant l'effet de la structure morphologique, les items formés de morphèmes (mots composés et pseudomots composés) étaient accompagnés d'une plus grande négativité que les mots non composés (mots monomorphémiques et non-mots) dans cette même fenêtre temporelle. Ces résultats suggèrent un traitement différent pour les mots et les non-mots, mais aussi pour les mots composés et monomorphémiques avant 400 ms suivant la présentation du stimulus. Dans la fenêtre temporelle de 400 ms à 700 ms, l'effet de lexicalité était toujours détecté. Pour l'effet de la structure morphologique des mots, il a agi différemment. Alors que les effets principaux rapportés permettaient de constater que les mots polymorphémiques (mots composés et pseudomots composés) étaient associés à une plus grande négativité que les mots non composés (mots monomorphémiques et non-mots), l'interaction entre la structure et la lexicalité indiquait que cet effet était mené principalement par les mots inexistants. C'est donc dire que l'effet de structure à cette fenêtre temporelle concernait spécifiquement les mots inexistants, les pseudomots entraînant une plus grande négativité que les non-mots. Ces résultats représenteraient probablement l'établissement dans le lexique mental de nouvelles représentations morphosémantiques pouvant être associées à des processus combinatoires, réservant une place au traitement des morphèmes. Selon les auteurs, ce traitement serait constitué d'une étape de décomposition suivie d'une étape de (re)composition. Dans la fenêtre temporelle de 800 ms à 900 ms, l'effet de lexicalité était toujours présent, les mots inexistants étaient associés à une plus grande négativité que les mots existants. L'effet de structure n'était que marginal, et ce, seulement dans la région postérieure droite du cerveau.

Dans cette expérimentation, bien que la fréquence de chacun des morphèmes ainsi que celle du mot en entier aient été contrôlées lors de la sélection des items de la tâche, la transparence sémantique n'était pas un facteur contrôlé. Comme on le verra plus loin, cette variable exerce une influence considérable sur le traitement des mots composés. On ne sait donc pas si les effets sont uniquement dus à la structure des mots, ou s'ils peuvent être expliqués au moins en partie par la transparence sémantique.

Des effets similaires ont été observés par El Yagoubi et ses collègues (2008) toujours lors d'une tâche de décision lexicale en lecture de mots, mais auprès de participants italiens natifs. L'échantillon complet comptait 20 participants, mais deux ont été retirés des analyses dû à une trop grande quantité d'artéfacts dans le signal EEG. Au final, 18 participants (10 femmes et 8 hommes) droitiers âgés de 20 à

31 ans ($M = 25$ ans) ont participé à l'étude. Aucun participant n'a rapporté des troubles neurologiques ou des problèmes de vision non corrigée. Les auteurs ont souhaité vérifier l'effet de la lexicalité, l'effet de la structure morphologique ainsi que l'effet de la position de la tête lors de l'identification de mots composés. Les résultats concernant les effets de la lexicalité et de la structure morphologique seront d'abord rapportés. L'effet de la position de la tête est rapporté plus loin, conjointement à d'autres études portant sur ce même effet (section 2.7.4). Les auteurs ont donc élaboré quatre conditions expérimentales de 28 items chacun : 1) mots composés transparents formés de la tête en position initiale, 2) mots composés transparents formés de la tête en position finale, 3) mots monomorphémiques intégrant une suite orthographique rappelant un mot en position initiale et 4) mots monomorphémiques intégrant une suite orthographique rappelant un mot en position finale. Effectivement, une particularité de ces mots monomorphémiques est qu'ils étaient formés (en début ou en fin de mot) d'une suite orthographique pouvant correspondre à une racine libre, mais n'ayant pas le rôle de morphème dans le mot cible (p. ex. *dent* dans *accident*). Pour mesurer l'effet de la lexicalité, une liste de pseudomots a été générée en changeant la position des morphèmes aux conditions 1 et 2 ainsi que des syllabes aux conditions 3 et 4. À cela se sont aussi ajoutés des mots dérivés longs en guise de remplissage afin de dévier l'attention des participants de l'objectif de la tâche. D'abord, les résultats comportementaux ont permis de détecter un effet de la lexicalité. Les participants répondaient plus rapidement aux mots qu'aux non-mots, et de façon inattendue, les réponses étaient plus précises pour les non-mots que pour les mots. Un effet de structure a aussi été détecté grâce aux temps de réponse et à leur précision : les participants fournissaient plus rapidement et précisément une réponse pour les items non composés que pour les items composés. Donc, les items composés, qu'ils soient de réels mots composés en italien ou des pseudomots composés, étaient traités plus lentement et entraînaient un plus grand nombre d'erreurs.

L'amplitude des données psychophysologiques a été observée au regard de cinq fenêtres temporelles associées à différents marqueurs de potentiels évoqués : 0 ms à 270 ms (N1-P2), 270 ms à 370 ms (ANG), 370 ms à 500 ms (N400), 310 ms à 600 ms (P300) et 500 ms à 800 ms (P600). D'abord, aucun effet de lexicalité ni de structure n'a été détecté dans la fenêtre 0 ms à 270 ms. Dans les fenêtres temporelles de 270 ms à 370 ms et de 310 ms à 360 ms, cette dernière étant incluse dans la précédente, les résultats témoignaient d'un effet de lexicalité et de structure. Les non-mots généraient une plus grande négativité que les mots. De plus, les items composés entraînaient aussi une plus grande négativité que les items non composés. Comme le marqueur ANG est reconnu pour refléter l'établissement de représentations

morphosyntaxiques, les résultats favorisent les modèles attribuant une place à la décomposition morphologique. Dans la fenêtre temporelle de 370 ms à 500 ms associée au marqueur N400, l'effet de lexicalité était présent. Le traitement des non-mots produisait une plus grande négativité au N400 que celui des mots. Pour l'effet de structure, aucun effet principal n'a été trouvé. Or, l'interaction entre la lexicalité et la structure a permis de préciser que l'effet de lexicalité était plus important avec les mots monomorphémiques que les mots composés. Enfin, dans la fenêtre temporelle de 500 ms à 800 ms associée au P600, l'effet de lexicalité était toujours présent. Les non-mots généraient une plus grande positivité, cette fois-ci, que les mots, et ce, davantage pour les items non composés que composés. Donc, de plus grands N400 et P600 ont été observés pour les items non composés que composés. Comme certains mots monomorphémiques sélectionnés comportaient une suite orthographique pouvant correspondre à une racine libre existante en italien, il est possible que cette particularité ait entraîné de l'ambiguïté lors de la lecture des mots, ce qui expliquerait les résultats obtenus pour le N400 et le P600.

Auprès d'adultes anglophones, la recherche de Fiorentino et Poeppel (2007) jumelant des données comportementales et psychophysiologiques recueillies à l'aide de la MEG a aussi permis de montrer un traitement différent entre les mots composés et les mots monomorphémiques. Les participants ($n = 12$, 8 femmes et 4 hommes) étaient tous droitiers, monolingues et ont rapporté une vision normale ou corrigée à la normale. Ils étaient âgés de 18 à 26 ans ($M = 21$ ans). Une tâche de décision lexicale a permis de comparer les processus mis en œuvre lors du traitement de mots monomorphémiques, de mots composés formés de racines libres et de non-mots. Comme pour Fiorentino et ses collègues (2014), les résultats comportementaux ont démontré que le temps de réponse des mots composés était plus rapide que celui des mots monomorphémiques. Les non-mots étaient associés aux temps de réponse les plus longs. Aussi, les mots composés, tout comme les non-mots, étaient associés à une meilleure performance à la tâche de décision lexicale que les mots monomorphémiques. Parmi les items de la tâche de décision lexicale, trois sous-groupes de 12 items ont été sélectionnés pour représenter trois niveaux de fréquence des mots : fréquence basse, fréquence moyenne, fréquence élevée. Les analyses des résultats comportementaux ont été reproduites en ajoutant cette variable de fréquence du mot entier. Concernant les temps de réponse, un effet principal de la structure du mot ainsi qu'un effet principal de fréquence ont été observés. Aucune interaction n'a été trouvée. Les résultats informent sur le fait que, de façon générale, les participants répondaient plus rapidement aux mots plus fréquents qu'aux mots plus rares. Qui plus est, les temps de réponse associés au traitement des mots composés fréquents et moyennement fréquents étaient significativement plus rapides que ceux des mots

monomorphémiques du même niveau de fréquence. La même tendance se dessinait pour les mots de basse fréquence, sans que cette différence soit autant marquée. Pour ce qui est de la précision des réponses, les items fréquents étaient mieux réussis que les items rares. Cette différence était présente entre les mots composés et monomorphémiques autant de moyenne que de basse fréquence. Cependant, pour les mots de haute fréquence, aucune différence n'était observée entre ces deux structures, les deux ayant obtenu des résultats très hauts quant à la précision (mots composés fréquents, 99 %; mots monomorphémiques fréquents, 100 %). Les résultats comportementaux soutiennent le traitement morphologique décompositionnel, même pour les mots fréquents, entraînant de meilleurs temps de réponse et de meilleurs résultats en lecture.

L'analyse des résultats de la MEG a permis d'identifier trois marqueurs communs aux traitements des items présentés : M170, M250 et M350. Apparus pour chacune des conditions, les marqueurs plus précoces, soit le M170 associé au traitement de la forme visuelle des mots et le M250 associé au traitement phonologique ou orthophonologique, ne laissent présager aucune différence entre les conditions. Dans le même ordre d'idées que les résultats comportementaux, un traitement différent était observable pour les mots composés et monomorphémiques dans la fenêtre temporelle associée au marqueur M350 (300 ms à 400 ms) impliqué dans l'accès lexical. En observant la latence du M350, Fiorentino et Poeppel (2007) ont constaté que le M350 se présentait plus tôt pour les mots composés (et les non-mots), vers 333 ms, que pour les mots monomorphémiques, vers 360 ms. Les résultats comportementaux ainsi que les résultats au marqueur M350 soutiennent la décomposition morphologique des mots composés. Effectivement, les temps de réponse plus rapides ainsi que la présence du marqueur M350 plus tôt pour les mots composés que pour les mots monomorphémiques laisserait envisager un accès aux morphèmes avant l'accès au mot entier.

Aussi, la recherche de Stites et ses collaborateurs (2016) visait à vérifier si le traitement des mots composés relevait de la décomposition ou du traitement du mot en entier auprès de participants anglophones adultes. Deux expérimentations ont été élaborées par les chercheurs, une pour mesurer le suivi oculaire, et l'autre pour mesurer l'activité électrique cérébrale dans les fenêtres temporelles associées au N400 et au P600. Un total de 25 étudiants de premier cycle universitaire anglophones ont participé à la première expérimentation¹². Les participants de la deuxième expérimentation étaient au

¹² Aucune information supplémentaire n'est fournie sur ces participants.

nombre de 21 (8 femmes et 12 hommes âgés de 18 à 23 ans, $M = 19$ ans). Ils étaient aussi tous des étudiants de premier cycle universitaires, anglophones, droitiers, sans antécédents neurologiques. Aucun participant n'a rapporté avoir été exposé à une autre langue que l'anglais avant l'âge de 5 ans.

Dans chacune des expérimentations, les auteurs ont choisi de présenter aux participants une liste de mots composés en anglais correctement écrits (p. ex. *millepatte*) ou contenant des lettres inversées, soit au sein du même morphème (p. ex. **millepattte*) ou entre les deux morphèmes (p. ex. **millpeatte*). Il a été demandé aux participants de lire les phrases présentées mot à mot et de répondre, par la suite, à des questions de compréhension. Les auteurs ont observé une activation cérébrale plus importante pour les mots contenant des lettres inversées comparativement aux mots correctement écrits, mais aucune différence significative n'a été observée entre les deux conditions inversées, tout aussi perturbantes à la lecture l'une que l'autre. Il a alors été conclu que les mots n'étaient pas décomposés en morphèmes, mais plutôt traités en entièreté. Cependant, il convient de soulever que la tâche de lecture de phrases a pu aider les lecteurs à surmonter les perturbations occasionnées par l'inversement des lettres. Effectivement, le contexte de la phrase permet d'augmenter la contribution de la sémantique et d'accéder au sens du mot bien qu'il soit mal orthographié. Une autre limite de l'expérimentation réside dans les variables de contrôle lors du choix des items, soit la transparence sémantique et la fréquence. D'abord, la transparence sémantique n'a pas été contrôlée pour le choix des items. Certains items sont plus opaques que d'autres. On retrouve par exemple le mot opaque *bookworm* qui est la composition de deux racines libres, *livre* et *ver*. Or, il a été suggéré par d'autres auteurs (p. ex. Brooks et Cid de Garcia, 2015) que seuls les mots transparents seraient décomposés. Enfin, les auteurs ont veillé à calculer les indices de fréquence du mot entier ainsi que de chacun des morphèmes. Cependant, on remarque une grande variabilité dans les données. Effectivement, bien que la moyenne de fréquence du mot entier rapporté soit de 4,5 occurrences par un million de mots, l'étendue varie de 0 (très rare) à 64 (assez fréquent). Cet écart est d'autant plus important lorsqu'on s'intéresse à la fréquence de chacun des morphèmes. Bien que les fréquences moyennes des premiers morphèmes ($M = 104$ occurrences par un million de mots) ne soient pas significativement différentes de celles des deuxièmes morphèmes ($M = 122$ occurrences par un million de mots), ces fréquences varient énormément d'un mot à l'autre, de 0 à 1093 occurrences pour le premier morphème et de 0 à 1147 occurrences pour le deuxième. Certains mots composés sont donc formés de morphèmes extrêmement rares et d'autres, de morphèmes extrêmement fréquents. Comme la fréquence du mot entier (Kuperman *et al.*, 2009) et la fréquence des constituants (Duñabeitia *et al.*, 2007; Kuperman *et al.*, 2009; Vergara-Martínez *et al.*, 2009) sont des

facteurs qui se sont révélés particulièrement influents pour le traitement des mots composés dans d'autres recherches, un meilleur contrôle de ces variables aurait pu mener à des résultats différents.

2.7.2 Les études portant sur les effets de fréquence dans le traitement des mots composés

La recherche de Goodwin et ses collègues (2013) a veillé à identifier des facteurs linguistiques et individuels contribuant à l'identification des mots polymorphémiques chez 221 jeunes adolescents (193 élèves de 7^e et 28 élèves de 8^e année). L'échantillon était composé de 107 filles et de 112 garçons¹³.

Pour répondre à l'objectif de recherche, les élèves ont été amenés à lire 39 mots polymorphémiques tirés de leurs manuels scolaires ainsi que leurs racines. Plusieurs de ces mots étaient formés d'au moins une racine latine ou grecque (p. ex. *thermosphere*, *telegraph*, *dictator*, *phonetic*, etc.). La précision lors de la lecture de mots a été évaluée. Les résultats ont démontré que l'habileté des élèves à lire correctement la racine d'un mot favorisait l'identification du mot polymorphémique. Cette habileté interagissait avec l'opacité phonologique et orthographique de la racine au sein du mot polymorphémique. Effectivement, il semble qu'il soit plus difficile pour les élèves de s'appuyer sur leur connaissance de la racine pour lire un mot polymorphémique lorsque celle-ci subit des transformations phonologiques ou orthographiques (p. ex. *meter*, *biometric*).

En contrôlant pour la lecture de la racine, les auteurs ont identifié la fréquence du mot polymorphémique ainsi que la fréquence de la racine comme facteurs linguistiques prédisant la lecture précise de mots polymorphémiques. Pour les caractéristiques individuelles, la conscience morphologique (score composite de trois tâches : jugement de relation entre une racine et un mot polymorphémique, production de pseudomots suffixés et sélection d'un mot suffixé en contexte de phrase) et la connaissance du vocabulaire (profondeur) prédisaient significativement la lecture des mots polymorphémiques. Les élèves ayant une conscience morphologique développée et une bonne connaissance du vocabulaire lisaient avec plus de précision les mots polymorphémiques tirés de leurs manuels scolaires dont une grande majorité était formée de racines latines et grecques. Les autres facteurs linguistiques (nombre de voisins orthographiques des morphèmes, fréquence moyenne de la

¹³ La somme des filles et des garçons ne totalise pas celle de l'échantillon, mais nous rapportons les données fournies par les chercheurs.

famille morphologique, nombre de morphèmes dans le mot et opacité sémantique) et individuels (compréhension en lecture) évalués n'ont pas obtenu un seuil significatif.

Ainsi, parmi les facteurs linguistiques, la fréquence de la racine et celle du mot entier influencent la lecture de mots polymorphémiques formés, pour la plupart, de racines latines et grecques. Il s'agit donc de variables à mesurer dans la tâche expérimentale de la présente étude. Aussi, la profondeur du vocabulaire et la conscience morphologique sont tous deux des facteurs individuels influençant l'identification de ces mots polymorphémiques. C'est pourquoi ceux-ci seront aussi évalués dans l'expérimentation. Cette recherche en psycholinguistique permet de mieux comprendre pourquoi les modèles d'identification des mots écrits polymorphémiques les plus adaptés aux résultats de recherche récents sont ceux qui prévoient l'apport de la morphologie.

En effet, pour certains modèles de l'identification des mots polymorphémiques écrits (p. ex. modèles hybrides) la fréquence du mot entier et la fréquence des morphèmes sont des facteurs déterminants du traitement morphographique. Or, ces modèles ayant été empiriquement testés sur un échantillon de mots fléchis et dérivés, Kuperman et ses cochercheurs (2009) ont souhaité vérifier la concordance des modèles avec les résultats obtenus lors de l'identification de mots composés quant à ces effets de fréquence. Ils ont mené une recherche sur un échantillon de 19 étudiants universitaires (12 femmes, 7 hommes) locuteurs natifs du néerlandais, tous droitiers et avec une vision normale ou corrigée à la normale¹⁴. Les participants ont été soumis à une tâche de décision lexicale combinée au suivi oculaire. Les items étaient 1250 mots et 1250 pseudomots composés de trois morphèmes et plus en néerlandais (p. ex. *hautparleur* = racine + racine + suffixe). D'abord, les auteurs ont observé un effet de lexicalité relativement au temps de réaction et à la précision des réponses lors de la tâche de décision lexicale. Les mots composés entraînaient plus d'erreurs (23 %) que les pseudomots composés (7 %). Les auteurs ont expliqué ce résultat par le fait que de nombreux mots composés étaient rares et opaques sémantiquement. Parmi les items réussis, les mots étaient traités plus rapidement que les pseudomots. Les auteurs ont ensuite voulu vérifier les effets de fréquence sur le traitement des mots composés. Les variables de fréquence sont au nombre de cinq : les indices de fréquence du mot entier, de la première racine et de la deuxième racine ainsi que les indices de la taille de la famille morphologique de chacune des deux racines. Il s'avère que les temps de réponse aux mots composés étaient tous négativement

¹⁴ L'âge des participants n'est pas fourni.

corrélés aux cinq variables observées. De ce fait, plus les indices de fréquence du mot et des racines étaient élevés et plus les familles morphologiques de chacune des racines étaient étendues, plus le mot était traité rapidement. La fréquence du mot entier est la variable qui entraînait le plus grand effet sur les temps de réponse. Les résultats du suivi oculaire vont dans le même sens pour les effets de fréquence. Effectivement, la durée du regard diminuait lorsque les mots entiers et chacune des racines étaient plus fréquents.

L'influence de nombreuses variables dans la relation à l'étude a amené les auteurs à penser que différents facteurs influençaient simultanément et de façon interactive le traitement des mots composés. Selon Kuperman et ses collègues (2009), le traitement de l'information sur les mots composés (fréquence du mot entier, fréquence des racines et taille des familles morphologiques) ne serait pas séquentiel. De ce fait, ces résultats contrediraient les modèles décompositionnels qui ne prévoiraient l'accès au mot entier qu'à la suite du traitement de chacun des morphèmes (p. ex. Taft, 1994; Taft et Forster, 1975), les modèles holistiques qui ne considèreraient que l'accès au mot entier (p. ex. Butterworth, 1983), mais aussi les modèles supralexicaux qui n'accèderaient aux morphèmes et aux familles morphologiques qu'une fois le mot entier reconnu (Giraudo, 2005; Giraudo et Grainger, 2000, 2001, 2003; Grainger *et al.*, 1991). Les auteurs (Kuperman *et al.*, 2009) s'appuient sur les modèles à double voie, notamment sur celui de Schreuder et Baayen (1995), pour proposer différents ajustements. Effectivement, considérant les effets de fréquence du mot entier et des constituants observés au même moment dans le discours temporel (Kuperman *et al.*, 2009), plusieurs voies seraient empruntées au même moment et mèneraient au traitement du mot composé. Enfin, les auteurs accordent une place centrale aux familles morphologiques de chacun des morphèmes qui seraient activées dès l'accès audit morphème et non à la suite du mot entier. Pour ces raisons, Kuperman *et al.* (2009) envisage le traitement des mots composés selon un modèle de routes multiples (plutôt que doubles) où l'accès à la forme globale du mot, à chacun des morphèmes ainsi qu'à la famille morphologique de chaque racine se ferait de façon interactive et simultanée.

Enfin, Kuperman et ses collègues (2009) ont démontré que les mots composés étaient majoritairement accompagnés de plus d'une fixation (82 %). Ainsi, à chaque nouvelle fixation, de nouvelles informations sur le mot (et les morphèmes) sont traitées par le lecteur. Selon les auteurs, les informations récupérées de la fixation précédente seraient toujours en arrière-plan, traitées ou toujours en traitement (Kuperman

et al., 2009). Or, il est largement admis que le lecteur traite ce qu'il est en train de regarder (*eye-mind hypothesis*, Just et Carpenter, 1980), ce qui invalide l'hypothèse de Kuperman *et al.* (2009).

D'autres recherches ont porté sur l'effet de fréquence des morphèmes sur le traitement des mots composés. Lors de deux tâches de décision lexicale auprès de deux groupes de participants, l'une en basque (où la tête est surtout en début de mot) et l'autre en espagnol (où la tête est surtout en fin de mot), Duñabeitia et ses collaborateurs (2007) ont comparé le traitement de mots composés bimorphémiques. Le premier groupe de participants était formé de 52 étudiants universitaires basques et, l'autre, de 44 étudiants universitaires espagnols. Tous les participants étaient des locuteurs natifs de la langue associée à leur groupe respectif, le basque et l'espagnol, et avaient une vision normale ou corrigée à la normale¹⁵. Les résultats ont montré un effet de la fréquence du deuxième morphème, avec des temps de réponse plus rapides lorsque le deuxième morphème était fréquent plutôt que rare, indépendamment de la position de la tête et de la fréquence du premier morphème. Concernant la précision en lecture, aucun effet principal de fréquence des morphèmes n'a été trouvé. Néanmoins, l'analyse par participant a montré que les mots étaient mieux réussis lorsque le deuxième morphème était fréquent plutôt que rare. Les résultats suggèrent une place importante au deuxième morphème dans le traitement des mots composés écrits (Duñabeitia *et al.*, 2007). Les chercheurs ont expliqué leurs résultats en s'appuyant sur un modèle d'activation-vérification (Paap *et al.*, 1982). Lors de la lecture de gauche à droite, le premier morphème rencontré activerait différents candidats, y compris le morphème et le mot composé cible (p. ex. *mille*, *millepattes*, *millefeuilles*, *millenium*, etc.). Par la suite, lors de la lecture du deuxième morphème, de nouveaux processus permettraient aussi l'activation d'autres candidats, toujours ledit morphème ainsi que le mot composé cible (p. ex. *pattes*, *millepattes*, *crochepattes*, etc.). La redondance de l'activation du mot cible aux deux moments confirmerait l'hypothèse de départ et permettrait la reconnaissance du mot composé cible (p. ex. *millepattes*). L'identification du mot composé écrit ne pourrait donc pas aboutir avant la lecture du deuxième morphème.

Toujours en basque, Vergara-Martínez et ses collègues (2009) ont voulu vérifier si la fréquence de chacun des morphèmes d'un mot composé influençait leur traitement en contexte de lecture de phrases. Les participants étaient 23 locuteurs natifs du basque (16 femmes, 7 hommes) âgés de 20 ans (en

¹⁵ Aucune information supplémentaire n'est fournie sur ces participants.

moyenne, $\bar{ET} = 2,67$ ans). Tous les participants avaient une vision normale ou corrigée et étaient droitiers. Aucun n'a rapporté d'historique neurologique. En manipulant la fréquence (élevée et basse) de chaque morphème, les auteurs ont établi quatre conditions expérimentales : élevée-élevée; élevée-basse; basse-élevée; basse-basse. Tous les items ont été contrôlés pour leur longueur et leur nombre de voisins orthographiques. De plus, tous les items étaient transparents. Les résultats montrent que tous les mots sont caractérisés par la vague N1-P2 associée au traitement des informations visuelles (El Yagoubi *et al.*, 2008; Vergara-Martínez *et al.*, 2009), suivie d'une large négativité commençant vers 300 ms poststimulus jusqu'à environ 700 ms. Des différences ont été observées entre les premiers et deuxièmes morphèmes. De plus larges négativités ont été observées pour les premiers morphèmes fréquents comparativement aux rares, et ce, dès la fenêtre temporelle de 100 ms à 300 ms, puis de 450 ms à 650 ms. Dans la fenêtre temporelle de 300 ms à 450 ms, aucun effet principal de la fréquence du premier morphème n'a été observé. Aussi, des effets liés à la fréquence du deuxième morphème ont été observés dans l'hémisphère droit, avec les mots composés formés d'un morphème rare en deuxième position associés à une plus grande négativité au N400 qu'un morphème fréquent, et ce, de 300 ms à environ 650 ms. De 450 ms à 650 ms, l'effet de la fréquence du premier morphème et celui du deuxième morphème étaient opposés. Effectivement, l'effet de fréquence du premier morphème était plus important lorsque ce morphème était fréquent, alors que pour le deuxième morphème, cet effet était plus grand lorsque ce deuxième morphème était rare. Les auteurs ont suggéré que chaque morphème d'un mot composé serait associé à des sous-processus différents lors de l'identification des mots composés écrits. Selon eux, leurs résultats sont aussi compatibles avec le modèle d'activation-vérification (Paap *et al.*, 1982) qui prévoit que le premier morphème d'un mot composé activerait différents candidats comme mot cible en vertu de sa fréquence, et que le deuxième morphème faciliterait la reconnaissance du mot composé. Les auteurs ont interprété l'amplitude au N400 comme l'effort requis pour intégrer chacun des morphèmes dans la représentation globale du mot.

Même si les auteurs (Vergara-Martínez *et al.*, 2009) n'ont pas contrôlé pour la position de la tête dans le mot composé, il est peu probable que l'effet constaté soit dû à ce facteur comme la tête des mots basques est presque toujours en début de mot (Duñabeitia *et al.*, 2007). Pour plus de certitude, il aurait tout de même été pertinent de contrôler ce facteur. Enfin, les auteurs ont choisi de vérifier l'identification des mots dans un contexte plus écologique, soit en contexte de lecture de phrases. Bien que ce choix méthodologique ait été critiqué pour la recherche de Stites et ses collègues (2016), Vergara-

Martínez et ses cochercheurs (2009) ont veillé à toujours présenter le mot composé comme premier mot de la phrase. Ainsi, les effets dus à l'apport du contexte étaient évités.

2.7.3 Les études portant sur les effets de transparence sémantique dans le traitement des mots composés

Pour continuer, les effets de transparence sémantique durant l'identification des mots composés écrits ont aussi fait l'objet d'études. Brooks et Cid de Garcia (2015) ont voulu vérifier l'effet de la transparence sémantique sur le traitement de mots composés lors d'une tâche de lecture de mots à haute voix. Dans le cadre de cette étude, l'amorçage combiné à l'enregistrement MEG a été utilisé auprès de 18 participants anglophones adultes âgés de 18 à 30 ans¹⁶. Tous les participants étaient droitiers et avaient une vision normale ou corrigée à la normale. Trois participants ont été retirés des analyses des données de la MEG dû à des interférences dans plus du quart de l'enregistrement.

Les items comprenaient des mots composés transparents, des mots composés opaques et des mots monomorphémiques contrôlés pour leur longueur et leur fréquence. Les résultats montraient que les mots composés (autant transparents qu'opaques) étaient décomposés avant l'accès lexical puisque, lorsqu'ils étaient amorcés par un de leurs morphèmes, les temps de réponse des mots composés étaient plus courts que ceux des mots monomorphémiques. Aussi, seuls les mots composés transparents seraient associés aux processus de composition morphologique comme une plus grande activation a été observée pour ces mots par rapport aux mots monomorphémiques. Aucune différence n'a été constatée pour les mots opaques. Les différences présentes de 250 ms à 470 ms pourraient être associées à une étape de l'accès lexical où l'accès aux morphèmes contribue à l'accès au sens du mot. Les différences observées de 430 ms à 600 ms réfèreraient à l'accès lexical dont l'accès aux représentations sémantiques et phonologiques. Les résultats suggèrent une sensibilité à la structure des mots, même en production de mots, en deux étapes. La première étape impliquerait la décomposition, une étape indépendante de la sémantique, et la deuxième étape, celle de composition, serait régie par la sémantique.

Contrairement à la recherche énoncée précédemment qui a observé la transparence sémantique du mot entier, Davis et ses collègues (2019) ont plutôt affiné ce critère en contrôlant pour la transparence sémantique de chacun des morphèmes du mot composé. Les participants sélectionnés étaient 22

¹⁶ Aucune information sur le sexe des participants n'est disponible.

étudiants de premier cycle universitaire, tous des locuteurs natifs de l'anglais, et âgés de 20,5 ans en moyenne (15 femmes, 7 hommes). Aucun participant n'a rapporté de difficultés langagières, attentionnelles, motrices ou psychiatriques et seulement trois participants étaient gauchers. Tous les participants avaient une vision normale ou corrigée à la normale. Parmi ces 22 participants, trois ont été sélectionnés pour valider la tâche expérimentale réalisée avec l'EEG.

Quatre conditions expérimentales ont été élaborées en alternant entre morphème transparent et opaque: transparent-transparent; opaque-transparent; transparent-opaque; opaque-opaque. À l'aide d'une tâche de décision lexicale, les auteurs ont vérifié les effets de la fréquence du mot composé et de la transparence sémantique des morphèmes en contexte de lecture de mots en anglais. Les auteurs ont constaté des temps de réponse plus rapides lorsque le deuxième morphème était transparent comparativement à opaque. Aussi, les temps de réponse étaient influencés par la fréquence du mot entier ainsi que la fréquence du deuxième morphème. Une fréquence globale élevée du mot jumelée à une basse fréquence du deuxième morphème entraînait des temps de réponse plus rapides. Concernant les résultats recueillis à la suite de l'enregistrement électroencéphalographique, les auteurs ont témoigné d'un effet de fréquence du mot entier et de chacun des morphèmes sur l'amplitude du P100. Les auteurs ont interprété ces résultats comme un accès simultané au mot entier et à chacun des morphèmes, et ce, aussi tôt qu'environ 100 ms poststimulus. Les marqueurs précoces dans le déroulement temporel d'identification des mots écrits composés seraient sensibles aux variables lexicosémantiques. Cependant, la fréquence du mot entier n'a pas contribué de façon significative à l'amplitude du N400 contrairement à la transparence du deuxième morphème. En bref, le deuxième morphème d'un mot composé joue un rôle central dans l'identification des mots écrits. Dans ce contexte, ce morphème correspond à la tête du mot. Effectivement, en anglais, cette dernière est toujours à droite (Dressler, 2006; Libben, 2006). Les résultats ne sont donc pas nécessairement généralisables à une langue comme le français où la position de la tête est variable. Les auteurs ont suggéré que leurs résultats soutiennent les modèles de routes multiples interactives (Kuperman *et al.*, 2009).

2.7.4 Les études portant sur les effets de la position de la tête dans le traitement des mots composés

La tête est un élément caractéristique des mots composés, agissant comme morphème donneur des informations sémantiques et même syntaxiques (Dal et Amiot, 2008; Dressler, 2006). Il s'agit d'un facteur qui n'est pas toujours contrôlé dans les études qui s'intéressent à l'identification des mots

composés, mais qui pourraient influencer l'interprétation des résultats obtenus. Des auteurs ont donc souhaité mesurer l'effet de la position de la tête sur le traitement des mots composés.

Un des objectifs de la recherche d'El Yagoubi et ses collègues (2008) visait à évaluer le décours temporel de mots composés formés de deux noms en italien et de l'effet de la position de la tête (étude décrite plus en détail à la section 2.7.1). Dans une tâche de décision lexicale combinée à l'enregistrement électroencéphalographique, les 20 participants adultes (10 femmes, 8 hommes) ont eu à lire 56 mots composés transparents (28 mots formés de la tête en position initiale et 28 mots formés de la tête en position finale) en plus de mots non composés et de non-mots. Alors qu'aucune différence quant à la position de la tête n'a été observée dans les résultats comportementaux (temps de réaction des participants et précision des réponses à la tâche de décision lexicale) ni avant 300 ms après la présentation du stimulus, une différence au marqueur P3 (vers 300 ms) a été révélée. Effectivement, les mots composés dont la tête est en position finale étaient associés à une plus grande positivité que les mots composés dont la tête est le morphème initial. Le même effet s'est maintenu au marqueur P600 (500 ms à 800 ms). Cette différence n'a pas été observée pour les mots non composés ni pour les non-mots. Il a donc été suggéré que la position de la tête était un élément déterminant dans le traitement des mots composés. En italien, les mots composés formés d'une tête comme dernier morphème ne reproduisent pas l'ordre syntaxique habituel. Ces mots ont pour la plupart été empruntés au latin ou à d'autres langues, comme l'anglais, où les têtes sont positionnées à droite. De ce fait, les auteurs suggèrent que la présence d'une tête en fin de mot, plutôt qu'en début de mot, demanderait plus de ressources cognitives au lecteur pour traiter ce mot composé.

Toujours sur les mots composés de deux noms en italien, Arcara et ses cochercheurs (2013) ont mené une recherche similaire à El Yagoubi et ses collègues (2008) aussi auprès de 24 étudiants universitaires italiens à développement typique. Tous les étudiants étaient des locuteurs natifs de l'italien et droitiers, en plus de rapporter une vision normale ou corrigée à la normale. Deux participants ont été retirés de l'échantillon dû à la mauvaise qualité de l'enregistrement électroencéphalographique. L'échantillon final était constitué de 22 étudiants universitaires (15 femmes, 7 hommes) âgés de 19 à 31 ans ($M = 21,36$ ans). Arcara et ses cochercheurs (2013) ont vérifié si le traitement visuel des mots était influencé par la position de la tête. Lors d'une tâche de décision lexicale, les mots composés formés de deux noms variaient quant à la position de la tête, en première ou dernière position. Comme El Yagoubi et ses collègues (2008), des mots monomorphémiques dans lesquels on retrouve, en début ou en fin de mot,

une suite orthographique qui correspond à un autre mot dans la langue (p. ex. *dent* dans *accident*) ont été ajoutés aux items. Enfin, les auteurs ont aussi ajouté des mots composés formés d'un verbe et d'un nom (p. ex. *portemanteau*) et des non-mots. Les items ont été contrôlés pour leur longueur, leur fréquence ainsi que le nombre de voisins orthographiques du mot composé et de chacun des morphèmes. Or, contrairement à El Yagoubi et ses collègues (2008), la transparence sémantique n'a pas été contrôlée. Les résultats comportementaux ont montré un traitement plus long pour les mots composés d'un verbe et d'un nom ainsi que pour ceux dont la tête est en dernière position plutôt qu'en première. Les résultats des potentiels évoqués vont aussi dans ce sens, signalant une négativité au marqueur ANG plus grande pour les mots composés d'un verbe et d'un nom ainsi que ceux formés d'une tête en position finale plutôt qu'initiale. Cet effet du ANG a été observé dans la fenêtre temporelle débutant vers 400 ms poststimulus jusqu'à 600 ms. Ces résultats ont permis de confirmer ceux d'El Yagoubi et ses collaborateurs (2008). Bien que ces derniers n'eussent pas pu trouver de résultats comportementaux significatifs, ils avaient aussi observé une différence au marqueur P3 suggérant un traitement visuel plus important pour les mots composés formés d'une tête en position finale plutôt qu'initiale.

Parmi les limites de l'étude, le nombre élevé de variables contrôlées par les auteurs a occasionné un nombre restreint d'items dans chacune des conditions (24 par condition). Les auteurs ont aussi précisé que les effets de la position de la tête observés dans leur étude ne pouvaient pas être généralisés à d'autres mots composés que ceux formés de deux noms en italien. D'autres recherches devront être réalisées pour confirmer leurs résultats. Enfin, bien que les auteurs se soient intéressés à la fenêtre temporelle de 400 ms à 600 ms, ils n'ont pas porté leur attention sur le marqueur N400, mais seulement sur le marqueur ANG.

Il convient de rappeler qu'en français comme en italien, la position de la tête est variable d'un mot à l'autre, parfois comme premier morphème, parfois comme dernier morphème. Cependant, pour les mots composés formés de racines latines et grecques en français, la tête est habituellement à droite (p. ex. *morphologie*). De ce fait, l'effet de la position de la tête ne devrait pas influencer le traitement des mots composés formés de racines latines et grecques en français. Toutefois, en s'appuyant sur les résultats d'El Yagoubi et ses collègues (2008) et d'Arcara et ses collègues (2013), il est possible que les mots formés de racines latines et grecques nécessitent plus de ressources cognitives que les mots composés où la tête serait placée en début de mot.

Une limite importante des études présentées dans les dernières sections (2.7.1 à 2.7.4) concerne la structure des items sélectionnés. Tous les mots composés de ces études sont formés de racines libres. Aucune n'a observé le traitement de mots composés formés de racines liées, comme les racines latines et grecques. Comme ces morphèmes ne sont pas autonomes dans la langue, il est possible que le traitement des mots qu'ils composent ne soit pas similaire aux mots composés formés de racines libres. À défaut d'avoir trouvé de recherches qui explorent le traitement morphographique de mots composés formés de racines liées, les quelques études qui prennent en considération des racines liées, mais au sein de mots dérivés, ont été recensées. Ces études peuvent renseigner sur le traitement morphographique des mots formés de racines latines et grecques.

2.7.5 Les études portant sur les effets du type de racine dans le traitement des mots dérivés

Une recherche anglophone a veillé à vérifier le traitement des mots préfixés formé de racines libres et liées (Forster et Azuma, 2000). Par une tâche de décision lexicale combinée au paradigme d'amorçage masqué, trois expérimentations ont été menées auprès d'étudiants du premier cycle universitaire (respectivement 54, 56 et 30 étudiants)¹⁷. Les items comprenaient des mots formés de la même racine liée en guise d'amorce et de cible (p. ex. *implore* – *EXPLORE*); de la même racine libre (p. ex. *trust* – *DISTRUST*); des mots monomorphémiques utilisés à la fois comme amorce et comme cible en guise de contrôle (p. ex. *gallery* – *GALLERY*); et des non-mots pour le bon déroulement de la tâche de décision lexicale. Dans les expérimentations 2 et 3, les items de contrôle ont été remplacés par des mots non reliés sémantiquement ou morphologiquement, mais présentant un chevauchement orthographique (p. ex. *ferment* – *CEMENT*) afin d'isoler le traitement orthographique. Les résultats ont suggéré un traitement similaire pour tous les mots polymorphémiques, qu'ils soient formés d'une racine libre ou liée, et ce même après le contrôle du chevauchement orthographique (Forster et Azuma, 2000). Effectivement, les résultats ont montré des effets d'amorçage masqué, c'est-à-dire que lorsque l'amorce et la cible étaient reliées morphologiquement, la cible était traitée plus rapidement et avec plus de précision. Cependant, aucune différence n'a été observée entre les types d'amorces. Des analyses post hoc ont aussi révélé que les effets d'amorçage étaient partiellement contrôlés par la productivité de la racine (Forster et Azuma, 2000). Une racine peu productive ne produisait pas d'effet d'amorçage. Il est toutefois à noter que seuls les mots préfixés (p. ex. *erupt*) ont été considérés dans le calcul de la productivité de la racine, qu'elle soit libre ou liée. Les mots formés de plusieurs affixes (p. ex. *corruptive*)

¹⁷ Aucune information supplémentaire n'est fournie sur les participants.

ou les mots composés (p. ex. *bankrupt*) n'ont pas été considérés. Ce choix méthodologique fausserait le calcul de productivité de la racine.

Aussi, Forster et Azuma (2000) ont estimé que les effets observés étaient de l'ordre du traitement morphologique plutôt que du traitement sémantique. Effectivement, ils ont énoncé que les mots dérivés de la même racine liée n'étaient pas sémantiquement reliés. Cependant, il est impossible de soutenir avec conviction cet argument comme les racines liées sont aussi porteuses de sens. Deux mots partageant une même racine liée partagent un sens commun (p. ex. *inclus, exclus ; photosynthèse, photographie*). La recherche gagnerait à mieux contrôler les relations sémantiques entre les mots. Enfin, une autre limite méthodologique mérite d'être soulevée quant au choix des items pour l'amorçage masqué. Effectivement, les cibles préfixées formées d'une racine libre (p. ex. *DISTRUST*) étaient précédées de leur racine comme amorce (p. ex. *trust*). Au contraire, les cibles préfixées formées d'une racine liée (p. ex. *REVIVE*) étaient précédées d'un autre mot préfixé formé de la même racine comme amorce (p. ex. *survive*) et non de leur racine (p. ex. *vive*). Il aurait été intéressant de vérifier si les résultats obtenus se maintiennent malgré cet ajustement.

Contrairement aux résultats de Forster et Azuma (2000), l'étude de Coch et ses cochercheurs (2020) a montré que les mots formés de racines liées et libres n'étaient pas traités de la même façon. Pour ce faire, 80 étudiants anglophones (40 femmes, 40 hommes) âgés de 18 à 23 ans ($M = 20$ ans, $ÉT = 1$ an et 4 mois) ont vécu une tâche de décision lexicale. Tous les participants étaient droitiers, monolingues et des locuteurs natifs de l'anglais. Aucun participant n'a rapporté des difficultés de développement. Sept participants ont rapporté des difficultés de prononciation avec certains phonèmes en bas âge, des difficultés qui se sont résorbées.

Les items de la tâche sont directement tirés de l'étude de Coch et ses cochercheurs (2013), présentée plus loin. Ainsi, la tâche de décision lexicale comprenait des mots et des pseudomots formés de racines liées (p. ex. *discern; *disject*), des mots et des pseudomots formés de racines libres (p. ex. *cobweb; *coblinc*) ainsi que des mots monomorphémiques (p. ex. *garlic*) et de non-mots (p. ex. **gartus*). Tous les items ont été contrôlés pour leur longueur, leur nombre de voisins orthographiques ainsi que la fréquence des bigrammes et trigrammes. De plus, les mots ont été contrôlés pour leur fréquence. Selon les auteurs, des analyses de variance ont permis de constater des effets de lexicalité et de structure morphologique des mots en précision et en temps de réaction durant la tâche de décision lexicale.

D'abord, concernant l'effet de lexicalité, les participants réussissaient mieux les mots que les pseudomots et les non-mots et y répondaient plus rapidement. Aucune différence n'a été observée entre les pseudomots et les non-mots. Concernant l'effet de la structure, les participants fournissaient des réponses plus précises et plus rapides aux mots de la condition de contrôle (mots monomorphémiques et non-mots) qu'aux mots et aux pseudomots formés de racines liées, suivis ensuite par les mots et les pseudomots formés de racines libres. Une interaction entre la lexicalité et la structure a aussi été trouvée. Pour les mots, la lecture des mots monomorphémiques était mieux réussie et plus rapide que celle des mots formés de racines liées. Les mots formés de racines libres entraînaient plus d'erreurs et un plus long temps de réponse. Pour les non-mots et les pseudomots, un patron légèrement différent s'est dessiné. Les non-mots étaient réussis plus rapidement et plus précisément que les pseudomots formés de racines libres. Dans cette condition, ce sont les pseudomots formés de racines liées qui étaient les plus difficiles et qui prenaient le plus de temps à être traités.

Pour plus de puissance statistique, les auteurs ont repris ces mêmes analyses dans des modèles mixtes pour mieux rendre compte des effets fixes et aléatoires de façon simultanée. Des modèles différents pour les mots et les non-mots (ou pseudomots) ont été réalisés. Des résultats légèrement différents de leurs analyses précédentes sont ressortis. Pour les mots, alors que les analyses de variance suggéraient des différences significatives entre les trois conditions relatives à la structure morphologique des items (monomorphémique, racine libre, racine liée), les modèles mixtes ne témoignaient d'aucune différence entre les mots monomorphémiques et les mots formés de racines liées. De ce fait, ces deux conditions étaient similairement mieux réussies, et plus rapidement, que les mots formés de racines libres. Pour les items inexistant dans la langue, la différence s'observait plutôt entre les non-mots et les pseudomots, sans de différence selon le type de racine. Ainsi, sur les plans du temps de réaction des participants et de la précision des réponses à la tâche de décision lexicale, les non-mots entraînaient de meilleurs résultats que les pseudomots. Pour résumer ces modèles mixtes, les mots formés de racines libres et ceux formés de racines liées ne seraient pas traités de façon similaire. Or, aucune différence quant à la précision et au temps de réaction ne permettait de distinguer le traitement des pseudomots formés de racines libres ou liées. Globalement, des deux types d'analyse (analyses de variance et modèles mixtes), il est ressorti que les mots formés de racines liées étaient traités plus rapidement et avec plus de précision que ceux formés de racines libres, suggérant un traitement plus efficace pour les mots formés de racines liées. Les auteurs ont recommandé un enseignement différencié pour ces deux types de racines (Coch *et al.*, 2020). Cependant, comme les mots formés de racines liées étaient tous des mots dérivés et que les mots

formés de racines libres étaient tous des mots composés, on ne sait pas si les résultats sont déterminés par le type de racine (libre ou liée) ou encore par le procédé de formation des mots (dérivation ou composition). On ne sait donc pas si ces mots de la condition *racines liées* sont traités plus efficacement parce qu'ils sont formés de racines liées ou bien parce qu'ils sont dérivés. Un meilleur contrôle des items permettrait d'éclaircir ce questionnement.

Les études de Forster et Azuma (2000) et de Coch et ses collègues (2020) permettent de constater des résultats contradictoires quant à la précision et au temps de réaction en lecture de mots dérivés formés de racines liées. Le peu d'études dans le domaine et le contrôle des items peut expliquer la divergence des résultats, soit que, d'une part, les mots formés de racines libres et liées sont traités de la même façon par le lecteur (Forster et Azuma, 2000) et, d'autre part, qu'ils sont traités différemment (Coch *et al.*, 2020). Les études suivantes combinent des mesures comportementales à des mesures psychophysiques pour mieux comprendre le traitement de ces mots et ainsi situer avec plus de précision le décours temporel relatif à ce traitement.

En contexte anglophone, l'étude de McKinnon et ses collègues (2003) est, à notre connaissance, la première étude en neuroimagerie à s'intéresser au traitement des mots polymorphémiques formés de racines liées. Grâce à l'enregistrement électroencéphalographique, les chercheurs ont élaboré trois expérimentations afin de comparer l'activation du N400 lors du traitement de mots et de pseudomots formés de racines libres et liées durant une tâche de décision lexicale. Les participants étaient 36 étudiants universitaires droitiers (25 femmes et 11 hommes) âgés de 18 à 29 ans. L'échantillon a été subdivisé en trois sous-groupes de 12 participants, un sous-groupe pour chaque expérimentation.

Les items de la tâche de décision lexicale comprenaient des mots dérivés formés de racines liées (p. ex. *submit*), des mots composés (p. ex. *bookmark*) ou dérivés (p. ex. *muffler*) formés de racines libres, des pseudomots formés en combinant des racines liées (p. ex. **promit*), ainsi que des non-mots créés en modifiant la position des lettres des mots formés de racines libres (p. ex. **moobkark*). Les résultats comportementaux ont porté exclusivement sur la précision en lecture. Les non-mots ont entraîné significativement moins d'erreurs que toutes les autres conditions expérimentales. Les résultats psychophysiques ont d'abord montré que tous les items ont déclenché une activation aux marqueurs N1 et P2 associés au traitement des caractéristiques visuelles des mots. Uniquement pour les pseudomots formés de racines liées, une légère hausse dans l'amplitude du P2 (150 ms à 300 ms) est

observée, à certaines électrodes seulement. Dans la fenêtre temporelle associée au marqueur N400 (300 ms à 500 ms), les non-mots ont entraîné une négativité plus grande que toutes les autres conditions. Aucune différence n'a été observée quant à l'amplitude du N400 pour les mots formés de racines libres, pour ceux formés de racines liées ainsi que pour les pseudomots, eux aussi formés de racines liées. La réponse cérébrale obtenue lors du traitement des mots formés de racines liées était donc similaire à celle des pseudomots formés de ces mêmes racines. Cela pourrait s'expliquer par le fait que ces pseudomots sont formés de morphèmes porteurs de sens, contrairement aux non-mots. Il serait alors possible d'attribuer un sens à ces pseudomots, même s'ils n'existent pas dans la langue cible, ici l'anglais. Les résultats équivalents pour les mots polymorphémiques et les pseudomots formés de racines liées suggèrent que les racines, qu'elles soient libres ou liées, seraient stockées dans le lexique mental pour permettre la décomposition morphologique de tous les mots ou pseudomots polymorphémiques. Aussi, la deuxième et la troisième expérimentation de la même étude ont permis de contrôler la similarité orthographique entre les mots et les pseudomots ainsi que la répétition des affixes d'un item à l'autre. Donc, ces résultats pourraient être interprétés comme éclairant le traitement morphographique des mots formés de racines liées, sans influence du traitement orthographique. Comme les auteurs pensent qu'il est difficile, voire impossible, d'attribuer un sens à un mot formé de racines liées en s'appuyant uniquement sur le sens de ses morphèmes, McKinnon et ses collègues (2003) s'appuient sur les modèles hybrides pour expliquer leurs résultats. Effectivement, selon leur hypothèse, une représentation globale du mot serait de mise. Il est à noter qu'il aurait été préférable de ne s'intéresser qu'à un seul procédé de formation des mots : l'hétérogénéité des items a pu influencer l'interprétation des résultats obtenus.

Comme il a été observé que des mots et des pseudomots partageant les mêmes racines entraînaient des réponses cérébrales semblables (McKinnon *et al.*, 2003), Coch et ses collaborateurs (2013) ont voulu vérifier si le marqueur N400 identifié lors du traitement morphographique des mots polymorphémiques était un indice de décomposition morphologique ou de composition sémantique toujours auprès de participants adultes ($n = 16$, 8 femmes, 8 hommes) âgés de 18 à 22 ans ($M = 20$ ans et 8 mois, $ÉT = 1$ an et 4 mois). Tous les participants étaient monolingues anglophones, à développement typique et droitiers. Les auteurs ont contrôlé trois conditions expérimentales. Les mots formés de racines liées étaient tous des mots dérivés (p. ex. *discern*) auxquels des pseudomots étaient appariés en interchangeant les morphèmes (p. ex. **disject*). Des mots composés formés de racines libres constituaient la deuxième condition expérimentale (p. ex. *cobweb*). Encore une fois, pour former des pseudomots, des morphèmes ont été interchangés entre les items (p. ex. **coblinc*). Enfin, la condition contrôle comprenait des mots

monomorphémiques dissyllabiques (p. ex. *garlic*) et des non-mots appariés en interchangeant, cette fois-ci, les syllabes (p. ex. **gartus*). D'abord, un effet principal de lexicalité a été observé pour le temps de réaction, mais pas pour la précision. Les mots étaient lus plus rapidement que les non-mots ou les pseudomots. Des effets principaux quant au type de racine étaient aussi présents, autant pour la précision que pour le temps de réaction. Pour la précision, les mots de la condition de contrôle étaient mieux réussis que ceux formés de racines liées. Les mots formés de racines libres entraînaient le plus d'erreurs. Pour le temps de réaction, les participants répondaient plus rapidement aux mots de la condition de contrôle qu'aux mots polymorphémiques, sans distinction quant au type de racine, libre ou liée. Une interaction entre l'effet de lexicalité et celui du type de racine était aussi présente, rapportant un patron similaire séparément pour les mots et les non-mots.

L'effet principal dû au type de racine a entraîné des différences d'amplitude au marqueur N400. Les items formés de racines libres produisaient une plus grande négativité que les items de contrôle. La plus petite négativité était générée par les items formés de racines liées. L'effet de lexicalité est aussi observable au N400 puisque les pseudomots (et non-mots) entraînaient une plus grande négativité que les mots. Aucune interaction entre le type de racine et la lexicalité n'est ressorti significatif. Selon l'interprétation des auteurs, cette sensibilité lexicale du N400 associerait ce marqueur à la composition sémantique plutôt qu'à la décomposition morphologique (Coch *et al.*, 2013). Si le N400 avait plutôt été un indice de décomposition morphologique, l'amplitude aurait été similaire dans chaque condition, sans distinction entre les mots et les pseudomots sauf dans la condition de contrôle (Coch *et al.*, 2013). Par exemple, le mot *discern* et le pseudomot **disject* auraient entraîné le même traitement puisqu'ils partagent une même racine liée. Or, cela ne signifie pas nécessairement que les mots polymorphémiques formés de racines liées ne sont pas décomposés avant l'accès lexical, mais simplement que ce ne serait pas le marqueur N400 qui serait associé à cette étape. Un autre marqueur pourrait y être associé comme le N250 (Leminen *et al.*, 2018).

Il est à noter que, dans la recherche de Coch et ses collaborateurs (2013), plus d'une variable est manipulée au travers des conditions : le type de racine (libre et liée) et le procédé de formation des mots (dérivation et composition). Afin de mieux interpréter les résultats de l'étude, il serait pertinent de ne manipuler qu'une seule variable à la fois. En ce sens, si l'intérêt est de vérifier le traitement des mots formés de racines liées et libres, il serait pertinent de comparer des items qui sont tous composés ou tous dérivés.

Toujours en contexte anglophone, grâce à la MEG, la recherche de Solomyak et Marantz (2010) souhaitait vérifier le traitement morphologique précoce des mots dérivés formés de différents types de racine auprès de 9 participants anglophones droitiers âgés de 19 à 29 ans ($M = 23,3$ ans)¹⁸. Les auteurs identifient trois types de racine : libre, liée et unique. Pour les auteurs, les racines uniques sont des racines qui n'apparaissent pas dans d'autres mots de la langue (Solomyak et Marantz, 2010). Comme explicité plus loin, ce n'est pas réellement le cas selon les items sélectionnés (p. ex. *lateral*). Par une tâche de décision lexicale, Solomyak et Marantz (2010) ont pu constater que les résultats comportementaux étaient significativement semblables pour les mots formés de tous les types de racines. Le temps de réponse était cependant affecté par la fréquence du mot et de son lemme. Les résultats au marqueur M130 indiquaient un effet significatif de la fréquence de l'afixe pour toutes les conditions, ce qui reflèterait la détection des propriétés orthographiques des mots affixés. Plus tard, au marqueur M170, l'effet de la fréquence de l'afixe était ici interprété comme la véritable décomposition du mot affixé pour toutes les conditions. Compte tenu de l'effet de la fréquence du lemme observé uniquement pour les racines libres et liées, et non pour les racines uniques, le marqueur M350 a été associé à l'accès lexical par les lemmes. Aucun effet de fréquence du mot entier n'a été observé dans les résultats recueillis par la MEG. De ce fait, les résultats de la présente recherche soutiennent le fait que les mots polymorphémiques formés de racines libres et liées seraient décomposés de façon similaire, ce qui assurerait ensuite l'accès lexical, contrairement aux racines uniques (Solomyak et Marantz, 2010).

Les résultats de cette recherche doivent être considérés avec prudence. D'abord, l'échantillon est relativement petit, ne comptant que neuf adultes. Aussi, les auteurs ont veillé à calculer la fréquence de l'afixe, mais aussi sa fréquence orthographique, c'est-à-dire la fréquence à laquelle la suite orthographique dudit affixe se retrouve en fin de mot, qu'il soit ou non un affixe (p. ex. *cahier*, *banquier*). Malgré le fait que ce contrôle soit rigoureux, les mots composés n'ont pas été considérés dans ces analyses de fréquence. Enfin, les listes de mots proposées pour chacune des catégories d'items demeurent la limite majeure de l'étude. Premièrement, une analyse fine des items suggérés a permis de constater que parmi la liste de mots formés de racines libres se trouvaient des mots formés de racines liées (p. ex. *dict* dans *predictable*). Les auteurs avaient probablement identifié *predict* comme racine libre, ce qui ne correspond pas aux définitions données aux sections 2.1.2 et 2.1.3. Deuxièmement, il est intéressant que la liste de mots formés de racines liées contienne des racines latines (p. ex. *equ* dans

¹⁸ Aucune information sur le sexe des participants n'est fournie.

equable et grecques (p. ex. *kin* dans *kinetic*). Cependant, il est curieux d'apercevoir des mots comme *valuable* dans la catégorie des racines liées, alors que d'autres mots comme *figural* se retrouvent dans les racines libres. Pourtant, ces items sont formés par la même règle morphologique, soit le retrait du *e* final de la racine (*value*, *figure*) pour adjoindre le suffixe. Troisièmement, on retrouve de nombreux mots formés de racines latines et grecques (p. ex. *lateral*, *rational*, *culinary*, *mammary*), donc de racines liées, au sein des racines uniques. Ces racines ne sont pas des morphèmes uniques puisqu'ils sont à la base d'autres mots. Voici quelques exemples en français avec la racine latine *lat* signifiant « côté » qui s'appliqueraient aussi à l'anglais : *latéral*, *latéralement*, *bilatéral*, *quadrilatère*, etc. Quatrièmement et dernièrement, même certains mots de contrôle sont formés de racines liées d'origine latine ou grecque (p. ex. *menopause*, *miniature*, *hypotenuse*, *voluptuous*). Dans la recherche, il n'est pas clairement indiqué sur quels critères et dans quelle mesure se sont basés les choix pour sélectionner les mots de contrôle, sauf le fait qu'ils ont été appariés selon la fréquence des suites orthographiques finales. Cependant, ces mots de contrôle auraient tous dû être monomorphémiques plutôt que formés de racines liées. À cause des limites énumérées, mais surtout compte tenu de la limite importante de la mauvaise catégorisation des items, les résultats de la recherche de Solomyak et Marantz (2010) n'aident pas à mieux comprendre le traitement des mots composés formés de racines liées, particulièrement les racines latines et grecques.

2.7.6 La synthèse des facteurs qui influencent l'identification des mots écrits polymorphémiques

À l'aide d'une combinaison de mesures comportementales et psychophysiques, les études recensées sur l'identification des mots composés permettent de constater que différents facteurs influencent leur traitement morphographique : la structure morphologique des items (Duñabeitia *et al.*, 2009; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014; Fiorentino et Poeppel, 2007; Stites *et al.*, 2016), la fréquence du mot entier (Goodwin *et al.*, 2013; Kuperman *et al.*, 2009), la fréquence des morphèmes (Duñabeitia *et al.*, 2009; Goodwin *et al.*, 2013; Kuperman *et al.*, 2009; Vergara-Martínez *et al.*, 2009), la transparence sémantique du mot entier (Brooks et Cid de Garcia, 2015) et de chacun des morphèmes (Davis *et al.*, 2019) ainsi que la position de la tête (Arcara *et al.*, 2013; El Yagoubi *et al.*, 2008). Cependant, aucun de ces effets n'a été vérifié sur des mots composés formés de racines liées. Étant donné le statut linguistique particulier de ces racines, il est probable que le traitement des mots composés formés de racines liées soit différent de ceux formés de racines libres.

Les résultats des études qui portent sur le traitement des mots dérivés formés de racines liées sont plutôt contradictoires, notamment parce que les listes de mots sont mal contrôlées. D'un côté, il est

annoncé que les mots formés de racines libres et de racines liées seraient traités de façon similaire (Forster et Azuma, 2000). De l'autre, on suggère un traitement différent pour ces deux types de mots (Coch *et al.*, 2013). Cependant, les résultats obtenus pourraient être attribués au procédé de formation des mots et non au type de racine. Du côté des études en neuroimagerie, davantage de recherches devraient d'être menées pour élucider la question relative au traitement des mots polymorphémiques formés de racines liées. En effet, il est suggéré que les mots formés de racines liées soient décomposés avant l'accès lexical, au même titre que les pseudomots formés de ces mêmes racines (McKinnon *et al.*, 2003). Il s'avère que cette décomposition ne serait pas marquée par le marqueur N400, qui serait davantage une marque de composition sémantique plus sensible à la lexicalité des items qu'au type de morphème présenté (Coch *et al.*, 2013). Des manques subsistent dans les études recensées. Effectivement, aucune étude n'a observé le traitement de racines liées au sein de mots composés.

De plus, les recherches recensées ont été menées dans des contextes linguistiques différents du français, soit le basque (Duñabeitia *et al.*, 2009; Duñabeitia *et al.*, 2007; Vergara-Martínez *et al.*, 2009), l'espagnol (Duñabeitia *et al.*, 2007), le néerlandais (Kuperman *et al.*, 2009), l'italien (Arcara *et al.*, 2013; El Yagoubi *et al.*, 2008) et, principalement, l'anglais (Brooks et Cid de Garcia, 2015; Coch *et al.*, 2013; Coch *et al.*, 2020; Davis *et al.*, 2019; Fiorentino *et al.*, 2014; Fiorentino et Poeppel, 2007; Forster et Azuma, 2000; Goodwin *et al.*, 2013; McKinnon *et al.*, 2003; Solomyak et Marantz, 2010; Stites *et al.*, 2016). D'ailleurs, les recherches sur les racines liées ont été menées exclusivement en contexte anglophone (Coch *et al.*, 2013; Coch *et al.*, 2020; Forster et Azuma, 2000; McKinnon *et al.*, 2003; Solomyak et Marantz, 2010). Aucune recherche recensée n'a été menée en contexte francophone.

Aussi, il est à noter qu'à l'exception d'une seule (Goodwin *et al.*, 2013), toutes les études énumérées ci-dessus ont été menées auprès de participants adultes, des lecteurs experts. On ne sait pas si les mêmes résultats seraient attendus avec des lecteurs en contexte d'apprentissage qui n'ont peut-être pas encore automatisé le traitement de tous les mots écrits. Plus particulièrement, à partir du 3^e cycle du primaire, les élèves rencontrent de plus en plus de nouveaux mots du vocabulaire disciplinaire formés de racines latines et grecques (Bar-Ilan et Berman, 2007; Green, 2008). Pallier ces manques permettrait de mieux comprendre le traitement des mots composés formés de racines latines et grecques, des mots particulièrement présents dans les manuels scolaires des élèves de 3^e cycle du primaire et à la base des savoirs disciplinaires. De plus, situer ces résultats par rapport à ceux obtenus auprès des adultes permettrait d'envisager une trajectoire développementale.

2.8 Les questions et les hypothèses de recherche

Les racines latines et grecques occupent une place importante dans la formation des mots en français, particulièrement ceux du vocabulaire disciplinaire. Or, on ne sait pas si les élèves du 3^e cycle du primaire connaissent ces racines et s'ils les utilisent pour identifier les mots écrits qui les contiennent. On ne sait pas non plus quels processus cognitifs sont mis en œuvre lors de l'identification des mots écrits polymorphémiques composés formés de racines liées. Enfin, pour bien comprendre le traitement de ces mots particuliers, il est d'intérêt d'observer comment les facteurs linguistiques et individuels interviennent dans la relation à l'étude. La présente recherche vise donc à répondre à deux questions de recherche. La première question de recherche est :

- 1) Lors de l'identification des mots écrits durant une tâche de décision lexicale, quel est l'effet sur la précision, le temps de réaction ainsi que l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus chez des élèves du 3^e cycle du primaire des facteurs linguistiques suivants :
 - a. la lexicalité (mots, pseudomots/non-mots);
 - b. la structure morphologique des items (polymorphémique, monomorphémique);
 - c. le type de racine (libre, liée);
 - d. l'origine des racines (latine, grecque); et
 - e. la fréquence (mot entier, première racine, deuxième racine)?

D'abord, un effet principal de lexicalité se traduirait par des temps de réponse plus courts (Coch *et al.*, 2013; Coch *et al.*, 2020; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014; Kuperman *et al.*, 2009) et des réponses plus précises (Coch *et al.*, 2020) pour les mots comparativement aux non-mots, c'est-à-dire des unités formés de suites de lettres qui ne forment pas des mots existants en français. Aussi, un effet de lexicalité se manifesterait très tôt dans le décours temporel, soit dès la fenêtre temporelle de 200 ms à 350 ms poststimulus associé à la décomposition morphologique, mais aussi dans la fenêtre temporelle de 350 ms à 500 ms poststimulus, associé à la récupération du sens du mot. Cet effet se traduirait par une négativité plus forte pour les non-mots que pour les mots dans ces deux fenêtres temporelles (El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014).

Ensuite, un effet de structure morphologique est attendu. Cependant, la direction de l'effet reste incertain pour le traitement des mots. Si on s'appuie sur les résultats d'El Yagoubi et ses collaborateurs

(2008), les participants répondraient plus rapidement et plus précisément aux items monomorphémiques qu'aux items polymorphémiques. Cependant, l'hypothèse inverse se dessine si on s'appuie sur les résultats de Fiorentino et ses collaborateurs (2007, 2014). Dans ce cas, les participants répondraient plus rapidement et plus précisément aux mots composés qu'aux mots monomorphémiques. Pour le traitement des pseudomots, soit les non-mots qui sont formés de morphèmes existants en français, ils entraîneraient un traitement plus long que les mots, qu'ils soient monomorphémiques ou polymorphémiques (Fiorentino *et al.*, 2014). Un effet de la structure morphologique des mots serait observé de 200 ms à 350 ms par une négativité plus forte pour les mots polymorphémiques comparativement aux mots monomorphémiques (El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014). Une interaction entre la lexicalité et la structure morphologique des items de 350 ms à 500 ms se traduirait par une négativité plus importante pour les pseudomots comparativement aux non-mots (Fiorentino *et al.*, 2014).

Selon les études antérieures (Coch *et al.*, 2013; Coch *et al.*, 2020), le traitement des items de la présente étude serait affecté par un effet du type de racine. Dans ces études, l'identification des mots dérivés formés de racines liées a été comparée à celle des mots composés formés de racines libres. Si l'effet du type de racine est indépendant du procédé de formation des mots, les participants répondront plus rapidement et plus précisément à la tâche de décision lexicale pour les items formés de racines liées que pour ceux formés de racines libres. Aussi, par manque de données empiriques, il est difficile d'émettre des hypothèses sur l'effet du type de racine sur l'amplitude et la latence de l'activité électrique dans les fenêtres temporelles identifiées (200 ms à 350 ms et 350 ms à 500 ms). Seule l'étude de Coch et ses collègues (2013) suggère que les items composés formés de racines libres entraîneraient une plus grande négativité vers 400 ms que les items de contrôle (mots monomorphémiques et non-mots), et ces derniers entraîneraient à leur tour une plus grande négativité que les items dérivés formés de racines liées. Comme le type de racine et le procédé de formation des mots sont indissociables dans cette étude, il est impossible d'émettre des hypothèses robustes. Donc, cet effet du type de racine ne pourra s'observer dans les résultats de la présente étude que s'il se distingue de celui du procédé de formation des mots.

Aussi, un effet dû à l'origine des racines dans les résultats comportementaux est attendu. En effet, comme les racines grecques sont généralement plus transparentes que les racines latines (Templeton, 1983) et que la transparence des morphèmes influence le traitement des mots polymorphémiques

composés (Davis *et al.*, 2019), il est attendu que les élèves répondent plus rapidement et précisément aux items formés de racines grecques que latines. Or, sur la base de résultats empiriques, il est impossible d'appuyer ces hypothèses ni de fournir des hypothèses éclairantes quant aux mesures psychophysiologiques.

De plus, sur la base des résultats des études antérieures (Duñabeitia *et al.*, 2007; Goodwin *et al.*, 2013; Kuperman *et al.*, 2009; Vergara-Martínez *et al.*, 2009), la fréquence du mot et la fréquence des racines contribueraient au temps de réaction des participants et à la précision des réponses à la tâche de décision lexicale lors de la lecture de mots composés. Plus le mot serait fréquent et plus ses racines seraient fréquentes, plus les réponses des élèves seraient rapides et exactes. Aussi, si le traitement de chacune des racines est associé à des processus cognitifs différents comme c'est le cas pour l'étude basque de Vergara-Martínez et ses collaborateurs (2009), la fréquence des racines aura des effets différents sur l'amplitude de chacune des fenêtres temporelles observées. Si la fréquence des racines influence de façon similaire l'amplitude dans chacune des fenêtres temporelles, on peut inférer que le traitement de la première et de la deuxième racine fait appel aux mêmes processus cognitifs.

Enfin, aucune hypothèse spécifique n'est formulée quant aux effets de chacune des variables indépendantes linguistiques sur la latence des potentiels évoqués. Cependant, un effet détecté sur la latence pourrait témoigner d'un temps de traitement différent pour chaque niveau d'une même variable. Cette interprétation serait alors aussi confirmée par un effet sur le temps de réaction des participants à la tâche de décision lexicale.

La deuxième question de recherche est :

2) Lors de l'identification des mots écrits durant une tâche de décision lexicale, quel est l'effet sur la précision, le temps de réaction ainsi que l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus chez des élèves du 3^e cycle du primaire des facteurs individuels suivants :

- a. la connaissance du vocabulaire;
- b. la reconnaissance des racines latines et grecques;
- c. les habiletés de décomposition morphologique;
- d. la connaissance du sens des racines latines et grecques; et
- e. les habiletés d'analyse morphologique?

À l'instar de Goodwin et ses collègues (2013), les habiletés morphologiques des élèves (connaissance des racines latines et grecques et analyse morphologique) et la connaissance du vocabulaire seraient susceptibles de favoriser le traitement des mots composés formés de racines latines et grecques. Les élèves avec de meilleures habiletés morphologiques identifieraient plus facilement et rapidement les items polymorphémiques de la tâche de décision lexicale. Aussi, comme le marqueur de potentiel évoqué N250 est attribuable à la décomposition morpho-orthographique des mots avant l'accès lexical (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008), la négativité dans la fenêtre temporelle de 200 ms à 350 ms poststimulus serait plus importante pour les élèves avec de meilleures habiletés morphologiques.

CHAPITRE 3

MÉTHODOLOGIE

Ce chapitre expose les différents choix méthodologiques qui émanent du cadre conceptuel. Pour commencer, le devis de recherche expérimental intra-sujets qui en découle est décrit (3.1). Par la suite, les participants de l'étude et les critères de sélection sont présentés (3.2). Ensuite, sont décrits les instruments de mesure (3.3) : la tâche de décision lexicale (3.3.1) ainsi que le test dynamique mesurant la connaissance du vocabulaire disciplinaire, la connaissance des racines latines et grecques et quelques habiletés morphologiques (3.3.2). Ce chapitre se poursuit avec la description de la procédure de collecte de données (3.4). Pour terminer, les étapes réalisées pour préparer les données électroencéphalographiques avant leur analyse sont expliquées (3.5), puis le plan d'analyse prévu pour répondre à chacune des questions de recherche est exposé (3.6).

3.1 Le devis de recherche

Cette recherche s'appuyait sur un devis expérimental intra-sujets (protocole factoriel) afin d'évaluer « simultanément l'influence de plusieurs interventions ou variables indépendantes d'une manière croisée » (Mercier *et al.*, 2005, p. 133). Dans le cadre de cette étude, différentes mesures permettaient d'observer la variable dépendante, le traitement morphographique. Le temps de réaction et la précision ainsi que la latence et l'amplitude aux fenêtres temporelles de 200 ms à 350 ms ainsi que de 350 ms à 500 ms poststimulus ont été mesurées. Des variables indépendantes de deux types ont été mesurées : linguistiques, soit propres aux items, et individuelles, c'est-à-dire propres aux participants. Les variables indépendantes linguistiques manipulées lors de la tâche de décision lexicale sont la lexicalité (mot ou pseudomot et non-mot), la structure morphologique de l'item (polymorphémique ou monomorphémique), le type de racine (libre ou liée) et l'origine des racines (latine ou grecque). La fréquence du mot ainsi que la fréquence des racines ont aussi été mesurées après la sélection des items de la tâche de décision lexicale. Les variables indépendantes individuelles sont la connaissance du vocabulaire disciplinaire, la connaissance des racines latines et grecques et les habiletés morphologiques des participants, plus particulièrement la reconnaissance des racines, la décomposition morphologique et l'habileté à utiliser l'analyse morphologique pour déduire le sens d'un mot inconnu.

3.2 Les participants

Cette section sert à présenter le profil des participants de la présente étude. D'abord, l'échantillon et la méthode d'échantillonnage retenue sont décrits (3.2.1). Puis, les critères d'inclusion et d'exclusion des participants dans l'échantillon ainsi que les mesures ayant permis de déterminer ces critères sont exposés (3.2.2).

3.2.1 La description de l'échantillon et de la méthode d'échantillonnage

L'échantillon de l'étude comptait 29 participants à développement typique (16 filles et 13 garçons)¹⁹. La collecte de données ayant eu lieu aux mois de mai à juillet 2022, 16 d'entre eux terminaient leur 5^e année du primaire tandis que les 13 autres terminaient leur 6^e année du primaire au moment de la collecte de données. Les participants étaient âgés de 10 ans et 10 mois à 12 ans et 9 mois ($M = 11$ ans et 7 mois; $ÉT = 7$ mois). Dix-huit participants habitaient la région administrative de Montréal, huit participants habitaient la Montérégie et un participant chacun habitait les régions administratives du Centre-du-Québec, de Laval et de Lanaudière.

L'échantillonnage non probabiliste a été adopté dans la réalisation de ce projet de recherche. Pour minimiser le biais dû à la sélection non aléatoire des participants, l'échantillon a été sélectionné de façon à être le plus fidèle possible à la population qu'il représente dans le respect des critères d'inclusion et d'exclusion présentés dans la prochaine section. Le recrutement des participants s'est fait par le biais des réseaux sociaux et par le bouche-à-oreille. Les parents qui désiraient participer au projet de recherche avec leur enfant ont pu prendre connaissance du projet et des critères d'admissibilité puis s'inscrire par

¹⁹ Étant donnée la taille de l'échantillon, le sexe n'a pas été considéré comme variable de contrôle bien que des résultats montrent des différences entre les sexes quant à différentes tâches en littératie, notamment aux épreuves obligatoires de français en lecture et en écriture en 6^e année du primaire (Desrosiers et Tétreault, 2012). Cependant, le recrutement des participants a été effectué en assurant une certaine représentativité des deux sexes. Selon Ullman (2007), les femmes auraient une plus grande facilité avec les tâches qui relèvent de la mémoire déclarative que procédurale, alors qu'on dénote l'inverse chez les hommes. Cette hypothèse pourrait s'expliquer par le fait que l'oestrogène, plus présent chez les femmes, favorise la performance de la mémoire déclarative (Ullman, 2007). Cette différence serait aussi observable dans l'analyse des potentiels évoqués où lors d'une tâche en morphologie flexionnelle l'activité électrique générée lors d'une violation morphologique ou syntaxique (mauvais morphème flexionnel) s'apparente à un N400 chez les femmes et à un marqueur ANG chez les hommes (Steinhauer et Ullman, 2002; Ullman *et al.*, 2002). Selon ces auteurs, le N400 s'observerait en région postérieure et refléterait l'accès lexical et la sollicitation de la mémoire déclarative. En contrepartie, le ANG en région antérieure gauche serait plutôt associé au traitement grammatical et aux tâches relevant de la mémoire procédurale (Ullman, 2001a; 2001b), comme la décomposition morphologique.

un formulaire en ligne. Les parents ont ensuite été contactés par téléphone pour vérifier à nouveau les critères d'admissibilité du participant et pour prendre rendez-vous pour la collecte de données.

Conformément aux exigences relatives aux recherches auprès d'êtres humains, un certificat éthique a été émis par le comité institutionnel d'éthique de la recherche avec des êtres humains de l'Université du Québec à Montréal (#2022-4717; Annexe A). De plus, le consentement libre et éclairé, non seulement des participants mineurs, mais aussi de leurs parents, a été obtenu. Pour mieux renseigner les parents et les participants, un lien vers une vidéo d'un enfant réalisant une activité à NeuroLab a été envoyé dès le premier contact téléphonique avec les parents. Enfin, une compensation financière de 50\$ a été remise à chaque famille pour leur participation.

3.2.2 Les critères et les mesures d'inclusion et d'exclusion

Pour être inclus dans l'échantillon, les participants ne devaient correspondre à aucun de ces profils : dyslexie, dysorthographe, dyscalculie, trouble du déficit de l'attention avec ou sans hyperactivité (TDA/H), trouble développemental du langage, trouble du spectre de l'autisme, déficience intellectuelle, trouble moteur ou trouble neurologique (p. ex. épilepsie). Tout autre diagnostic devait être déclaré. Les participants ne devaient pas non plus avoir redoublé une année scolaire. Tous les participants avaient une vision normale ou corrigée à la normale à l'aide de lunettes ou de lentilles cornéennes. Aussi, pour des raisons techniques liées à la tâche de décision lexicale (voir section 3.3.1.2), tous les participants devaient être droitiers.

De plus, le statut socioéconomique des participants a été contrôlé. Effectivement, cette variable est liée à leurs habiletés scolaires, notamment en lecture (Fluss *et al.*, 2009; Sirin, 2005). Entre autres, sans constater de différences cognitives non verbales, on observe des différences entre les élèves de milieux défavorisés et moyens quant à diverses composantes de la littératie, dont le vocabulaire (Bara *et al.*, 2008; Gentaz *et al.*, 2015; Gentaz *et al.*, 2013). De ce fait, afin d'avoir une meilleure homogénéité de l'échantillon, un statut socioéconomique moyen a été exigé. Des recherches ont identifié le niveau d'éducation de la mère comme le principal indicateur du statut socioéconomique de la famille (p. ex. Fluss *et al.*, 2009). Étant donné la variabilité possible des structures familiales, le niveau d'éducation le plus élevé des deux parents a été recueilli. Les données issues du recensement de 2016 de Statistique Canada indiquent que 48,3 % de la population québécoise de 25 à 64 ans a niveau d'éducation équivalent ou supérieur au collégial (Statistique Canada, 2017). Ainsi, au moins un des deux parents de la

structure familiale devait avoir obtenu un diplôme d'études collégiales ou l'équivalent pour que son enfant soit admissible au projet de recherche. Le Tableau 3.1 expose la distribution du niveau de scolarité le plus élevé obtenu par un des parents des participants.

Tableau 3.1 Niveau de scolarité le plus élevé obtenu par un des parents des participants

Niveau de scolarité	<i>n</i>
Diplôme d'études collégiales (ou équivalent)	4
Certificat (1 ^{er} cycle universitaire)	1
Baccalauréat (1 ^{er} cycle universitaire)	12
Maitrise (2 ^e cycle universitaire)	8
Doctorat (3 ^e cycle universitaire)	4
Total	29

Aussi, les participants qui parlaient une autre langue que le français avec les membres de leur famille ou avec leurs amis ou qui n'ont pas été scolarisés en milieu francophone depuis le début de leur scolarité ont été exclus de l'étude. En effet, étant donné les racines communes entre le français et d'autres langues latines et germaniques, la connaissance des racines latines et grecques pourrait être influencée par la connaissance d'une autre langue. Pour contrôler cette variable, le parent du participant mineur devait répondre à ces différentes questions dans un questionnaire autorapporté : « Quelle(s) langue(s) parle l'enfant avec le parent 1? », « Quelle(s) langue(s) parle l'enfant avec le parent 2? », « Quelle(s) langue(s) parle l'enfant avec ses frères et sœurs ou ses amis à la maison? », « Quelle(s) langue(s) parlent les parents entre eux? ». Sur une échelle de *Toujours* à *Jamais* les parents devaient aussi répondre aux questions suivantes : « À quelle fréquence votre enfant utilise-t-il le français à la maison? », « Si votre enfant parle une autre langue que le français, à quelle fréquence utilise-t-il cette langue ou ces autres langues à la maison? ». Il est toutefois à noter que l'anglais comme langue seconde est enseigné au Québec dès la 1^{re} année du primaire. Les participants sélectionnés n'étaient donc pas entièrement monolingues, mais aucun n'était fluide dans une autre langue que le français. Ainsi, après une discussion téléphonique avec la chercheuse principale, les participants étaient inclus dans l'échantillon même s'ils parlaient rarement anglais à la maison, plutôt que jamais, et que leur compétence ne s'était développée que durant les cours d'anglais obligatoires à l'école primaire selon les propos rapportés par les parents.

Comme la présente étude porte sur des élèves à développement typique, la correspondance de l'échantillon à la population cible a été assurée quant aux habiletés d'identification des mots écrits et à

l'intelligence non verbale des participants. Ainsi, les résultats des participants devaient correspondre à la norme pour leur groupe d'âge. Comme dans certaines recherches utilisant une procédure similaire (p. ex. Coch *et al.*, 2020), deux prétests (identification des mots écrits et intelligence non verbale) ont été passés aux participants afin d'exclure ceux qui auraient des résultats déviants par rapport à la moyenne fournie dans les normes des tests. L'échec à un des deux prétests entraînait une exclusion de l'échantillon.

D'abord, le premier prétest utilisé était le test d'identification des mots écrits qui est une tâche standardisée provenant de la batterie d'évaluation du langage écrit (BELEC) de Mousty et ses collaborateurs (1994). Il s'agit d'un test individuel d'au maximum dix minutes pour mesurer la vitesse et la précision lors de la lecture de 72 mots et non-mots courts et longs. Les élèves étaient conservés dans l'échantillon si leurs résultats en vitesse et en précision, autant pour les mots que les non-mots, ne divergeaient pas de plus de deux écarts types par rapport à la moyenne standardisée disponible pour les élèves de 4^e année du primaire, normes standardisées les plus près de l'âge des participants (Mousty et Leybaert, 1999).

Puis, les matrices progressives standards de Raven (1983) constituaient le deuxième prétest. Elles ont permis de mesurer l'intelligence non verbale des participants afin d'assurer que le raisonnement des participants correspondait à la norme selon leur groupe d'âge. Ce test contient cinq séries de douze images présentées en ordre croissant de difficulté. L'élève devait sélectionner le bon morceau parmi six à huit pièces pour reconstituer correctement l'image. Il s'agit d'un test individuel et qui se réalise en moins de dix minutes. La consistance interne (α) de cet instrument de mesure est de 0,88 (Pearson, 2007). Les participants devaient obtenir un résultat leur permettant d'être classés au-delà du 25^e percentile selon les normes estimées pour leur groupe d'âge (Orme, 1966), ce qui correspond au seuil de l'intelligence moyenne pour ce test (Vanier, 1991).

Après que les parents aient témoigné l'intérêt de leur enfant à participer à la recherche, neuf élèves ont été refusés parce qu'ils ne répondaient pas à l'un ou l'autre des critères d'inclusion : trois élèves étaient bilingues, deux élèves présentaient des troubles d'apprentissage, un élève était gaucher, le niveau de scolarité des parents d'un élève n'était pas suffisamment élevé, un élève a échoué le prétest d'identification des mots écrits et, enfin, un élève a échoué le prétest d'intelligence non verbale.

3.3 Les instruments de mesure

D'abord, la variable dépendante, soit le traitement morphographique, a été mesurée lors d'une tâche de décision lexicale (3.3.1). À cet effet, des données comportementales (temps de réaction et précision) et psychophysiques (amplitude et latence de potentiels évoqués) ont été recueillies lors de la tâche de décision lexicale durant laquelle les variables indépendantes linguistiques suivantes ont été manipulées : la lexicalité (mot ou pseudomot et non-mot), la structure morphologique de l'item (polymorphémique ou monomorphémique), le type de racine (libre ou liée) et l'origine des racines (latine ou grecque). Puis, un test dynamique (3.3.2) a permis de mesurer les variables indépendantes individuelles : la connaissance du vocabulaire disciplinaire, la connaissance des racines latines et grecques, la reconnaissance des racines, les habiletés de décomposition morphologique et les habiletés d'analyse morphologique des participants. Cette section décrit chacune des mesures de la recherche et les choix méthodologiques qui en découlent.

3.3.1 La tâche de décision lexicale : le traitement morphographique

La tâche de décision lexicale permet de recueillir des données nécessaires pour inférer sur le traitement morphographique lors de l'identification des mots écrits. Les prochaines sections décrivent les stimulus choisis pour cette tâche de décision lexicale (3.3.1.1) ainsi que le déroulement de la tâche (3.3.1.2).

Afin de valider la compréhension des consignes ainsi que le temps de passation, un pilote du test a été passé à deux élèves du 3^e cycle du primaire qui répondent aux mêmes critères d'inclusion que les participants de l'échantillon. Ces participants étaient une fille en 5^e année du primaire (11 ans et 0 mois) et un garçon en 6^e année du primaire (11 ans et 10 mois). Aucun enregistrement des réponses ni enregistrement électroencéphalographique n'a été effectué. À la suite de ce petit pilote, aucune modification n'a été apportée à la tâche de décision lexicale puisque celui-ci était satisfaisant : les participants semblaient avoir bien compris les consignes.

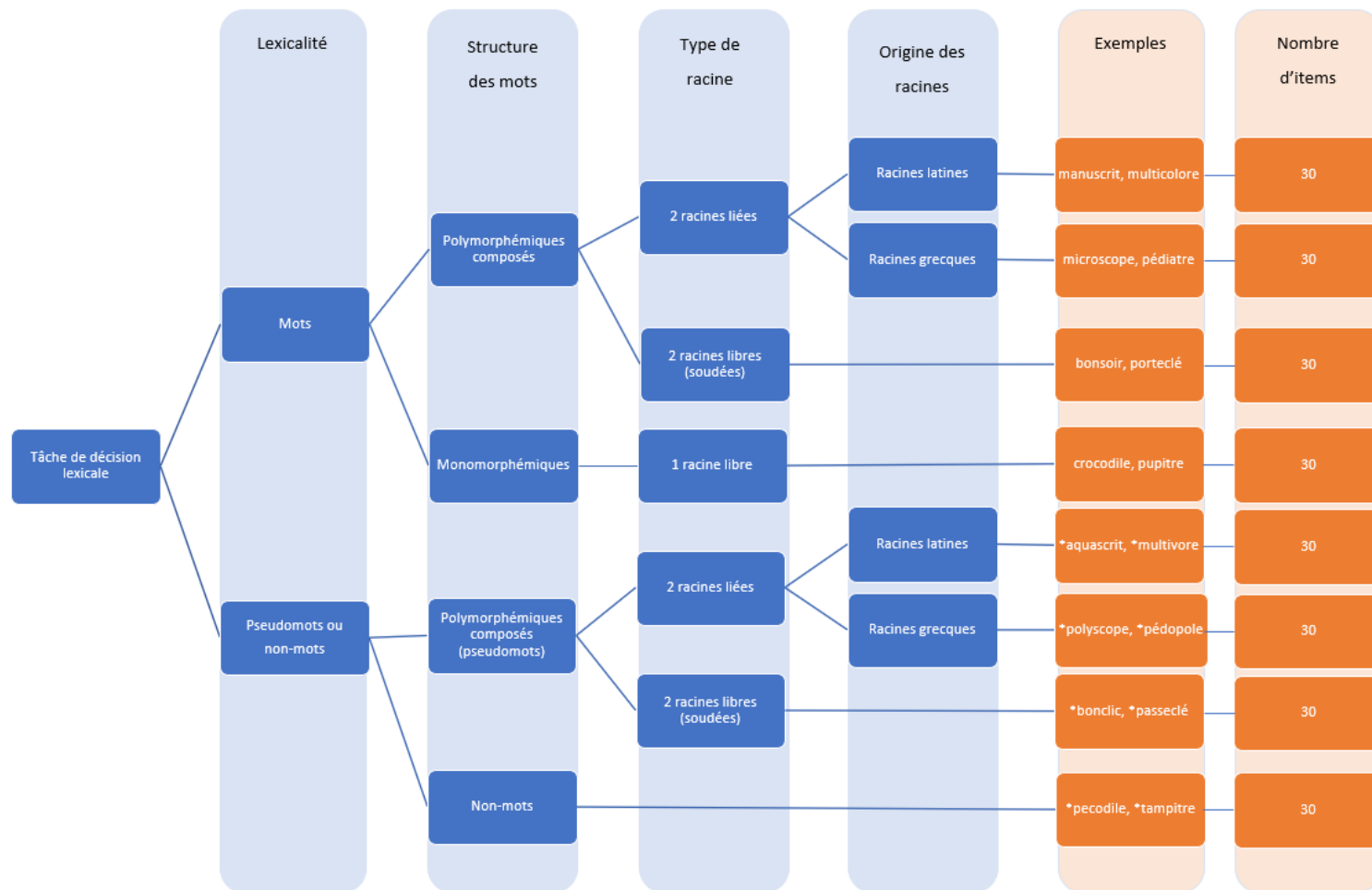
3.3.1.1 Les stimulus de la tâche de décision lexicale

La tâche de décision lexicale comprenait 120 mots et 120 non-mots présentés aléatoirement au participant en quatre blocs de 60 items (voir Annexe B). L'ordre de présentation des items était différent pour chaque participant. La présentation aléatoire des stimulus réduisait l'effet d'habituation des participants à la tâche, ce qui aurait entraîné des temps de réponse de plus en plus courts pour les

mêmes items d'un participant à l'autre à mesure qu'ils auraient progressé dans la tâche (Kuperman *et al.*, 2009).

La sélection des stimulus pour la tâche de décision lexicale a été faite dans le respect de variables susceptibles d'intervenir dans le traitement lexical. La lexicalité (intrinsèque à une tâche de décision lexicale), la structure du mot (polymorphémique ou monomorphémique), le type de racine (libre ou liée) et l'origine des racines (latine ou grecque) étaient les variables appariées dans le plan factoriel de la tâche de décision lexicale (Figure 3.1). Sur la base de ces variables, les 240 items étaient répartis en huit conditions expérimentales de trente stimulus chacun : 1) mots polymorphémiques composés formés de deux racines latines liées (p. ex. *manuscrit*) ; 2) mots polymorphémiques composés formés de deux racines grecques liées (p. ex. *microscope*) ; 3) mots polymorphémiques composés formés de deux racines libres (p. ex. *bonsoir*) ; 4) mots monomorphémiques (p. ex. *crocodile*) ; 5) pseudomots polymorphémiques composés formés de deux racines latines liées (p. ex. **aquascrit*) ; 6) pseudomots polymorphémiques composés formés de deux racines grecques liées (p. ex. **polyscope*) ; 7) pseudomots polymorphémiques composés formés de deux racines libres (p. ex. **bonclit*) ; 8) non-mots (p. ex. **pecodile*).

Figure 3.1 Plan factoriel des stimulus pour la tâche de décision lexicale



Pour sélectionner les mots polymorphémiques composés formés de racines liées, l'ouvrage de Cellard (2013), qui répertorie les principales racines latines et grecques dans le vocabulaire français, a été consulté. Le dictionnaire d'étymologie du français (Picoche, 2015) et le Trésor de la Langue française informatisé (Analyse et traitement informatique de la langue française [ATILF], s.d.) ont aussi été des ressources incontestables. Afin de veiller à une grande représentativité des racines latines et grecques, mais surtout d'éviter la surexposition à certaines racines, aucune racine n'a été répétée d'un mot à l'autre. Effectivement, la répétition d'un stimulus peut entraîner une réduction de la négativité au N400 (Kutas et Federmeier, 2011). Par exemple, comme le stimulus *omnivore* a été sélectionné, les stimulus *apivore* ou *piscivore* ne pouvaient pas l'être. De ce fait, 60 racines latines et 60 racines grecques sont présentées au sein des 60 mots polymorphémiques composés formés de racines latines ou grecques.

Afin de respecter la même forme orthographique que les mots polymorphémiques composés formés de racines liées, les mots polymorphémiques composés formés de deux racines libres étaient soudés, c'est-à-dire qu'ils ne pouvaient pas être séparés d'un trait d'union (p. ex. *cerf-volant*) ou d'un espace (p. ex. *papier journal*). D'ailleurs, certains de ces stimulus sont des mots qui s'écrivent avec un trait d'union en graphie traditionnelle (p. ex. *bien-être, essuie-tout*), mais sans trait d'union en graphie rectifiée (p. ex. *bienêtre, essuietout*). Les graphies rectifiées ont été privilégiées, d'autant plus que depuis la rectification de l'orthographe (Conseil supérieur de la langue française, 1990) les élèves ont de plus en plus d'occasions de rencontrer ces graphies dans leurs manuels scolaires. Ainsi, une éventuelle différence dans le traitement des mots polymorphémiques composés formés de racines libres et liées ne pourra pas être associée aux différences typographiques.

Aussi, aucun mot n'était formé d'affixes dérivationnels ou flexionnels afin de respecter le procédé de formation des mots sélectionnés, la composition. Par exemple, le stimulus *décennie*, bien que formé de deux racines latines, n'a pas été sélectionné comme il contient aussi le suffixe *-ie*. De plus, sans nécessairement y associer d'indice, les mots polymorphémiques composés sélectionnés étaient relativement transparents sémantiquement puisque cette variable est susceptible d'intervenir dans la relation à l'étude (Brooks et Cid de Garcia, 2015; Davis *et al.*, 2019). Le sens de chacune des racines devait être lié, au moins partiellement, au sens du mot entier. Par exemple, le mot *croquemonsieur* n'a pas été sélectionné parmi les mots polymorphémiques composés formés de deux racines libres puisque sa relation sémantique avec ses racines libres est opaque.

Les mots monomorphémiques ont été sélectionnés avec une attention particulière à leur nombre de lettres afin de s'approcher le plus possible de celui des autres conditions. Afin de faciliter la formation de non-mots (voir plus loin), les mots monomorphémiques devaient être composés d'au moins deux syllabes orales. Par exemple, les mots comme *rouille* et *nouille* n'ont pas pu être sélectionnés. Aussi, les syllabes des mots monomorphémiques ne devaient pas correspondre à la forme orthographique d'un autre mot monomorphémique dans la langue, appelé *mot incorporé* (p. ex. *terne* dans *lanterne* ou *mer* et *veille* dans *merveille*). Effectivement, la présence de mots incorporés, que ce soit dans des mots ou des non-mots, influence le traitement des mots écrits, notamment le temps de réaction lors d'une tâche de décision lexicale (Beyersmann *et al.*, 2018). Enfin, il s'agissait d'une tâche de décision lexicale où les participants avaient à fournir leur jugement de décision lexicale sur les items présentés. Dans des études antérieures, des auteurs ont émis l'hypothèse qu'un plus grand nombre de mauvaises réponses était fourni par les enfants que les adultes comme ceux-ci ont un vocabulaire moins développé (Henderson et Chard, 1980). Comme les mauvaises réponses n'étaient pas analysées et pour avoir suffisamment de puissance statistique, les mots sélectionnés devaient être susceptibles d'avoir déjà été rencontrés par des élèves de 3^e cycle du primaire. Par exemple, même si le mot *macrocéphale* répond à tous les critères énumérés, il n'a pas été sélectionné puisqu'il est improbable qu'un élève de 5^e ou de 6^e année du primaire l'ait déjà rencontré dans ses lectures.

Aux mots étaient appariés des pseudomots et des non-mots aussi contrôlés pour leur structure, leur type de racine ainsi que leur origine. La formation des pseudomots polymorphémiques composés s'est faite en jumelant aléatoirement deux racines des mots au sein de chaque catégorie (p. ex. **aquascrit*). La position (initiale et finale) de chaque racine demeurait inchangée. Ainsi, si la racine *multi* était présentée en position initiale dans un mot (p. ex. *multicolore*), elle l'était aussi dans son pseudomot homologue (p. ex. **multivore*). Pour les non-mots de contrôle, soit ceux appariés aux mots monomorphémiques, ceux-ci ont été formés en interchangeant aléatoirement les syllabes des mots monomorphémiques (p. ex. **tampitre* formé par *tambour* et *pupitre*). Ainsi, on s'assurait de former des mots constitués de syllabes existantes en français. Quelques modifications ont été apportées manuellement à la suite de ce jumelage automatique de façon que les pseudomots et les non-mots présentés soient orthographiquement plausibles (p. ex. dans les non-mots, *pin* ne pouvait pas précéder *pitre* puisque la voyelle nasale [ɛ̃] s'écrivait avec un *m* au lieu d'un *n* devant un *p*). Enfin, aucun pseudomot ou non-mot ne comportait de *pseudoaffixes*, c'est-à-dire une suite orthographique en début ou en fin de mot

rappelant un préfixe ou un suffixe existants en français (p. ex. *re* et *ette* dans *recette*) afin d'éviter de déclencher un traitement morphographique dérivationnel, même erroné.

Étant donné toutes les conditions de sélection des stimulus décrites précédemment, le choix des items était restreint. Il a donc été impossible de contrôler pour d'autres propriétés psycholinguistiques des mots en amont, comme pour la fréquence des mots ainsi que la fréquence des racines. Ces indices de fréquence ont été recueillis après la sélection des stimulus. Pour la fréquence du mot entier, une recension des principales listes de fréquences en français (Annexe C) a permis de constater que peu de listes sont adaptées à la population ciblée. Effectivement, la plupart des listes recensées ont été bâties à l'aide de corpus destinés aux adultes (p. ex. Drouin, 2007; New *et al.*, 2001; Tutin, 2008). Parmi les listes élaborées à partir d'ouvrages destinés aux élèves du primaire (Lété *et al.*, 2004; Ministère de l'Éducation du Loisir et du Sport, 2014; Stanké *et al.*, 2019; Tremblay *et al.*, 2023; Tremblay *et al.*, 2018), le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018) répondait aux besoins de la présente étude puisqu'elle recense les mots disciplinaires et transdisciplinaires de manuels scolaires québécois aux trois cycles du primaire. C'est donc l'indice de fréquence correspondant à l'occurrence du mot à travers tous les manuels recensés qui a été utilisé pour les mots de la tâche de décision lexicale. La fréquence totale des trois cycles a été sélectionnée plutôt que celle au 3^e cycle uniquement puisque les élèves du 3^e cycle du primaire ont déjà été en contact avec les mots présentés dans les manuels des cycles précédents.

Cependant, le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018) ne répertorie pas la fréquence des racines au sein du corpus. À notre connaissance, la seule base de données qui répertorie certaines données sur quelques racines latines et grecques est Polymots (LIF-TALEP, 2012). Cette base, qui recense principalement des mots dérivés, compte 8 000 mots regroupés par racine en 600 familles morphologiques (Gala et Rey, 2008). La productivité de chaque racine au sein de la base de données est fournie, soit le pourcentage de mots de la base de données qui contiennent cette racine. Il ne s'agit donc pas d'un indice de fréquence d'apparition de la racine. La productivité s'apparenterait davantage à un indice de taille de la famille morphologique au sein de cette base de données. Aussi, comme les auteurs ont fait le choix de ne conserver que des mots communs, les mots du vocabulaire disciplinaire n'ont pas été retenus (Gala et Rey, 2008). De ce fait, plusieurs racines latines et grecques à la base de la formation de mots du vocabulaire disciplinaire ne sont pas répertoriées (p. ex. *hydr*, *log*, *géo*, *api*). En s'appuyant sur la base de données du Lexique Scientifique Transdisciplinaire

Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018), la fréquence des racines a été calculée manuellement. Pour ce faire, pour chaque racine, tous les mots de la base de données formés de ladite racine ainsi que leur fréquence respective ont été répertoriés. La fréquence de tous ces mots a ensuite été cumulée pour obtenir la fréquence de la racine. Les indices de fréquence pour chacun des items sont disponibles à l'Annexe D. Le tableau 3.2 reprend les fréquences moyennes et les écarts types des items et de leurs racines pour chacune des catégories d'items de la tâche de décision lexicale. Dans le cadre de cette étude, la fréquence brute des items a été rapportée, c'est-à-dire le nombre d'occurrences de l'item dans le corpus de la base de données. Comme tous les indices de fréquence sont récupérés de la même base de données, l'utilisation de la fréquence brute n'est pas un enjeu. Cependant, comme cette fréquence brute dépend du nombre de mots dans le corpus de la base de données, une éventuelle comparaison avec des indices de fréquence provenant de bases de données différentes est impossible. À ce moment, la standardisation de l'indice sous un même dénominateur serait nécessaire. Dans la plupart des bases de données lexicales, les indices de fréquence sont rapportés sur 1 million d'occurrences (p. ex. Manulex, Lété *et al.*, 2004; Lexique 3, New *et al.*, 2001).

Tableau 3.2 Moyennes et écarts types des fréquences des items et de leurs racines pour chacune des catégories d'items de la tâche de décision lexicale

	Fréquence du stimulus entier		Fréquence de la 1 ^{re} racine		Fréquence de la 2 ^e racine	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
(1) Mots polymorphémiques formés de deux racines liées latines	16,00	30,92	748,24	1355,48	1296,27	2253,04
(2) Mots polymorphémiques formés de deux racines liées grecques	18,20	37,08	258,53	418,54	783,93	2552,77
(3) Mots polymorphémiques formés de deux racines libres	32,27	69,47	2735,97	2905,56	2221,23	2927,18
(4) Mots monomorphémiques	45,27	72,00	—	—	—	—
(5) Pseudomots polymorphémiques formés de deux racines latines	—	—	748,24	1355,48	1296,27	2253,04
(6) Pseudomots polymorphémiques formés de deux racines grecques	—	—	258,53	418,54	783,93	2552,77
(7) Pseudomots polymorphémiques formés de deux racines libres	—	—	2735,97	2905,56	2221,23	2927,18
(8) Non-mots	—	—	—	—	—	—

Note. — = non disponible.

Des ANOVAS à un facteur ont permis de mesurer la différence des moyennes entre les catégories pour la fréquence du stimulus entier, la fréquence de la première racine et la fréquence de la deuxième racine.

Pour les comparaisons multiples, la correction Bonferroni est appliquée. Si le postulat d'homogénéité des variances n'est pas respecté, les résultats de l'ANOVA de Welch est présenté et les comparaisons multiples rapportées sont corrigées en fonction de la correction Games-Howell. Étant donné la grande quantité de comparaisons multiples, seules les comparaisons significatives sont rapportées afin d'éviter d'alourdir le texte.

Pour la fréquence du stimulus entier, l'ANOVA ne permet pas de conclure de différences entre les quatre catégories de mots, $F(3, 61,43) = 1,68$, $p = 0,180$, $\eta^2 = 0,05$. Cependant, les écarts types élevés témoignent d'une grande variabilité quant à la fréquence des stimulus entiers dans chacune des catégories. Pour la fréquence de la première racine, des différences entre les catégories sont significatives, $F(5, 76,84) = 9,41$, $p < 0,001$, $\eta^2 = 0,26$. Selon l'analyse des comparaisons multiples, les premières racines des mots et pseudomots polymorphémiques formés de deux racines libres sont plus fréquentes que celles des mots et pseudomots polymorphémiques formés de racines latines ($p = 0,018$, $d = 0,51$), ainsi que de celles des mots et pseudomots polymorphémiques formés de racines grecques ($p < 0,001$, $d = 0,69$). Pour la fréquence de la deuxième racine, les résultats ne montrent aucune différence significative entre les catégories, $F(5, 174) = 1,90$, $p = 0,097$, $\eta^2 = 0,05$.

D'autres propriétés psycholinguistiques des stimulus sont rapportées dans l'annexe D : le nombre de lettres; le nombre de syllabes orales; le nombre de graphèmes, indice récupéré de Manulex-infra (Peereman et al., 2007) ou calculé manuellement lorsque le stimulus n'était pas dans la base de données; la taille de la famille morphologique du stimulus entier ainsi que de chacune des racines, indice calculé manuellement en comptant le nombre de mots partageant la même racine à partir du Lexique Scientifique Transdisciplinaire Junior (Tremblay et al., 2023; Tremblay et al., 2018); ainsi que le cycle scolaire (1^{er} cycle, 2^e cycle ou 3^e cycle) d'apparition du stimulus entier et de chacune des racines dans les manuels scolaires du primaire, indice extrait à partir des données brutes inédites du Lexique Scientifique Transdisciplinaire Junior (Tremblay et al., 2023; Tremblay et al., 2018). Des analyses de comparaison des moyennes entre chacune des catégories ont été effectuées pour toutes ces variables. Par souci d'espace et de pertinence, seuls les comparaisons significatives sont rapportées à l'Appendice B.

3.3.1.2 Le déroulement de la tâche de décision lexicale

La tâche de décision lexicale s'est déroulée en deux temps. D'abord, une fois le casque d'électroencéphalographie correctement installé sur la tête de l'élève (voir section 3.4), une base de

référence a été enregistrée afin de recueillir les principaux artéfacts susceptibles d'être produits durant la tâche de décision lexicale et ainsi faciliter leur retrait lors du nettoyage de données. Il a été demandé aux élèves de cligner des yeux, de regarder à gauche puis à droite, de regarder vers le haut puis vers le bas, de serrer la mâchoire et de bâiller. Puis, les consignes spécifiques relatives à la tâche de décision lexicale étaient lues aux participants avant de démarrer la tâche. Le protocole de passation détaillé de la tâche de décision lexicale, comprenant les consignes à lire aux participants, est disponible à l'annexe E. Cependant, les principaux éléments sont présentés ci-dessous.

Avant la tâche de décision lexicale, un entraînement de huit mots et de huit pseudomots et non-mots (deux items par condition) a été présenté aux participants (Duñabeitia *et al.*, 2009). Durant cet entraînement, des rétroactions étaient fournies afin d'assurer que la tâche et les consignes étaient bien comprises. Pour chaque item de la tâche de décision lexicale, les participants devaient lire silencieusement le stimulus et décider si la suite de lettres présentée était un mot qui existe en français. Les participants avaient comme consigne de répondre le plus rapidement possible. Les participants étaient aussi informés qu'il était possible qu'un non-mot soit formé de deux mots existants en français (Davis *et al.*, 2019). Les participants devaient appuyer sur la touche *Ctrl gauche* du clavier s'il s'agissait d'un mot et *Ctrl droit* s'il s'agissait d'un non-mot. Pour aider les participants, un cercle vert était collé sur la touche *Ctrl gauche* et un cercle rouge était collé sur la touche *Ctrl droit*. Comme la main dominante des participants pouvait influencer à la baisse les temps de réponse, deux solutions étaient possibles pour contourner cette variable : la contrebalancer entre les participants ou la contrôler. Certains chercheurs (p. ex. Coch *et al.*, 2013; McKinnon *et al.*, 2003) ont effectivement fait le choix de contrebalancer cette variable de sorte que pour la moitié des participants la réponse comportementale au mot devait être fournie par la main droite et, pour l'autre moitié, par la main gauche. Dans le cadre de cette étude, pour éviter toute confusion lors de la collecte de données et lors de l'interprétation des résultats, cette variable a été contrôlée en ne sélectionnant que des élèves droitiers. Ainsi, la main dominante a été associée aux items qui entraînent habituellement des temps de réponse plus lents, soit les non-mots (Coch *et al.*, 2013; Coch *et al.*, 2020; Fiorentino *et al.*, 2014; Fiorentino et Poeppel, 2007).

Afin de concevoir la tâche de décision lexicale, une comparaison des modalités temporelles de différentes tâches de décision lexicale a été effectuée. À cet égard, on peut remarquer dans le Tableau 3.3, que les tâches comprennent généralement les mêmes étapes d'un auteur à l'autre. Une tâche de décision lexicale commence généralement par la présentation d'une croix au centre de l'écran pour

positionner de façon optimale le regard du participant en vue de la présentation du stimulus et éviter un balayage visuel préalable à l'identification du mot, ce qui pourrait bruite les données psychophysiques. Puis, le stimulus est présenté. Dans la plupart des cas (Arcara *et al.*, 2013; Coch *et al.*, 2013; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014), ce dernier est présenté durant 2 500 ms à 3 000 ms, ou jusqu'à la réponse de jugement lexical qui représente la troisième étape. Enfin, dans certaines tâches, un court intervalle entre les stimulus est prévu.

Tableau 3.3 Modalités temporelles de tâches de décision lexicale dans des recherches mesurant le traitement morphographique

Auteurs	Étape 1 Croix au centre	Étape 2 Présentation du stimulus	Étape 3 Jugement lexical	Étape 4 Intervalle entre les stimulus
McKinnon <i>et al.</i> (2003)	500 ms	700 ms	Écran vierge pour 1 450 ms suivi d'une incitation au jugement lexical par pression d'un bouton	Aucun
El Yagoubi <i>et al.</i> (2008)	500 ms	Max. 3 000 ms ou jusqu'à réponse	Pression d'un bouton	#### pour 2 000 ms
Arcara <i>et al.</i> (2013)	1 500 ms	Max. 2 500 ms ou jusqu'à réponse	Pression d'un bouton	Écran vierge pour 2 000 ms
Coch <i>et al.</i> (2013)	Jusqu'à pression d'un bouton	Max. 3 000 ms ou jusqu'à réponse	Pression d'un bouton	Aucun
Fiorentino <i>et al.</i> (2014)	750 ms	Max. 3 000 ms ou jusqu'à réponse	Pression d'un bouton	Aucun
Davis <i>et al.</i> (2019)	Aucun	250 ms	Pression d'un bouton	Pseudorandomisation de 400 ms à 1 600 ms

Aussi, les modalités visuelles utilisées lors de ces tâches de décision lexicale pour la présentation des stimulus ont été comparées (voir Tableau 3.4). Au sein des tâches observées, tous les chercheurs disposent leurs items centrés à l'écran. Aussi, la plupart utilisent un fond noir et une écriture blanche pour la présentation des stimulus en conservant toujours une police d'écriture lisible. Enfin,

lorsqu'indiquées, les informations relatives à la taille d'écriture sont relativement variées d'une tâche à l'autre.

Tableau 3.4 Modalités visuelles utilisées lors de la présentation des stimulus dans les tâches de décision lexicale de recherches mesurant le traitement morphographique

Auteurs	Logiciel	Couleur de l'arrière-plan	Couleur du stimulus	Police	Taille / caractères	Disposition
Forster et Azuma (2000)	DMDX	Blanc	Noir	Courrier New	Taille 10 / amorce en minuscules, cible en majuscules	Centré
McKinnon <i>et al.</i> (2003)	Non indiqué	Non indiqué	Non indiqué	Non indiqué	Non indiqué / non indiqué	Centré
El Yagoubi <i>et al.</i> (2008)	E-Prime 1.1	Argent	Noir	Courrier New	Taille 32 / non indiqué	Centré
Arcara <i>et al.</i> (2013)	Non indiqué	Noir	Blanc	Courrier	Taille 18 / majuscules	Centré
Coch <i>et al.</i> (2013)	Presentation software (Neuro-behavioral Systems)	Noir	Blanc	Arial	Taille 36 / non indiqué	Centré
Fiorentino <i>et al.</i> (2014)	Paradigm (Tagliaferri, 2005)	Noir	Blanc	Courrier New	Non indiqué / non indiqué	Centré
Davis <i>et al.</i> (2019)	E-Prime 1.2	Noir	Blanc	Courrier	Non indiqué / minuscules	Centré

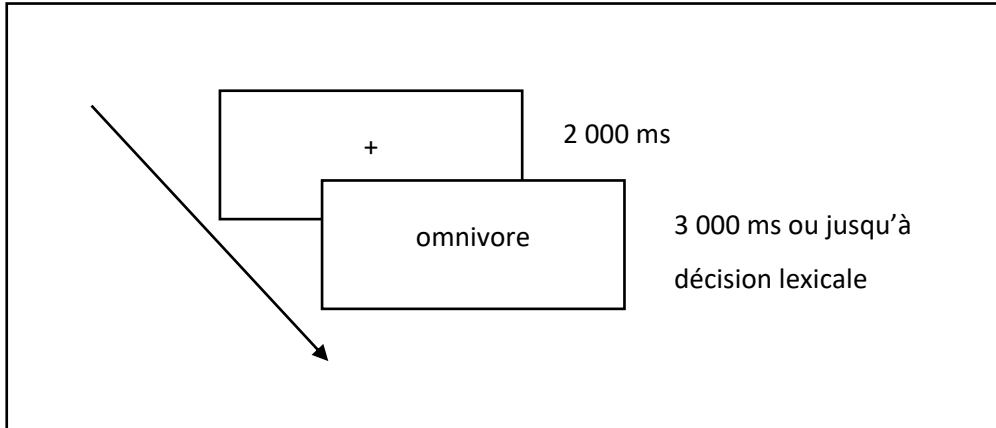
La recension des études antérieures a permis de prendre des décisions méthodologiques quant à la tâche de décision lexicale pour la présente étude. Bien que plusieurs utilisent le logiciel E-Prime, le logiciel DMDX (version 6.1.1.1; Forster et Forster, 2003) sera utilisé lors de la tâche de décision lexicale. Il s'agit d'un logiciel plus précis pour calculer le temps de réaction durant l'identification des mots écrits. Alors que la plupart des recherches utilisent un arrière-plan foncé et une écriture pâle, un choix différent a été effectué pour plus de validité écologique. Tous les items ont été présentés en minuscules de

couleur noire et étaient disposés au centre de l'écran sur un fond blanc. Cette présentation des stimulus rejoint davantage l'expérience qu'ont les enfants avec l'écrit notamment dans leurs manuels scolaires, l'écriture foncée sur fond pâle. La police d'écriture Arial de taille 30 a été utilisée pour la présentation des stimulus. Cette taille était suffisamment grande pour que le stimulus soit lu facilement, mais sans nécessiter de balayage visuel afin d'identifier les stimulus en une seule fixation. Pour présenter les stimulus de la tâche de décision lexicale, un écran cathodique plutôt qu'un écran DEL (diode électroluminescente) pour lequel le temps de rafraîchissement est trop lent, bien que ce soit non visible à l'œil nu, a été utilisé.

La présentation des stimulus durant la tâche de décision lexicale est représentée à la Figure 3.2. La tâche commençait par une croix de fixation (+) au centre de l'écran qui restait affichée 2 000 ms. Puis, l'item était présenté pendant 3 000 ms ou jusqu'au jugement lexical des participants qui répondaient en appuyant sur les touches identifiées. Ce temps de présentation a été sélectionné comme des temps de présentation du stimulus trop brefs sont habituellement accompagnés d'une moins bonne exactitude chez les participants (Davis *et al.*, 2019). Un intervalle supplémentaire entre les stimulus n'était pas nécessaire puisque cette tâche était déjà remplie par la croix de fixation présentée avant chaque stimulus. C'est d'ailleurs pourquoi la présentation est de 2 000 ms, ce qui est plus long que ce qui est réalisé dans les recherches du tableau 3.2. Aucune rétroaction n'était fournie entre les items. Si les participants ne fournissaient aucune réponse avant les 3 000 ms imparties, le stimulus disparaissait de l'écran pour afficher la croix de fixation du prochain stimulus. Afin d'éviter la fatigue des participants, chaque bloc de 60 items était séparé d'une pause. Les participants pouvaient mettre fin à la pause lorsqu'ils se sentaient prêts à reprendre la tâche en appuyant sur la barre d'espacement du clavier comme indiqué à l'écran. Avec l'entraînement, le temps moyen de la tâche est de 22 minutes et 44 secondes ($\bar{ET} = 2$ minutes et 5 secondes). Le temps moyen de la tâche de décision lexicale uniquement est de 19 minutes et 32 secondes ($\bar{ET} = 1$ minute 56 secondes), ce qui comprend aussi les pauses prises par les participants. La passation la plus longue a duré 29 minutes et 8 secondes, avec l'entraînement.

Durant l'enregistrement, il a été demandé aux participants d'éviter tout mouvement et de ne pas toucher au casque d'électroencéphalographie. Les participants devaient aussi demeurer muets.

Figure 3.2 Présentation des stimulus durant la tâche de décision lexicale



3.3.2 Le test dynamique pour mesurer la connaissance du vocabulaire, la connaissance des racines latines et grecques et les habiletés morphologiques

Dans des recherches précédentes (Crosson et McKeown, 2016; Crosson *et al.*, 2019; Crosson et Moore, 2017), un test dynamique (dans lequel les questions posées dépendent des réponses aux questions antérieures) a été utilisé pour mesurer la connaissance des racines latines et grecques des participants et, par le fait même, leur capacité à procéder à une analyse morphologique pour déduire le sens d'un mot inconnu en s'appuyant sur le sens de la racine. Les tests dynamiques sont reconnus pour leur intérêt à révéler des informations sur les connaissances des participants qui sont toujours en développement (Burton et Watkins, 2007; Poehner et Lantolf, 2010). Selon les différentes adaptations utilisées dans quelques études anglophones (Crosson et McKeown, 2016; Crosson *et al.*, 2019; Crosson et Moore, 2017), ce test compte de 16 à 18 racines latines liées.

Un test similaire a été conçu en sélectionnant 20 stimulus de la tâche de décision lexicale, cinq mots et cinq pseudomots formés de deux racines grecques ainsi que cinq mots et cinq pseudomots formés de deux racines latines (voir Annexe F). Ainsi, ce test permet de mesurer la connaissance de quarante racines latines et grecques. La présence de pseudomots formés de racines latines et grecques existantes permet de mesurer les connaissances morphologiques explicites des élèves (Fejzo, 2011), une limite d'autres tests existants. Le test comprenait deux items de pratique pour exercer les participants à la tâche, un item avant la série de mots (*midi*) et un item avant la série de pseudomots (**biblionaute*).

Le test consistait à demander aux participants de déduire le sens des 20 mots ou pseudomots. Les items cibles étaient soulignés dans des phrases acontextuelles, c'est-à-dire des phrases qui ne fournissaient aucun indice contextuel permettant de déduire le sens du mot ou du pseudomot. Au fil du test individuel, différents indices étaient donnés aux élèves s'ils ne fournissaient pas de réponse suffisamment précise dès la première tentative. Inspiré de Crosson et McKeown (2016), le protocole du test se lisait comme suit : 1) montrer la phrase écrite aux élèves et la lire à haute voix; 2) leur demander de déduire le sens du mot cible ; 3) Amorce 1 : attirer l'attention des élèves sur la structure du mot; 4) Amorce 2 : demander aux élèves le sens des racines; et 5) Amorce 3 : demander à nouveau aux élèves de déduire le sens du mot dans le contexte de la phrase grâce à ces nouveaux indices. Le Tableau 3.5 présente la structure du test dynamique en utilisant l'item *ovipare* comme exemple.

Tableau 3.5 Structure du test dynamique avec l'item *ovipare* comme exemple

Interventions de l'expérimentateur	
Lecture de la phrase	« Dans cet enclos, on retrouve un animal ovipare. »
Question initiale	« Que signifie le mot <i>ovipare</i> ? »
	<u>Relances :</u> <ul style="list-style-type: none"> ✓ Si l'élève utilise le mot cible dans sa définition : « Que veux-tu dire? » ✓ Si l'élève utilise un exemple : « Essaie de le définir. Que signifie le mot <i>ovipare</i> ? » ✓ Si l'élève définit adéquatement le mot, identifie les racines et utilise explicitement le sens de chacune des racines, passer à l'item suivant.
Amorce 1	« Regarde bien le mot <i>ovipare</i> . Y vois-tu des parties de mot que tu connais ? »
	<u>Relances :</u> <ul style="list-style-type: none"> ✓ Si l'élève peut identifier une ou les racines : « Tu peux les encadrer. ». Puis, passer à l'amorce 2. ✓ Si l'élève nomme d'autres mots qui contiennent la même racine, l'amener vers l'amorce 2 en commençant par cette racine : « Oui, tous ces mots ont la même partie de mot en commun. Que crois-tu qu'elle signifie. Quel est son sens? » ✓ Si l'élève n'identifie aucune racine, séparer le mot en morphèmes à l'aide d'un trait, puis passer à l'amorce 2.
Amorce 2	« Que signifient ces parties du mot ? »
	<u>Relances :</u> <ul style="list-style-type: none"> ✓ Si l'élève donne une réponse : « Tu peux l'écrire au-dessus. ». Puis, passer à l'amorce 3. ✓ Si l'élève n'identifie qu'une racine : « Sais-tu ce que signifie l'autre partie du mot? ». Puis, passer à l'amorce 3.

Interventions de l'expérimentateur

Amorce 3 Fournir uniquement les réponses qui n'ont pas été données par l'élève ou les réponses erronées et écrire le sens au-dessus de chaque racine.
« Je vais te donner un indice. *Ovi* est une partie de mot qui signifie "œuf" et *pare* est une partie de mot qui signifie "mettre au monde". »
« Est-ce que cela te donne d'autres indices sur le mot *ovipare*? Que signifie le mot *ovipare* ?»

Relances :

- ✓ Si l'élève définit adéquatement le mot en utilisant explicitement le sens de chacune des racines, passer à l'item suivant.
 - ✓ Si l'élève ne donne pas de réponse ou donne une réponse incomplète ou erronée, passer à l'item suivant.
-

Dans ce test dynamique, la question initiale permet de mesurer la connaissance du vocabulaire en questionnant les participants sur le sens du mot cible. Les deux premières amorces, quant à elles, permettent de renseigner sur la connaissance des racines latines et grecques : l'habileté à les reconnaître et à décomposer le mot en morphèmes (Amorce 1) et à leur attribuer un sens (Amorce 2). Enfin, la dernière amorce renseigne sur l'habileté d'analyse morphologique des participants, soit d'utiliser le sens des racines pour déduire le sens d'un mot inconnu (Amorce 3).

La connaissance du vocabulaire a été codifiée selon la complétude de la définition des participants. Une définition complète s'est vu attribuer une note de 2 (p. ex. pour *ovipare* : « les bébés naissent dans des œufs »). Une définition partielle qui ne contenait, par exemple, qu'un des deux concepts centraux du mot cible recevait une note de 1 (p. ex. pour *ovipare* : « qui mange des œufs »). Si la définition était erronée ou si le mot était inconnu, la note 0 était attribuée (p. ex. « aucune idée »). Si la définition de l'élève était complète dès la première question, mais qu'il n'utilisait pas explicitement le sens des racines dans sa définition les questions en guise d'amorce étaient posées afin de mesurer les autres connaissances et habiletés. Cependant, si le participant fournissait une définition complète en utilisant explicitement le sens des racines (p. ex. pour *ovipare* : « Je sais qu'*ovi* veut dire *œuf* et que *pare* peut dire *mettre au monde*. Donc, un animal *ovipare* est un animal qui met au monde des bébés dans des œufs. »), les questions en guise d'amorce n'étaient pas posées au participant et tous les points étaient accordés pour toutes les amorces. L'administrateur du test passait ensuite à l'item suivant. La connaissance du vocabulaire n'a pas été notée pour les pseudomots.

À la suite de la première amorce (Amorce 1), deux mesures distinctes étaient codifiées : la reconnaissance des racines et la décomposition morphologique. La reconnaissance des racines réfère à

l'habileté des participants à nommer une racine, à constater sa présence dans le mot cible. Durant la codification du test, il a été constaté que certains participants ne verbalisaient pas nécessairement les deux racines du mot, mais le décomposaient adéquatement en tirant un trait entre les deux morphèmes (p. ex. « peut-être si on le sépare ici, au milieu »). Alors, il était impossible de présumer que ces élèves reconnaissaient automatiquement les deux racines. Il est probable que ces participants ne reconnaissaient qu'une seule racine et séparaient le mot avant ou après celle-ci. Sinon, il est aussi probable que les participants aient scindé le mot en deux sur la base d'autres critères non verbalisés (p. ex. nombre de syllabes, longueur du mot), mais que cela entraîne une décomposition morphologique juste. De ce fait, afin d'éviter de confondre l'habileté à reconnaître une racine dans le mot cible et l'habileté des élèves à décomposer le mot en morphèmes, ces deux habiletés ont été codifiées séparément. Bien entendu, l'attribution d'une note exacte à la reconnaissance des morphèmes entraînait aussi une note exacte à la décomposition morphologique, mais l'inverse n'était pas toujours le cas. De façon détaillée, pour la reconnaissance des racines, lorsque les participants nommaient les racines dans le mot cible, ils recevaient un point pour chaque racine reconnue (p. ex. pour *auditorium* : « je séparerais, mais j'enlèverais le *T*, donc *audi* et *orium* »). La délimitation des frontières des racines devait être claire. Une délimitation inexacte entraînait une note de 0 (p. ex. pour *funambule* : « *funam* et *bule* »). Pour l'habileté à décomposer le mot en morphèmes, un point a été attribué à une décomposition exacte (p. ex. pour *hydrophobe* : « moi *hydrophobe*, je vois *hydro* et *phobe* ») et aucun point pour une mauvaise décomposition (p. ex. pour *funambule* : « *funam* et *ambule* »). D'ailleurs, dans ce dernier exemple, le participant recevait 1 point sur un maximum de 2 pour la reconnaissance des racines (ayant reconnu la racine *ambule*), mais aucun point pour la décomposition morphologique. Aucun point n'était accordé pour une absence de réponse.

Le sens des racines du mot (Amorce 2) a été codé grâce à une échelle à quatre valeurs. Les participants recevaient 3 points lorsqu'ils étaient en mesure de fournir le sens exact de la racine ou un synonyme (p. ex. pour *audi* dans *auditorium* : « *audi* c'est que tu entends »). Les participants recevaient 2 points lorsque la définition était partielle, en utilisant un mot de la même famille morphologique pour définir le mot cible (p. ex. pour *audi* dans *auditorium* : « *audi* veut dire *audition* ») ou en utilisant un mot appartenant au champ lexical du sens de la racine (p. ex. pour *audi* dans *auditorium* : « c'est en rapport avec les oreilles »). Les participants recevaient un point lorsqu'ils n'étaient pas en mesure de définir le sens de la racine, mais qu'ils nommaient un autre mot qui contient cette même racine (p. ex. pour *ambule* dans *funambule* : « ça ressemble à *somnambule*, mais je ne sais pas ce que ça veut dire »). Enfin,

aucun point n'était attribué lorsque les participants fournissaient un sens erroné (p. ex. pour *pare* dans *ovipare* : « *pare* ça veut dire *proie* peut-être ») ou ne fournissaient aucune définition (p. ex. « je ne sais pas »). Ainsi, les participants pouvaient recevoir un maximum de 6 points par mot ou pseudomot pour mesurer leur connaissance du sens des racines.

Enfin, l'analyse morphologique (Amorce 3), soit la capacité à utiliser le sens des racines pour déduire le sens du mot, pouvait recevoir un maximum de 2 points par mot. Si les participants fournissaient une réponse juste du sens du mot en utilisant le sens des deux racines (que le sens leur soit déjà connu ou qu'il ait été fourni par l'expérimentateur), ils recevaient 2 points (p. ex. pour *trimestre* : « dans trois mois »). Si les participants fournissaient une réponse partielle en n'utilisant le sens que d'une seule racine, ils recevaient 1 point (p. ex. pour *trimestre* : « au mois prochain »). Enfin, si les participants n'arrivaient pas à définir le mot à l'aide du sens des racines latines et grecques, ils ne recevaient aucun point (p. ex. pour *trimestre* : « une partie de l'année »).

3.3.2.1 La validation du test dynamique

Comme le test dynamique était à concevoir, un aspect de la validité de construit a été vérifié en faisant passer une première version du test à deux élèves répondant à tous les critères d'inclusion, mais ne participant pas à la recherche. Il s'agit des mêmes élèves que ceux ayant participé au pilote de la tâche de décision lexicale (voir section 3.3.1). La question initiale du test reprenait textuellement la formulation de Crosson et McKeown (2016) : « Que peux-tu me dire au sujet de... » Par exemple, pour l'item *trimestre*, la phrase était : « La réunion aura lieu au trimestre prochain. » La chercheuse principale posait la question initiale à l'élève : « Que peux-tu me dire au sujet de la réunion? » Lors de la validation, les élèves répondaient : « Elle aura lieu au trimestre prochain. » La question initiale n'était donc pas suffisamment orientée vers le sens du mot. À cet effet, elle a été reformulée : « Que signifie le mot *trimestre*? » Aussi, le mot ciblé a été souligné afin que les élèves y portent exclusivement leur attention.

3.4 La procédure de la collecte de données

La collecte de données a eu lieu du mois de mai à juillet 2022. Dès que le contact a été établi avec les parents des participants, une première rencontre a été fixée afin de faire passer les deux prétests aux participants, le test d'identification des mots écrits (Mousty *et al.*, 1994) et le test d'intelligence non verbale (Raven, 1983) (voir section 3.2.2). Cette rencontre s'est déroulée à distance, en visioconférence par l'application Zoom ou Teams selon la préférence des parents, et durait au maximum 30 minutes.

C'est aussi à ce moment que les participants et leurs parents signaient le formulaire d'information et de consentement.

Si les résultats des participants aux prétests correspondaient aux normes recherchées, une autre rencontre était fixée pour la visite au laboratoire et la deuxième série de tests. Accompagnés de leurs parents, les participants se sont rendus au laboratoire de recherche NeuroLab, à l'Université du Québec à Montréal. Durant cette seconde rencontre, les participants ont réalisé la tâche de décision lexicale ainsi que le test dynamique. Deux types de données ont été recueillies durant la tâche de décision lexicale : des données comportementales et des données psychophysiologiques par l'analyse des potentiels évoqués. Les données comportementales recueillies pour cette tâche de décision lexicale concernent la précision et le temps de réaction. En synchronie avec la tâche de décision lexicale, les données psychophysiologiques relatives à l'activité électrique cérébrale ont été recueillies par l'électroencéphalographie.

À leur arrivée au laboratoire, les participants et les parents ont été accueillis et une visite rapide des installations a été effectuée. Par la suite, la chercheuse revoyait le formulaire d'information et de consentement déjà signé avec les participants et les parents. Puis, les participants accompagnés de leurs parents entraient dans la salle de collecte de données et le bonnet préalablement assemblé avec les électrodes était installé sur la tête des participants. L'enregistrement électroencéphalographique a été réalisé avec un montage à 64 électrodes disposées selon le système topographique 10-10 (Chatrian *et al.*, 1985) et référencé en ligne au FCz. Utiliser 64 électrodes plutôt que 32 permet d'obtenir une meilleure correction des mouvements oculaires après l'enregistrement. Dans chacune des électrodes, du gel conducteur a été inséré à l'aide d'une petite seringue pour assurer un contact entre le scalp et l'électrode. Bien que le système recommande une impédance de 25 kOhm pour assurer un bon signal, une impédance sous 10 kOhm a été ciblée pour toutes les électrodes. Cette précaution assure de conserver une impédance suffisamment basse et stable tout au long de l'expérimentation. L'impédance des deux électrodes de référence (mise à la terre et référence) était inférieure à 10 kOhm pour tous les participants (moyenne de 5,00 kOhm pour la mise à la terre et de 2,62 kOhm pour la référence). En moyenne pour tous les participants, l'impédance de 92,08 % des électrodes du scalp est demeurée sous 10 kOhm et ce pourcentage augmente à 98,00 % sous 25 kOhm.

Pendant l'installation du bonnet, les participants pouvaient jouer sur une tablette numérique fournie par l'équipe de recherche. Une fois les participants bien installés, les parents et les assistants de recherche quittaient la salle pour attendre dans une salle adjacente. Puis, les consignes de la tâche de décision lexicale étaient données aux participants. La chercheuse accompagnait les participants pour un court entraînement avant de se déplacer derrière un miroir sans tain pour la réalisation de la tâche. Au besoin, la chercheuse pouvait communiquer avec les participants à l'aide d'un micro. Une fois la tâche terminée, le bonnet était retiré et les participants pouvaient laver leurs cheveux.

À la suite d'une pause pour les participants où une collation était fournie, cette seconde séance de tests se terminait avec le test dynamique (voir section 3.3.2). Ce test s'est déroulé dans un local de recherche près du laboratoire. À la fin de la collecte de données, la compensation financière a été remise aux familles avant leur départ du laboratoire.

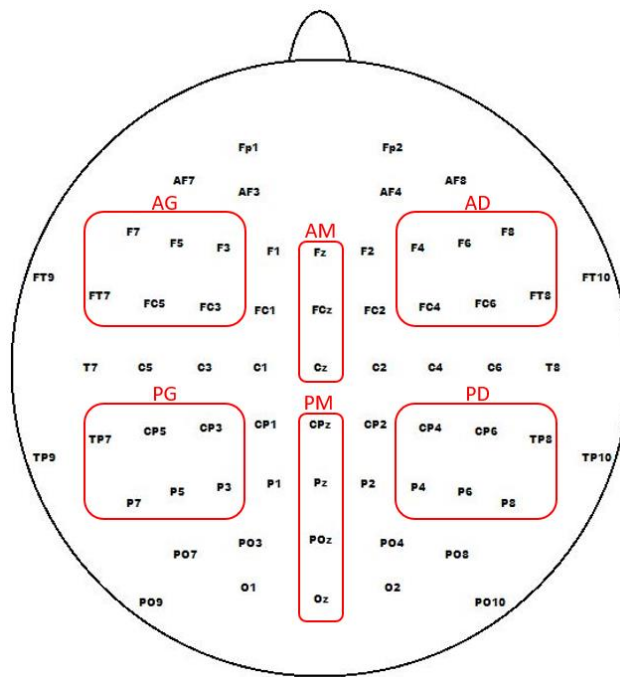
3.5 La préparation des données électroencéphalographiques

Les données électroencéphalographiques ont été collectées grâce au logiciel *BrainVision Recorder Professional* version 1.21.0402 (Brain Vision Solutions, 2020) et amplifiées à 1 000 Hz. Le webinaire de Brain Products GmbH (2018) a servi de guide pour le nettoyage des données électroencéphalographiques et l'extraction des potentiels évoqués. À la suite de l'enregistrement électroencéphalographique, un filtre a été appliqué pour retirer les fréquences inférieures à 1,1 Hz et supérieures à 30 Hz. Un filtre coupe-bande à 60 Hz a aussi été appliqué. Puis, une inspection semi-automatique des données brutes a été effectuée pour retirer automatiquement les portions du signal électrique à haut voltage (supérieures à 15 $\mu\text{V}/\text{ms}$) ou à très bas voltage (inférieures à 0,5 μV). Le signal entier a été inspecté manuellement pour permettre le retrait d'autres artefacts qui n'auraient pas été automatiquement détectés, comme ceux associés à un mouvement moteur. Puis, une correction oculaire par l'analyse de composantes indépendantes [ma traduction de *independent component analysis* (ICA)] a permis de rétablir les artefacts causés par les clignements et les mouvements latéraux des yeux en utilisant l'électrode FP1 comme référence pour les mouvements oculaires verticaux et l'électrode F8 comme référence pour les mouvements oculaires horizontaux.

À partir de ce nouveau fichier, les données ont été référencées hors ligne à la moyenne des électrodes des deux mastoïdes, les électrodes TP9 et TP10 (Arcara *et al.*, 2013; Coch *et al.*, 2013; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014). La référence en ligne (FCz) a été conservée. À l'instar de Fiorentino et ses

collègues (2014), six régions topographiques d'intérêt ont été définies: antérieure gauche (F3, F5, F7, FC3, FC5, FT7), antérieure médiane (Fz, FCz, Cz), antérieure droite (F4, F6, F8, FC4, FC6, FT8), postérieure gauche (P3, P5, P7, CP3, CP5, TP7), postérieure médiane (CPz, Pz, POz, Oz) et postérieure droite (P4, P6, P8, CP4, CP6, TP8). Les électrodes utilisées dans le montage de Fiorentino et de ses collègues (2014) ont servi de frontières pour déterminer celles à intégrer dans les six régions topographiques d'intérêt de la présente étude. La Figure 3.3 présente les six régions topographiques sélectionnées ainsi que les électrodes sélectionnées pour chacune d'elles.

Figure 3.3 Représentation schématique des électrodes dans chacune des six régions topographiques d'intérêt



Note. AG = antérieure gauche; AM = antérieure médiane; AD = antérieure droite; PG = postérieure gauche; PM = postérieure médiane; PD = postérieure droite.

Par la suite, pour chaque condition expérimentale, les données ont été initialement segmentées de 200 ms préstimulus (pour le niveau de base [ma traduction de *baseline*]) à 3 000 ms poststimulus de façon à conserver uniquement les essais correctement répondus par les participants. Une deuxième segmentation (de -200 ms à 1 000 ms) a permis de restreindre les fenêtres d'observation de chaque essai correctement répondu à au plus 1 000 ms poststimulus, puisque les processus cognitifs ciblés dans

la présente étude ne sont pas susceptibles de se produire au-delà de 1000 ms après la présentation du stimulus. Chaque segment a ensuite été corrigé pour le niveau de base puis les segments d'une même condition ont été rapportés à la moyenne par électrode et par région d'intérêt. À cette étape, les segments contenant un artéfact ont automatiquement été rejetés, soit une perte moyenne de 25,25 % des segments dans lesquels les participants fournissaient une réponse correcte. Le Tableau 3.6 présente le nombre de segments utilisés et perdus pour chacune des conditions expérimentales. Le Tableau 3.7 présente ces mêmes données par participant.

Tableau 3.6 Nombre de segments utilisés et perdus pour chacune des conditions expérimentales

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(1) Mots formés de deux racines latines	547	62,87	439	80,26	108	19,74
(2) Mots formés de deux racines grecques	596	68,51	474	79,53	122	20,47
(3) Mots formés de deux racines libres	553	63,56	436	78,84	117	21,16
(4) Mots monomorphémiques	733	84,25	586	79,95	147	20,05
(5) Pseudomots formés de deux racines latines	761	87,47	605	79,50	156	20,50
(6) Pseudomots formés de deux racines grecques	695	79,89	565	81,29	130	18,71
(7) Pseudomots formés de deux racines libres	793	91,15	634	79,95	159	20,05
(8) Non-mots	812	93,33	645	79,43	167	20,57
Moyenne (écart type)	686,25 (107,17)	78,88 (12,31)	548,00 (85,93)	79,84 (0,73)	138,25 (21,87)	20,16 (0,73)

Note. Le maximum de segments par condition est de 870 (30 stimulus × 29 participants).

^a Le pourcentage est calculé par rapport au nombre maximal de segments possible par condition.

^b Le pourcentage est calculé par rapport au nombre de segments corrects.

^c Le pourcentage est calculé par rapport au nombre de segments corrects.

Tableau 3.7 Nombre de segments utilisés et perdus par participant pour chacune des conditions expérimentales

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
Participant 1						
(1) Mots formés de racines latines	24	80,00	23	95,83	1	4,17
(2) Mots formés de racines grecques	26	86,67	26	100,00	0	0,00
(3) Mots formés de racines libres	15	50,00	15	100,00	0	0,00
(4) Mots monomorphémiques	27	90,00	26	96,30	1	3,70
(5) Pseudomots formés de racines latines	30	100,00	28	93,33	2	6,67
(6) Pseudomots formés de racines grecques	28	93,33	26	92,86	2	7,14
(7) Pseudomots formés de racines libres	29	96,67	28	96,55	1	3,45
(8) Non-mots	30	100,00	29	96,67	1	3,33
Participant 2						
(1) Mots formés de racines latines	11	36,67	10	90,91	1	9,09
(2) Mots formés de racines grecques	11	36,67	11	100,00	0	0,00
(3) Mots formés de racines libres	7	23,33	7	100,00	0	0,00
(4) Mots monomorphémiques	20	66,67	19	95,00	1	5,00
(5) Pseudomots formés de racines latines	23	76,67	21	91,30	2	8,70
(6) Pseudomots formés de racines grecques	23	76,67	23	100,00	0	0,00
(7) Pseudomots formés de racines libres	26	86,67	25	96,15	1	3,85
(8) Non-mots	29	96,67	28	96,55	1	3,45
Participant 3						
(1) Mots formés de racines latines	16	53,33	5	31,25	11	68,75
(2) Mots formés de racines grecques	18	60,00	4	22,22	14	77,78
(3) Mots formés de racines libres	17	56,67	1	5,88	16	94,12
(4) Mots monomorphémiques	23	76,67	8	34,78	15	65,22
(5) Pseudomots formés de racines latines	25	83,33	8	32,00	17	68,00
(6) Pseudomots formés de racines grecques	22	73,33	5	22,73	17	77,27
(7) Pseudomots formés de racines libres	27	90,00	8	29,63	19	70,37
(8) Non-mots	28	93,33	7	25,00	21	75,00
Participant 4						
(1) Mots formés de racines latines	15	50,00	14	93,33	1	6,67
(2) Mots formés de racines grecques	20	66,67	12	60,00	8	40,00
(3) Mots formés de racines libres	20	66,67	14	70,00	6	30,00
(4) Mots monomorphémiques	22	73,33	18	81,82	4	18,18
(5) Pseudomots formés de racines latines	27	90,00	24	88,89	3	11,11
(6) Pseudomots formés de racines grecques	25	83,33	20	80,00	5	20,00
(7) Pseudomots formés de racines libres	29	96,67	26	89,66	3	10,34
(8) Non-mots	29	96,67	23	79,31	6	20,69
Participant 5						
(1) Mots formés de racines latines	18	60,00	16	88,89	2	11,11
(2) Mots formés de racines grecques	18	60,00	15	83,33	3	16,67
(3) Mots formés de racines libres	15	50,00	15	100,00	0	0,00
(4) Mots monomorphémiques	27	90,00	23	85,19	4	14,81
(5) Pseudomots formés de racines latines	29	96,67	28	96,55	1	3,45
(6) Pseudomots formés de racines grecques	27	90,00	25	92,59	2	7,41
(7) Pseudomots formés de racines libres	27	90,00	24	88,89	3	11,11

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(8) Non-mots	26	86,67	26	100,00	0	0,00
Participant 6						
(1) Mots formés de racines latines	18	60,00	13	72,22	5	27,78
(2) Mots formés de racines grecques	25	83,33	16	64,00	9	36,00
(3) Mots formés de racines libres	19	63,33	12	63,16	7	36,84
(4) Mots monomorphémiques	28	93,33	19	67,86	9	32,14
(5) Pseudomots formés de racines latines	29	96,67	24	82,76	5	17,24
(6) Pseudomots formés de racines grecques	21	70,00	15	71,43	6	28,57
(7) Pseudomots formés de racines libres	30	100,00	23	76,67	7	23,33
(8) Non-mots	29	96,67	21	72,41	8	27,59
Participant 7						
(1) Mots formés de racines latines	18	60,00	18	100,00	0	0,00
(2) Mots formés de racines grecques	18	60,00	16	88,89	2	11,11
(3) Mots formés de racines libres	19	63,33	18	94,74	1	5,26
(4) Mots monomorphémiques	24	80,00	24	100,00	0	0,00
(5) Pseudomots formés de racines latines	29	96,67	28	96,55	1	3,45
(6) Pseudomots formés de racines grecques	25	83,33	24	96,00	1	4,00
(7) Pseudomots formés de racines libres	28	93,33	28	100,00	0	0,00
(8) Non-mots	29	96,67	26	89,66	3	10,34
Participant 8						
(1) Mots formés de racines latines	14	46,67	9	64,29	5	35,71
(2) Mots formés de racines grecques	21	70,00	13	61,90	8	38,10
(3) Mots formés de racines libres	15	50,00	11	73,33	4	26,67
(4) Mots monomorphémiques	27	90,00	21	77,78	6	22,22
(5) Pseudomots formés de racines latines	25	83,33	20	80,00	5	20,00
(6) Pseudomots formés de racines grecques	21	70,00	16	76,19	5	23,81
(7) Pseudomots formés de racines libres	30	100,00	25	83,33	5	16,67
(8) Non-mots	27	90,00	15	55,56	12	44,44
Participant 9						
(1) Mots formés de racines latines	22	73,33	21	95,45	1	4,55
(2) Mots formés de racines grecques	21	70,00	17	80,95	4	19,05
(3) Mots formés de racines libres	21	70,00	17	80,95	4	19,05
(4) Mots monomorphémiques	24	80,00	18	75,00	6	25,00
(5) Pseudomots formés de racines latines	19	63,33	19	100,00	0	0,00
(6) Pseudomots formés de racines grecques	18	60,00	16	88,89	2	11,11
(7) Pseudomots formés de racines libres	28	93,33	24	85,71	4	14,29
(8) Non-mots	26	86,67	22	84,62	4	15,38
Participant 10						
(1) Mots formés de racines latines	17	56,67	12	70,59	5	29,41
(2) Mots formés de racines grecques	19	63,33	16	84,21	3	15,79
(3) Mots formés de racines libres	19	63,33	16	84,21	3	15,79
(4) Mots monomorphémiques	20	66,67	19	95,00	1	5,00
(5) Pseudomots formés de racines latines	24	80,00	18	75,00	6	25,00
(6) Pseudomots formés de racines grecques	24	80,00	20	83,33	4	16,67
(7) Pseudomots formés de racines libres	20	66,67	16	80,00	4	20,00
(8) Non-mots	26	86,67	25	96,15	1	3,85
Participant 11						
(1) Mots formés de racines latines	24	80,00	22	91,67	2	8,33
(2) Mots formés de racines grecques	27	90,00	26	96,30	1	3,70

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(3) Mots formés de racines libres	25	83,33	21	84,00	4	16,00
(4) Mots monomorphémiques	27	90,00	23	85,19	4	14,81
(5) Pseudomots formés de racines latines	29	96,67	26	89,66	3	10,34
(6) Pseudomots formés de racines grecques	24	80,00	21	87,50	3	12,50
(7) Pseudomots formés de racines libres	28	93,33	25	89,29	3	10,71
(8) Non-mots	30	100,00	26	86,67	4	13,33
Participant 12						
(1) Mots formés de racines latines	17	56,67	17	100,00	0	0,00
(2) Mots formés de racines grecques	24	80,00	24	100,00	0	0,00
(3) Mots formés de racines libres	19	63,33	19	100,00	0	0,00
(4) Mots monomorphémiques	28	93,33	28	100,00	0	0,00
(5) Pseudomots formés de racines latines	28	93,33	28	100,00	0	0,00
(6) Pseudomots formés de racines grecques	25	83,33	25	100,00	0	0,00
(7) Pseudomots formés de racines libres	29	96,67	29	100,00	0	0,00
(8) Non-mots	30	100,00	30	100,00	0	0,00
Participant 13						
(1) Mots formés de racines latines	18	60,00	16	88,89	2	11,11
(2) Mots formés de racines grecques	12	40,00	12	100,00	0	0,00
(3) Mots formés de racines libres	12	40,00	11	91,67	1	8,33
(4) Mots monomorphémiques	26	86,67	22	84,62	4	15,38
(5) Pseudomots formés de racines latines	28	93,33	22	78,57	6	21,43
(6) Pseudomots formés de racines grecques	25	83,33	23	92,00	2	8,00
(7) Pseudomots formés de racines libres	30	100,00	27	90,00	3	10,00
(8) Non-mots	28	93,33	26	92,86	2	7,14
Participant 14						
(1) Mots formés de racines latines	19	63,33	16	84,21	3	15,79
(2) Mots formés de racines grecques	15	50,00	15	100,00	0	0,00
(3) Mots formés de racines libres	20	66,67	19	95,00	1	5,00
(4) Mots monomorphémiques	24	80,00	18	75,00	6	25,00
(5) Pseudomots formés de racines latines	28	93,33	23	82,14	5	17,86
(6) Pseudomots formés de racines grecques	26	86,67	22	84,62	4	15,38
(7) Pseudomots formés de racines libres	29	96,67	24	82,76	5	17,24
(8) Non-mots	30	100,00	25	83,33	5	16,67
Participant 15						
(1) Mots formés de racines latines	23	76,67	23	100,00	0	0,00
(2) Mots formés de racines grecques	22	73,33	21	95,45	1	4,55
(3) Mots formés de racines libres	28	93,33	27	96,43	1	3,57
(4) Mots monomorphémiques	29	96,67	28	96,55	1	3,45
(5) Pseudomots formés de racines latines	23	76,67	23	100,00	0	0,00
(6) Pseudomots formés de racines grecques	22	73,33	22	100,00	0	0,00
(7) Pseudomots formés de racines libres	27	90,00	25	92,59	2	7,41
(8) Non-mots	29	96,67	29	100,00	0	0,00
Participant 16						
(1) Mots formés de racines latines	23	76,67	19	82,61	4	17,39
(2) Mots formés de racines grecques	25	83,33	22	88,00	3	12,00
(3) Mots formés de racines libres	23	76,67	20	86,96	3	13,04
(4) Mots monomorphémiques	26	86,67	22	84,62	4	15,38
(5) Pseudomots formés de racines latines	30	100,00	27	90,00	3	10,00
(6) Pseudomots formés de racines grecques	27	90,00	26	96,30	1	3,70

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(7) Pseudomots formés de racines libres	30	100,00	28	93,33	2	6,67
(8) Non-mots	29	96,67	24	82,76	5	17,24
Participant 17						
(1) Mots formés de racines latines	20	66,67	17	85,00	3	15,00
(2) Mots formés de racines grecques	21	70,00	14	66,67	7	33,33
(3) Mots formés de racines libres	24	80,00	14	58,33	10	41,67
(4) Mots monomorphémiques	29	96,67	21	72,41	8	27,59
(5) Pseudomots formés de racines latines	24	80,00	11	45,83	13	54,17
(6) Pseudomots formés de racines grecques	20	66,67	10	50,00	10	50,00
(7) Pseudomots formés de racines libres	24	80,00	15	62,50	9	37,50
(8) Non-mots	29	96,67	21	72,41	8	27,59
Participant 18						
(1) Mots formés de racines latines	20	66,67	15	75,00	5	25,00
(2) Mots formés de racines grecques	23	76,67	18	78,26	5	21,74
(3) Mots formés de racines libres	23	76,67	17	73,91	6	26,09
(4) Mots monomorphémiques	24	80,00	18	75,00	6	25,00
(5) Pseudomots formés de racines latines	24	80,00	16	66,67	8	33,33
(6) Pseudomots formés de racines grecques	24	80,00	21	87,50	3	12,50
(7) Pseudomots formés de racines libres	23	76,67	19	82,61	4	17,39
(8) Non-mots	28	93,33	21	75,00	7	25,00
Participant 19						
(1) Mots formés de racines latines	20	66,67	17	85,00	3	15,00
(2) Mots formés de racines grecques	27	90,00	21	77,78	6	22,22
(3) Mots formés de racines libres	20	66,67	15	75,00	5	25,00
(4) Mots monomorphémiques	26	86,67	22	84,62	4	15,38
(5) Pseudomots formés de racines latines	18	60,00	15	83,33	3	16,67
(6) Pseudomots formés de racines grecques	13	43,33	12	92,31	1	7,69
(7) Pseudomots formés de racines libres	22	73,33	17	77,27	5	22,73
(8) Non-mots	16	53,33	14	87,50	2	12,50
Participant 20						
(1) Mots formés de racines latines	18	60,00	9	50,00	9	50,00
(2) Mots formés de racines grecques	16	53,33	9	56,25	7	43,75
(3) Mots formés de racines libres	22	73,33	14	63,64	8	36,36
(4) Mots monomorphémiques	25	83,33	15	60,00	10	40,00
(5) Pseudomots formés de racines latines	26	86,67	15	57,69	11	42,31
(6) Pseudomots formés de racines grecques	25	83,33	15	60,00	10	40,00
(7) Pseudomots formés de racines libres	27	90,00	15	55,56	12	44,44
(8) Non-mots	29	96,67	14	48,28	15	51,72
Participant 21						
(1) Mots formés de racines latines	14	46,67	6	42,86	8	57,14
(2) Mots formés de racines grecques	14	46,67	5	35,71	9	64,29
(3) Mots formés de racines libres	8	26,67	4	50,00	4	50,00
(4) Mots monomorphémiques	21	70,00	6	28,57	15	71,43
(5) Pseudomots formés de racines latines	30	100,00	14	46,67	16	53,33
(6) Pseudomots formés de racines grecques	26	86,67	11	42,31	15	57,69
(7) Pseudomots formés de racines libres	29	96,67	12	41,38	17	58,62
(8) Non-mots	26	86,67	13	50,00	13	50,00
Participant 22						
(1) Mots formés de racines latines	18	60,00	11	61,11	7	38,89

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(2) Mots formés de racines grecques	14	46,67	11	78,57	3	21,43
(3) Mots formés de racines libres	23	76,67	20	86,96	3	13,04
(4) Mots monomorphémiques	24	80,00	19	79,17	5	20,83
(5) Pseudomots formés de racines latines	29	96,67	23	79,31	6	20,69
(6) Pseudomots formés de racines grecques	27	90,00	21	77,78	6	22,22
(7) Pseudomots formés de racines libres	30	100,00	24	80,00	6	20,00
(8) Non-mots	30	100,00	24	80,00	6	20,00
Participant 23						
(1) Mots formés de racines latines	17	56,67	15	88,24	2	11,76
(2) Mots formés de racines grecques	21	70,00	17	80,95	4	19,05
(3) Mots formés de racines libres	13	43,33	11	84,62	2	15,38
(4) Mots monomorphémiques	24	80,00	20	83,33	4	16,67
(5) Pseudomots formés de racines latines	28	93,33	17	60,71	11	39,29
(6) Pseudomots formés de racines grecques	29	96,67	26	89,66	3	10,34
(7) Pseudomots formés de racines libres	29	96,67	18	62,07	11	37,93
(8) Non-mots	30	100,00	22	73,33	8	26,67
Participant 24						
(1) Mots formés de racines latines	25	83,33	12	48,00	13	52,00
(2) Mots formés de racines grecques	27	90,00	13	48,15	14	51,85
(3) Mots formés de racines libres	28	93,33	13	46,43	15	53,57
(4) Mots monomorphémiques	28	93,33	13	46,43	15	53,57
(5) Pseudomots formés de racines latines	28	93,33	14	50,00	14	50,00
(6) Pseudomots formés de racines grecques	25	83,33	15	60,00	10	40,00
(7) Pseudomots formés de racines libres	27	90,00	12	44,44	15	55,56
(8) Non-mots	28	93,33	12	42,86	16	57,14
Participant 25						
(1) Mots formés de racines latines	21	70,00	21	100,00	0	0,00
(2) Mots formés de racines grecques	19	63,33	18	94,74	1	5,26
(3) Mots formés de racines libres	21	70,00	18	85,71	3	14,29
(4) Mots monomorphémiques	26	86,67	26	100,00	0	0,00
(5) Pseudomots formés de racines latines	25	83,33	23	92,00	2	8,00
(6) Pseudomots formés de racines grecques	28	93,33	26	92,86	2	7,14
(7) Pseudomots formés de racines libres	29	96,67	26	89,66	3	10,34
(8) Non-mots	29	96,67	24	82,76	5	17,24
Participant 26						
(1) Mots formés de racines latines	15	50,00	12	80,00	3	20,00
(2) Mots formés de racines grecques	20	66,67	19	95,00	1	5,00
(3) Mots formés de racines libres	16	53,33	16	100,00	0	0,00
(4) Mots monomorphémiques	26	86,67	19	73,08	7	26,92
(5) Pseudomots formés de racines latines	24	80,00	21	87,50	3	12,50
(6) Pseudomots formés de racines grecques	27	90,00	25	92,59	2	7,41
(7) Pseudomots formés de racines libres	26	86,67	23	88,46	3	11,54
(8) Non-mots	28	93,33	22	78,57	6	21,43
Participant 27						
(1) Mots formés de racines latines	16	53,33	11	68,75	5	31,25
(2) Mots formés de racines grecques	21	70,00	17	80,95	4	19,05
(3) Mots formés de racines libres	18	60,00	11	61,11	7	38,89
(4) Mots monomorphémiques	23	76,67	19	82,61	4	17,39
(5) Pseudomots formés de racines latines	28	93,33	21	75,00	7	25,00

Condition expérimentale	Nombre de segments corrects		Nombre de segments utilisés pour la moyenne		Perte de segments	
	n	% ^a	n	% ^b	n	% ^c
(6) Pseudomots formés de racines grecques	27	90,00	16	59,26	11	40,74
(7) Pseudomots formés de racines libres	27	90,00	22	81,48	5	18,52
(8) Non-mots	29	96,67	23	79,31	6	20,69
Participant 28						
(1) Mots formés de racines latines	25	83,33	20	80,00	5	20,00
(2) Mots formés de racines grecques	25	83,33	23	92,00	2	8,00
(3) Mots formés de racines libres	25	83,33	23	92,00	2	8,00
(4) Mots monomorphémiques	29	96,67	27	93,10	2	6,90
(5) Pseudomots formés de racines latines	24	80,00	22	91,67	2	8,33
(6) Pseudomots formés de racines grecques	16	53,33	15	93,75	1	6,25
(7) Pseudomots formés de racines libres	28	93,33	26	92,86	2	7,14
(8) Non-mots	28	93,33	27	96,43	1	3,57
Participant 29						
(1) Mots formés de racines latines	21	70,00	19	90,48	2	9,52
(2) Mots formés de racines grecques	26	86,67	23	88,46	3	11,54
(3) Mots formés de racines libres	18	60,00	17	94,44	1	5,56
(4) Mots monomorphémiques	26	86,67	25	96,15	1	3,85
(5) Pseudomots formés de racines latines	27	90,00	26	96,30	1	3,70
(6) Pseudomots formés de racines grecques	25	83,33	23	92,00	2	8,00
(7) Pseudomots formés de racines libres	25	83,33	20	80,00	5	20,00
(8) Non-mots	27	90,00	26	96,30	1	3,70

Note. Le maximum de segments par condition pour chaque participant est de 30.

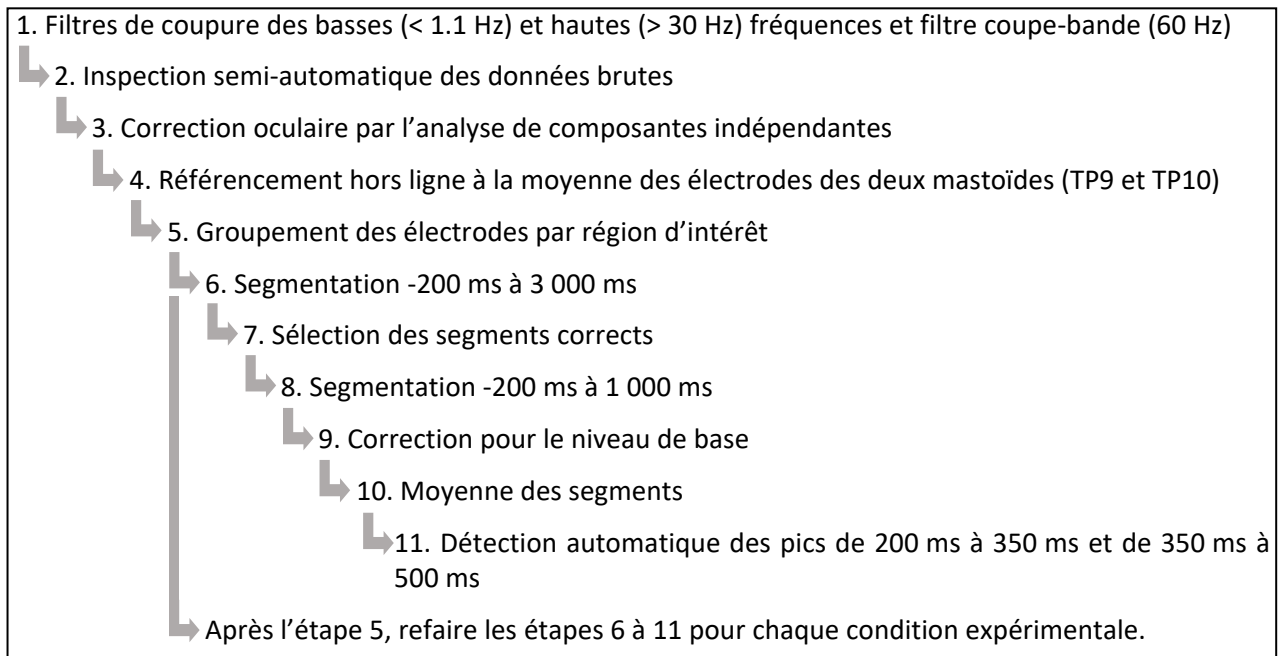
^a Le pourcentage est calculé par rapport au nombre maximal de segments possible par condition.

^b Le pourcentage est calculé par rapport au nombre de segments corrects.

^c Le pourcentage est calculé par rapport au nombre de segments corrects.

Par la suite, une détection automatique de l'amplitude la plus élevée dans les deux fenêtres temporelles d'intérêt, de 200 ms à 350 ms et de 350 ms à 500 ms, a été effectuée. L'amplitude et la latence de chacun de ces pics moyennés par condition ont enfin été exportées séparément pour chacun des participants et pour chacune des régions d'intérêt. La Figure 3.4 présente les étapes de préparation réalisées sur les enregistrements électroencéphalographiques bruts de chacun des participants.

Figure 3.4 Étapes de préparation réalisées sur les enregistrements électroencéphalographiques bruts de chacun des participants



3.6 Le plan d'analyse des résultats

Toutes les analyses statistiques ont été réalisées avec le logiciel SPSS (IBM Corp., 2021). La première question de recherche visait à vérifier l'effet des facteurs linguistiques sur la précision, le temps de réaction ainsi que l'amplitude et la latence de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus chez des élèves du 3^e cycle du primaire durant une tâche de décision lexicale. Quatre variables linguistiques sont manipulées durant la tâche de décision lexicale : la lexicalité (mots, pseudomots/non-mots), la structure morphologique des items (polymorphémique, monomorphémique), le type de racine (libre, liée) et l'origine des racines (latine, grecque).

Les résultats obtenus à la tâche de décision lexicale et les données extraites de l'enregistrement électroencéphalographique capté lors de cette même tâche permettent de répondre à une partie de la première question de recherche. Il convient de rappeler les huit conditions expérimentales de la tâche de décision lexicale : 1) mots composés formés de racines latines, 2) mots composés formés de racines grecques, 3) mots composés formés de racines libres, 4) mots monomorphémiques, 5) pseudomots formés de racines latines, 6) pseudomots formés de racines grecques, 7) pseudomots formés de racines

libres, et 8) non-mots. Ces huit conditions expérimentales ont été comparées quant au temps de réaction et à la précision des réponses, d'une part, puis, d'autre part, quant à l'amplitude et à la latence des potentiels évoqués à deux fenêtres temporelles poststimulus : de 200 ms à 350 ms et de 350 ms à 500 ms. L'activation cérébrale détectée dans la fenêtre temporelle de 200 ms à 350 ms permettrait de détecter les processus de décomposition morphologique habituellement attribués au marqueur de potentiel évoqué N250. Pour la fenêtre temporelle de 350 ms à 500 ms associée au marqueur de potentiel évoqué N400, une activation renseignerait sur la récupération du sens du mot, l'accès lexical. Les six régions topographiques ciblées (antérieure gauche, antérieure médiane, antérieure droite, postérieure gauche, postérieure médiane et postérieure droite) ont aussi été insérées dans les analyses statistiques. Les interactions entre les variables indépendantes (lexicalité, structure morphologique des items, type de racine et origine des racines) ainsi que les régions topographiques d'intérêt permettront de vérifier les effets de la lexicalité, de la structure morphologique des items, du type de racine et de l'origine des racines sur l'identification des mots écrits. L'interaction avec les régions topographiques est nécessaire afin de permettre l'identification et l'interprétation des potentiels évoqués.

Ainsi, de multiples analyses de variance (ANOVAs) à mesures répétées ont été effectuées lors desquelles un ajustement a permis de compenser pour les comparaisons multiples des effets simples, l'ajustement Sidak des intervalles de confiance. Aussi, lorsque le test de sphéricité était significatif, les corrections nécessaires ont été apportées afin de rapporter les résultats, soit la correction Greenhouse-Geisser lorsque l'épsilon était inférieur à 0,75 et Huynh-Feldt lorsque l'épsilon était supérieur à 0,75 (Field, 2013). En plus de la statistique F , des degrés de liberté et de la valeur p , la taille de l'effet est rapportée grâce à l'éta-carré partiel (η^2p) ou au d de Cohen (1988). Conformément aux suggestions de Cohen (1988), des éta-carrés partiels de 0,14, 0,06 et 0,01 représentent respectivement un grand, un moyen et un petit effet. Pour le d de Cohen, un grand effet est supérieur à 0,80, un moyen effet est supérieur à 0,50 et un petit effet est supérieur à 0,20 (Cohen, 1988).

Pour mesurer l'effet de la lexicalité, les conditions expérimentales attribuées aux mots (1, 2, 3, 4) ont été comparées aux pseudomots et aux non-mots (5, 6, 7, 8). D'abord, deux ANOVAs à mesures répétées à un seul facteur intrasujet (2; lexicalité) ont été réalisées avec les données comportementales (temps de réaction et précision des réponses à la tâche de décision lexicale). Puis, pour vérifier l'effet de la lexicalité sur l'amplitude et la latence de l'activité cérébrale à chacune des deux fenêtres temporelles, quatre ANOVAs à mesures répétées à deux facteurs intrasujets (2×6 ; lexicalité \times régions) ont été effectuées.

Pour valider l'hypothèse statistique alternative, la lexicalité doit avoir un effet significatif sur le temps de réaction et la précision des réponses à la tâche de décision lexicale, entraînant des temps de réponse plus courts et des réponses plus précises pour les mots que les pseudomots et les non-mots. Aussi, une négativité plus grande serait observée pour les pseudomots et les non-mots aux deux fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms. Étant donné que plusieurs recherches antérieures confirment l'effet de la lexicalité, la lexicalité a été ajoutée aux analyses subséquentes. Ainsi, un résultat significatif de l'interaction de la lexicalité avec une des trois autres variables (structure morphologique des items, type de racine, origine des racines) permettra de faire une analyse différenciée selon la lexicalité. Advenant que la lexicalité produise un effet important sur les variables dépendantes et qu'elle ne soit pas introduite dans les analyses subséquentes, il est possible que les effets des autres variables linguistiques sur les variables dépendantes soient effacés.

Pour mesurer l'effet de la structure morphologique des items, le traitement des mots polymorphémiques (1, 2, 3), des mots monomorphémiques (4), des pseudomots polymorphémiques (5, 6, 7) et des non-mots (8) a été comparé. Pour ce faire, deux ANOVAs à mesures répétées à deux facteurs intrasujets (2×2 ; lexicalité \times structure) avec les données comportementales (temps de réaction et précision des réponses à la tâche de décision lexicale) ont été effectuées. Puis, pour vérifier l'effet de la structure morphologique des items sur l'amplitude et la latence de l'activité cérébrale à chacune des deux fenêtres temporelles, quatre ANOVAs à mesures répétées à trois facteurs intrasujets ($2 \times 2 \times 6$; lexicalité \times structure \times régions) ont été réalisées. Selon les résultats des études antérieures, l'une ou l'autre des hypothèses alternatives pourraient être validées si l'effet de structure est significatif quant aux données comportementales. Si tel est le cas, l'hypothèse validerait que les participants fournissent une réponse plus rapide et plus précise lors de la décision lexicale des items monomorphémiques par rapport à celle des items polymorphémiques ou l'inverse, c'est-à-dire que les participants fournissent une réponse plus rapide et plus précise aux items polymorphémiques qu'aux items monomorphémiques. De plus, quant aux données psychophysiques, l'hypothèse alternative sera validée si les items polymorphémiques entraînent une plus grande négativité de 200 ms à 350 ms et, pour les pseudomots et les non-mots, de 350 ms à 500 ms.

Pour mesurer l'effet du type de racine, les items ont été regroupés de façon à obtenir six catégories : les mots formés de deux racines liées (1, 2), les mots formés de deux racines libres (3), les mots monomorphémiques (4), les pseudomots formés de deux racines liées (5, 6), les pseudomots formés de

deux racines libres (7) ainsi que les non-mots (8). Les mots monomorphémiques et les non-mots sont, ici, des items de contrôle. L'effet du type de racine sur les résultats comportementaux de temps de réaction et de précision des réponses à la tâche de décision lexicale a été mesuré grâce à deux ANOVAs à mesures répétées à deux facteurs intrasujets (2×3 ; lexicalité \times type de racine). Puis, quatre ANOVAs à mesures répétées à trois facteurs intrasujets ($2 \times 3 \times 6$; lexicalité \times type de racine \times régions) ont permis de vérifier l'effet du type de racine sur l'amplitude et la latence de l'activité cérébrale de 200 ms à 350 ms et de 350 ms à 500 ms. Pour valider l'hypothèse alternative sur le type de racine dans les mots composés, les items formés de racines liées devront entraîner des réponses significativement plus rapides et précises que les items formés de racines libres en plus d'être associés à de plus faibles négativités de 350 ms à 500 ms.

Enfin, l'effet de l'origine des racines a été mesuré en ne considérant que les conditions expérimentales concernées : mots formés de racines latines (1), mots formés de racines grecques (2), pseudomots formés de racines latines (5) et pseudomots formés de racines grecques (6). En suivant le même principe que pour les analyses précédentes, deux ANOVAs à mesures répétées à deux facteurs intrasujets (2×2 ; lexicalité \times origine) ont d'abord été effectuées avec les données comportementales de temps de réaction et de précision des réponses à la tâche de décision lexicale. Puis, l'effet de l'origine des racines sur l'amplitude et la latence de l'activité cérébrale à chacune des deux fenêtres temporelles a été mesurée en effectuant quatre ANOVAs à mesures répétées à trois facteurs intrasujets ($2 \times 2 \times 6$; lexicalité \times origine \times régions). Des différences entre les items formés de racines latines et grecques aux niveaux comportemental et psychophysique permettront de valider l'hypothèse alternative, et ce peu importe la direction de ces différences.

Pour continuer, la première question de recherche visait aussi à mesurer les effets de la fréquence du mot entier et de la fréquence des racines sur le traitement des mots composés formés de racines latines et grecques chez les élèves du 3^e cycle du primaire. Enfin, la deuxième question de recherche visait à identifier les facteurs individuels tels que la connaissance des racines latines et grecques, la connaissance du vocabulaire disciplinaire ainsi que des habiletés morphologiques qui pourraient influencer le traitement des mots composés formés de racines latines et grecques. Ces variables n'ont pas été manipulées lors de la tâche de décision lexicale. Comme les données sont nichées par participant, des analyses de régression multiniveaux ont été effectuées pour chacune des six variables dépendantes :

précision des réponses, temps de réaction, amplitude de 200 ms à 350 ms, latence de 200 ms à 350 ms, amplitude de 350 ms à 500 ms et latence de 350 ms à 500 ms.

Pour vérifier l'influence de ces variables sur la précision des réponses lors de la tâche de décision lexicale (variable binaire), deux régressions logistiques multiniveaux ont été réalisées, une pour les mots et l'autre pour les pseudomots et les non-mots. La régression logistique multiniveaux permet d'établir la probabilité d'obtenir une bonne réponse pour chaque item de la tâche de décision lexicale selon la valeur des variables indépendantes entrées dans le modèle de régression tout en considérant l'interdépendance des observations pour chacun des participants. Dans l'équation, les quatre conditions expérimentales (items formés de deux racines latines [1 ou 5], items formés de deux racines grecques [2 ou 6], items formés de deux racines libres [3 ou 7] et items de contrôle [4 ou 8]), le niveau scolaire des participants (5^e ou 6^e année), les trois indices de fréquence (fréquence du mot entier, fréquence de la 1^{re} racine, fréquence de la 2^e racine) ainsi que les cinq facteurs individuels (connaissance du vocabulaire, reconnaissance des racines, décomposition morphologique, connaissance du sens des racines latines et grecques et analyse morphologique) ont été insérés. Pour les mots monomorphémiques, le seul indice de fréquence entré dans l'équation est la fréquence du mot entier. Les indices de fréquence des racines ne s'appliquent pas dans ce contexte. Pour les pseudomots, au contraire, l'indice de fréquence du mot entier, étant inexistant, n'a pas été entré dans l'équation alors que les indices de fréquence de chacune des racines y sont demeurés. Enfin, pour les non-mots, aucun indice de fréquence n'était disponible.

Enfin, des régressions linéaires multiniveaux ont été effectuées pour vérifier l'influence des facteurs linguistiques de fréquence et des facteurs individuels sur les variables dépendantes continues : temps de réaction, amplitude de 200 ms à 350 ms, latence de 200 ms à 350 ms, amplitude de 350 ms à 500 ms et latence de 350 ms à 500 ms. Les mêmes variables que pour les régressions logistiques multiniveaux ont été insérées dans les régressions, autant pour les mots que pour les pseudomots et les non-mots. Pour les quatre régressions linéaires multiniveaux sur l'amplitude et la latence des potentiels évoqués, les régions topographiques ont aussi été insérées comme variable. L'hypothèse alternative sur les facteurs linguistiques de fréquence sera validée si la fréquence du mot entier et la fréquence des racines est associée à de meilleurs résultats en temps de réaction et en précision des réponses à la tâche de décision lexicale et à des variances d'amplitude de 200 ms à 350 ms et de 350 ms à 500 ms. Concernant les facteurs individuels, l'hypothèse alternative sera validée statistiquement si la connaissance du vocabulaire, la reconnaissance des racines, la décomposition morphologique, la connaissance du sens

des racines latines et grecques et l'analyse morphologique sont positivement liées à la précision en lecture et négativement liées au temps de réaction lors de la tâche de décision lexicale. Aussi, une relation positive entre la connaissance du vocabulaire, la reconnaissance des racines, les habiletés de décomposition morphologique, la connaissance du sens des racines latines et grecques, les habiletés d'analyse morphologique ainsi que l'amplitude de 200 ms à 350 ms permettrait de valider l'hypothèse alternative.

CHAPITRE 4

RÉSULTATS

D'abord, les statistiques descriptives des résultats de la tâche de décision lexicale (4.1) et du test dynamique (4.2) sont rapportées. Pour répondre à la première question de recherche, il a été vérifié si différentes variables linguistiques indépendantes, soit la lexicalité, la structure morphologique des items, le type de racine, l'origine des racines, la fréquence du mot entier et la fréquence des racines, ont un effet sur la précision et le temps de réaction des participants à la tâche de décision lexicale ainsi que sur l'amplitude maximale et la latence des potentiels évoqués aux fenêtres temporelles de 200 ms à 350 ms et de 350 ms à 500 ms susceptibles de détecter les processus cognitifs visés, comme la décomposition morpho-orthographique des mots polymorphémiques ainsi que l'accès sémantique. Les résultats de chacun de ces effets sont présentés successivement, (4.3). Puis, pour répondre à la deuxième question de recherche, l'influence des facteurs individuels sur le plan de la connaissance du vocabulaire, de la connaissance du sens de racines latines ou grecques et de certaines habiletés morphologiques est rapportée (4.4). Enfin, un résumé des principaux résultats de la recherche clôt cette section (4.5).

4.1 Les statistiques descriptives des résultats de la tâche de décision lexicale

Avant de procéder à l'analyse des données en lien avec les effets de la lexicalité, de la structure morphologique des items, du type de racine et de l'origine des racines, il convient de vérifier si les variables dépendantes étaient influencées par le niveau scolaire des participants. En effet, des élèves de 5^e année et de 6^e année du primaire ont été recrutés. Il s'est avéré que le niveau scolaire n'a aucun effet sur le temps de réaction, $t(27) = 1,589$, $p = 0,124$, sur l'amplitude de 200 ms à 350 ms, $t(27) = -1,579$, $p = 0,126$, sur la latence de 200 ms à 350 ms, $t(27) = 1,464$, $p = 0,155$, ainsi que sur la latence de 350 ms à 500 ms, $t(27) = -0,842$, $p = 0,407$. Cependant, les participants de 6^e année du primaire sont plus précis, donc réalisent moins d'erreurs, à la tâche de décision lexicale que leurs pairs de 5^e année, $t(27) = -2,090$, $p = 0,046$. Aussi, la négativité de 350 ms à 500 ms est plus importante chez les participants de 5^e année que ceux de 6^e année, $t(27) = -2,191$, $p = 0,037$. Par manque de puissance statistique, les analyses subséquentes (analyses de variance à mesures répétées) ne peuvent pas être différenciées selon le niveau scolaire (section 4.3). Toutefois, le niveau scolaire est inséré dans les analyses de régression logistique et linéaire multiniveaux (sections 4.4 et 4.5).

Le Tableau 4.1 résume les statistiques descriptives relatives à la précision des élèves durant la tâche de décision lexicale, soit leur pourcentage de bonnes réponses. La moyenne, l'écart type et l'intervalle de confiance à 95 % ont été calculés pour chaque catégorie d'items. En moyenne, les participants ont été en mesure de répondre correctement à 78,68 % ($ÉT = 6,40\%$) des items de la tâche de décision lexicale. Les items les moins bien réussis sont les mots polymorphémiques formés de deux racines latines ($M = 62,41\%$, $ÉT = 11,82\%$) ainsi que les mots polymorphémiques formés de deux racines libres ($M = 63,33\%$, $ÉT = 17,02\%$). Les items les mieux réussis sont les pseudomots polymorphémiques formés de deux racines libres ($M = 90,80\%$, $ÉT = 8,58\%$) et les non-mots ($M = 93,10\%$, $ÉT = 8,82\%$). À travers les différentes catégories d'items, l'écart type varie de 8,54 % à 17,02 % indiquant une grande variabilité d'un participant à l'autre, mais aussi d'une catégorie d'items à l'autre, relativement à la précision au test de décision lexicale.

Tableau 4.1 Pourcentage moyen, écart type et intervalle de confiance des bonnes réponses pour chaque catégorie d'items à la tâche de décision lexicale

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines latines	62,41	11,82	[57,92; 66,91]
(2) Mots polymorphémiques formés de deux racines grecques	68,28	15,34	[62,44; 74,11]
(3) Mots polymorphémiques formés de deux racines libres	63,33	17,02	[56,86; 69,81]
(4) Mots monomorphémiques	84,25	8,54	[81,00; 87,50]
(5) Pseudomots polymorphémiques formés de deux racines latines	87,47	10,38	[83,52; 91,42]
(6) Pseudomots polymorphémiques formés de deux racines grecques	79,77	12,21	[75,12; 84,42]
(7) Pseudomots polymorphémiques formés de deux racines libres	90,80	8,58	[87,54; 94,07]
(8) Non-mots	93,10	8,82	[89,75; 96,46]
Moyenne générale	78,68	6,40	[76,24; 81,11]

Note. IC 95 % = intervalle de confiance à 95 %.

La moyenne, l'écart type et l'intervalle de confiance à 95 % ont aussi été calculés pour le temps de réaction des élèves à la tâche de décision lexicale, uniquement pour les réponses correctes. Le Tableau 4.2 présente ces statistiques descriptives pour chaque catégorie d'items. En moyenne, les participants fournissent une réponse correcte aux items de la tâche de décision lexicale après 1422,70 ms

($\acute{E}T = 405,91$ ms). Les élèves fournissent une décision lexicale plus rapidement aux mots monomorphémiques et aux mots polymorphémiques formés de deux racines latines (respectivement 1115,40 ms et 1281,93 ms). Les trois catégories de pseudomots sont celles qui entraînent les réponses les plus lentes, soit 1681,34 ms pour ceux formés de deux racines grecques, 1553,43 ms pour ceux formés de deux racines libres et 1541,78 ms pour ceux formés de deux racines latines.

Tableau 4.2 Temps de réaction moyen (ms) pour chaque catégorie d'items réussis à la tâche de décision lexicale

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines latines	1281,93	407,28	[1192,04; 1371,83]
(2) Mots polymorphémiques formés de deux racines grecques	1406,37	432,38	[1304,86; 1507,88]
(3) Mots polymorphémiques formés de deux racines libres	1367,70	447,93	[1249,32; 1486,08]
(4) Mots monomorphémiques	1115,40	331,83	[1036,29; 1194,51]
(5) Pseudomots polymorphémiques formés de deux racines latines	1541,78	414,03	[1419,94; 1663,62]
(6) Pseudomots polymorphémiques formés de deux racines grecques	1681,34	433,50	[1553,44; 1809,24]
(7) Pseudomots polymorphémiques formés de deux racines libres	1553,43	409,99	[1432,00; 1674,87]
(8) Non-mots	1433,63	370,32	[1310,16; 1557,11]
Moyenne générale	1422,70	405,91	[1321,82; 1523,58]

Note. IC 95 % = intervalle de confiance à 95 %.

Conformément à la procédure décrite à la section 3.5, l'amplitude négative maximale et la latence des potentiels évoqués identifiés aux fenêtres temporelles de 200 ms à 350 ms et de 350 ms à 500 ms suivant la présentation de chaque stimulus réussi lors de la tâche de décision lexicale ont été récupérées. Les statistiques descriptives sont disponibles à l'annexe G.

4.2 Les statistiques descriptives des résultats du test dynamique

Durant le test dynamique, différentes habiletés ont été évaluées soit la connaissance du vocabulaire, l'habileté à reconnaître des racines latines ou grecques dans un mot ou dans un pseudomot, la capacité à décomposer un mot ou un pseudomot bimorphémique en ses deux morphèmes, la connaissance du sens des racines latines ou grecques ainsi que l'habileté à produire une analyse morphologique en se servant

du sens des racines pour définir le mot ou le pseudomot. Les statistiques descriptives de chacune de ces habiletés sont rapportées dans le Tableau 4.3.

Tableau 4.3 Moyennes, écarts types et intervalles de confiance des habiletés mesurées au test dynamique

Variable	Mots			Pseudomots		
	<i>M</i>	<i>ÉT</i>	IC 95 %	<i>M</i>	<i>ÉT</i>	IC 95 %
Connaissance du vocabulaire (max. 20)	10,07	3,62	[8,69; 11,44]	—	—	—
Reconnaissance des racines (max. 20)	10,83	3,86	[9,36; 12,30]	14,31	2,88	[13,21; 15,41]
Latines (max. 10)	4,41	2,16	[3,59; 5,24]	7,79	1,57	[7,20; 8,39]
Grecques (max. 10)	6,41	2,29	[5,54; 7,29]	6,52	1,74	[5,85; 7,18]
Décomposition morphologique (max. 10)	8,37	1,54	[7,79; 8,97]	8,90	0,72	[8,62; 9,17]
Sens des racines (max. 60)	23,66	10,14	[19,80; 27,51]	31,34	7,05	[28,66; 34,03]
Latines (max. 30)	10,52	5,64	[8,37; 12,66]	18,31	3,66	[16,92; 19,70]
Grecques (max. 30)	13,14	5,42	[11,08; 15,20]	13,03	4,09	[11,48; 14,59]
Analyse morphologique (max. 20)	14,76	2,53	[13,80; 15,72]	18,90	1,21	[18,44; 19,36]

Note. IC 95 % = intervalle de confiance à 95 %; — = non disponible.

D’abord, les résultats révèlent que les élèves du 3^e cycle du primaire ne connaissent que partiellement les mots du vocabulaire disciplinaire. Le score moyen visant à évaluer la capacité des élèves à fournir une définition précise ou partielle du mot s’élève à 10,07 sur un maximum de 20 points. Évidemment, cette habileté n’a pas été évaluée pour les pseudomots.

Ensuite, les élèves reconnaissent un peu plus de la moitié des racines latines et grecques présentées dans les mots ($M = 10,83/20$; 54,15 %) et dans les pseudomots ($M = 14,31/20$; 71,55 %). Pour les mots, les participants reconnaissent en moyenne 6,41/10 racines grecques et 4,41/10 racines latines. Pour les pseudomots, ils reconnaissent 6,52/10 racines grecques et 7,79/10 racines latines. Qu’ils reconnaissent ou non les racines dans les items présentés, les élèves sont majoritairement en mesure de décomposer adéquatement un mot bimorphémique ($M = 8,37/10$; 83,70 %) ou un pseudomot ($M = 8,90/10$; 89,00 %) en morphèmes.

Aussi, les élèves de 3^e cycle du primaire connaissent peu le sens des racines latines et grecques. Sur un score maximal de 60 points, les élèves atteignent 23,66 pour les mots et 31,34 pour les pseudomots, ce qui représente respectivement des taux de réussite de 39,43 % et de 52,23 %. Au sein des mots, cette connaissance du sens des racines semble plus élevée pour les racines grecques que latines alors que le profil inverse se dessinerait pour les pseudomots.

Enfin, lorsqu'ils connaissent le sens de chacune des racines d'un mot, les élèves semblent mieux réussir l'analyse morphologique sur des pseudomots ($M = 18,90/20$; 94,5 %) que sur des mots ($M = 14,76/20$; 73,85 %), probablement parce qu'avec des pseudomots insérés dans des phrases acontextuelles, les seuls indices disponibles demeurent le sens des racines latines ou grecques alors que, dans des mots, les connaissances antérieures du vocabulaire des élèves influencent l'analyse morphologique.

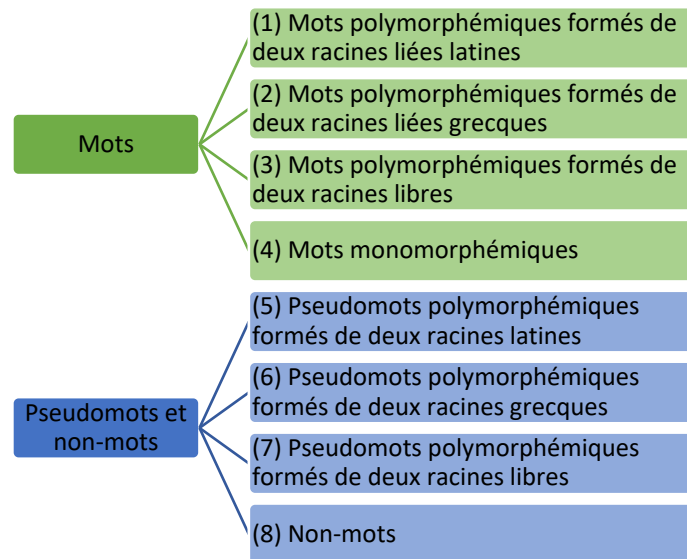
4.3 L'influence des facteurs linguistiques sur le traitement des mots écrits

Durant la tâche de décision lexicale, quatre variables linguistiques ont été manipulées : la lexicalité (4.3.1), la structure morphologique des items (4.3.2), le type de racine (4.3.3) et l'origine des racines (4.3.4). L'effet de chacune de ces variables est présenté dans les sections suivantes. La fréquence du mot entier et la fréquence des racines sont aussi des facteurs linguistiques susceptibles d'influencer le traitement des mots écrits. Ces variables n'ont pas été manipulées lors de la sélection des items pour la tâche de décision lexicale, mais elles ont été mesurées post hoc. Leur effet peut être considérable sur le traitement des mots écrits. Les résultats des effets de fréquence, autant ceux attribuables au mot entier que la fréquence de chacune des racines, sont présentés (4.3.5).

4.3.1 L'effet de la lexicalité

La Figure 4.1 schématise les groupements effectués pour vérifier l'effet de la lexicalité. Pour vérifier la différence de traitement entre les mots ainsi que les pseudomots et les non-mots, les quatre premières catégories d'items qui représentent les mots ont été regroupées : (1), (2), (3) et (4); ainsi que les quatre dernières composées de pseudomots et de non-mots : (5), (6), (7) et (8).

Figure 4.1 Groupements des catégories d'items pour vérifier l'effet de la lexicalité



Les moyennes et écarts types de la précision à la tâche de décision lexicale pour les mots ainsi que les pseudomots et les non-mots sont présentés dans le Tableau 4.4. Les résultats de l'ANOVA à mesures répétées avec une variable à deux niveaux (lexicalité) afin de mesurer l'effet de la lexicalité sur la précision sont présentés dans le Tableau 4.5. D'abord, ces résultats ont permis de confirmer que, en moyenne, les participants répondent avec plus d'exactitude aux pseudomots et aux non-mots ($M = 87,79\%$, $ÉT = 8,25\%$) qu'aux mots ($M = 69,57\%$, $ÉT = 11,08\%$). Cette différence est significative et représente un grand effet, $F(1, 28) = 44,24$, $p < 0,001$, $\eta^2p = 0,61$.

Tableau 4.4 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon la lexicalité

Lexicalité	<i>M</i>	<i>ÉT</i>
Mots	69,57	11,08
Pseudomots et non-mots	87,79	8,25

Note. $n = 29$.

Tableau 4.5 Résultats de l'ANOVA à mesures répétées pour l'effet de la lexicalité sur la précision (%) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	44,24	1, 28	< 0,001	0,61

Note. $n = 29$.

Concernant le temps de réaction à la tâche de décision lexicale, les moyennes et les écarts types pour les mots ainsi que les pseudomots et les non-mots sont présentés dans le Tableau 4.6 tandis que le Tableau 4.7 présente les résultats de l'ANOVA à mesures répétées pour vérifier l'effet de la lexicalité sur le temps de réaction. Les résultats montrent que la lexicalité influence le temps de réaction des participants à la tâche de décision lexicale. Les participants répondent plus rapidement aux mots ($M = 1292,85$ ms, $ÉT = 236,68$ ms) qu'aux pseudomots et aux non-mots ($M = 1552,55$ ms, $ÉT = 312,07$ ms). Encore une fois, cette différence est significative à $p < 0,001$, $F(1, 28) = 76,81$. Cela représente une grande taille d'effet, $\eta^2_p = 0,73$.

Tableau 4.6 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon la lexicalité

Lexicalité	<i>M</i>	<i>ÉT</i>
Mots	1292,85	236,68
Pseudomots et non-mots	1552,55	312,07

Note. $n = 29$.

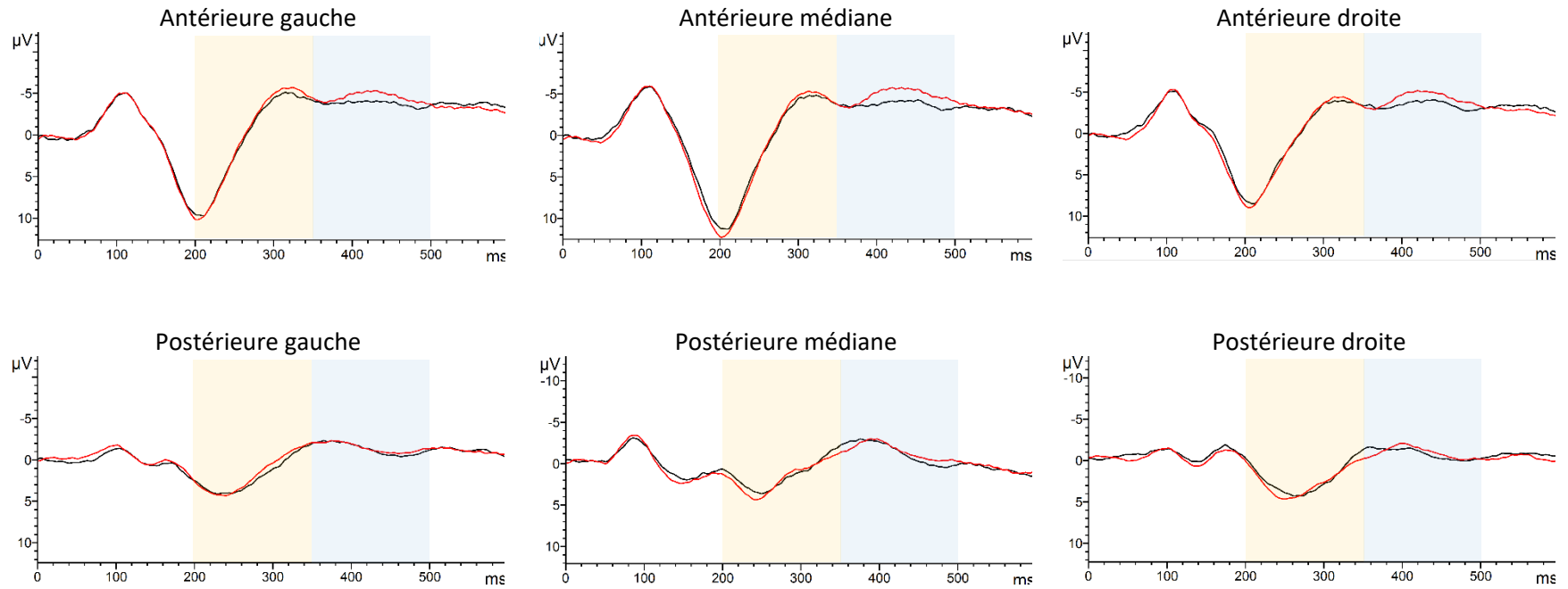
Tableau 4.7 Résultats de l'ANOVA à mesures répétées pour l'effet de la lexicalité sur le temps de réaction (ms) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2_p
Lexicalité	76,81	1, 28	< 0,001	0,73

Note. $n = 29$.

La Figure 4.2 présente les tracés des potentiels évoqués aux six régions d'intérêt en fonction de la lexicalité. L'inspection visuelle permet de constater une forte amplitude du marqueur N1-P2, particulièrement en régions antérieures. Cette signature est suivie de deux pics négatifs, plus visibles en régions antérieures que postérieures et pour les pseudomots et les non-mots que pour les mots. Le premier pic atteint son maximum vers 300 ms alors que le second l'atteint autour de 420 ms à 450 ms. Toujours en régions antérieures, les pseudomots et les non-mots semblent entraîner une plus grande négativité dans la fenêtre temporelle de 350 ms à 500 ms poststimulus.

Figure 4.2 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la lexicalité



Note. Tracés des potentiels évoqués pour les mots (ligne noire) et les pseudomots et les non-mots (ligne rouge) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Les moyennes et les écarts types de l'amplitude maximale de 200 ms à 350 ms et de 350 ms à 500 ms selon la lexicalité sont présentés dans le Tableau 4.8. Pour vérifier l'effet de la lexicalité sur l'amplitude, deux ANOVAs à mesures répétées 2 × 6 (lexicalité × région) ont été effectuées, une ANOVA pour chacune des deux fenêtres temporelles (de 200 ms à 350 ms et de 350 ms à 500 ms). Les résultats sont rapportés dans le Tableau 4.9. Ici et dans les sections suivantes, nous avons fait le choix de présenter côte à côte dans le même tableau les résultats aux deux fenêtres temporelles, que ce soit pour les statistiques descriptives (moyennes et écarts types) que pour les résultats des ANOVAs. Nous pensons que cette présentation permet de mieux observer le décours temporel de la variable dépendante.

Les résultats de l'ANOVA à mesures répétées pour l'amplitude maximale à la fenêtre temporelle de 200 ms à 350 ms poststimulus montrent un effet principal de région ainsi qu'une interaction significative entre la lexicalité et les régions topographiques. La décomposition de cette interaction en effets simples permet de constater qu'en région postérieure droite uniquement, les mots entraînent une plus grande négativité (-3,17 µV) que les pseudomots et les non-mots (-2,48 µV) à $p = 0,008$, une différence qui représente un grand effet ($d = 1,06$). Dans la fenêtre temporelle de 350 ms à 500 ms, seul un effet principal de régions est présent. L'interaction entre la lexicalité et les régions topographiques n'est pas significative.

Tableau 4.8 Moyennes et écarts types de l'amplitude (µV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la lexicalité pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Antérieure gauche	-6,40	3,55	-6,84	3,29
Antérieure médiane	-6,73	3,54	-7,05	2,62
Antérieure droite	-5,49	2,80	-6,19	2,36
Postérieure gauche	-2,61	2,04	-3,88	1,86
Postérieure médiane	-3,98	2,54	-4,98	2,47
Postérieure droite	-3,17	3,25	-3,56	2,78
Pseudomots et non-mots				
Antérieure gauche	-6,86	3,62	-7,34	3,17
Antérieure médiane	-6,83	3,69	-7,63	3,03
Antérieure droite	-5,49	2,91	-6,51	2,56
Postérieure gauche	-2,95	2,04	-3,73	2,14
Postérieure médiane	-3,40	3,18	-4,66	2,47
Postérieure droite	-2,48	3,24	-3,38	2,51

Note. $n = 29$.

Tableau 4.9 Résultats des ANOVAs à mesures répétées pour l'effet de la lexicalité sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	0,06	1, 28	0,802	0,00	0,38	1, 28	0,545	0,01
Région	23,01	2,87, 80,42 ^a	< 0,001	0,45	35,52	3,53, 98,88 ^a	< 0,001	0,56
Lexicalité*Région	3,99	3,67, 102,70 ^a	0,006	0,13	2,11	3,56, 99,56 ^a	0,094	0,07

Note. *n* = 29.

^a Correction Greenhouse-Geisser.

En ce qui concerne la latence, les moyennes et écarts types selon la lexicalité pour chacune des fenêtres temporelles (de 200 ms à 350 ms et de 350 ms à 500 ms) sont présentés dans le Tableau 4.10. Deux analyses de variance à mesures répétées 2×6 (lexicalité \times région) ont permis de vérifier l'effet de la lexicalité sur la latence, une analyse pour chacune des deux fenêtres temporelles (de 200 ms à 350 ms et de 350 ms à 500 ms). Les résultats sont rapportés dans le Tableau 4.11. De 200 ms à 350 ms, les résultats montrent un effet principal de régions ainsi qu'une interaction significative entre la lexicalité et les régions. La décomposition de l'interaction en effets simples montre que la latence diffère selon la lexicalité seulement en région antérieure médiane où on observe une latence plus rapide pour les pseudomots et les non-mots ($M = 314,31$ ms) que pour les mots ($M = 322,69$) à $p = 0,047$, ce qui représente un grand effet ($d = 0,77$). Dans la fenêtre temporelle de 350 ms à 500 ms, ni les effets principaux ni l'interaction entre la lexicalité et les régions topographiques ne sont significatifs.

Tableau 4.10 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la lexicalité pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Antérieure gauche	321,00	17,34	424,48	49,08
Antérieure médiane	322,69	18,79	416,31	49,63
Antérieure droite	326,72	18,45	416,24	43,08
Postérieure gauche	313,93	50,37	404,76	44,93
Postérieure médiane	299,66	62,40	400,03	36,32
Postérieure droite	286,52	66,42	397,69	42,45
Pseudomots et non-mots				
Antérieure gauche	321,66	17,72	416,90	45,05
Antérieure médiane	314,31	18,79	427,72	41,27
Antérieure droite	322,14	21,58	412,79	37,15
Postérieure gauche	331,00	25,94	427,72	56,13
Postérieure médiane	289,31	56,48	419,24	43,99
Postérieure droite	270,28	62,92	412,24	38,85

Note. *n* = 29.

Tableau 4.11 Résultats des ANOVAs à mesures répétées pour l'effet de la lexicalité sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	1,14	1, 28	0,295	0,04	3,74	1, 28	0,063	0,12
Région	9,46	2,48, 69,54 ^a	< 0,001	0,25	1,67	3,11, 86,97 ^a	0,178	0,06
Lexicalité*Région	2,86	2,92, 81,82 ^a	0,043	0,09	2,34	4,42, 123,78 ^b	0,053	0,08

Note. *n* = 29.

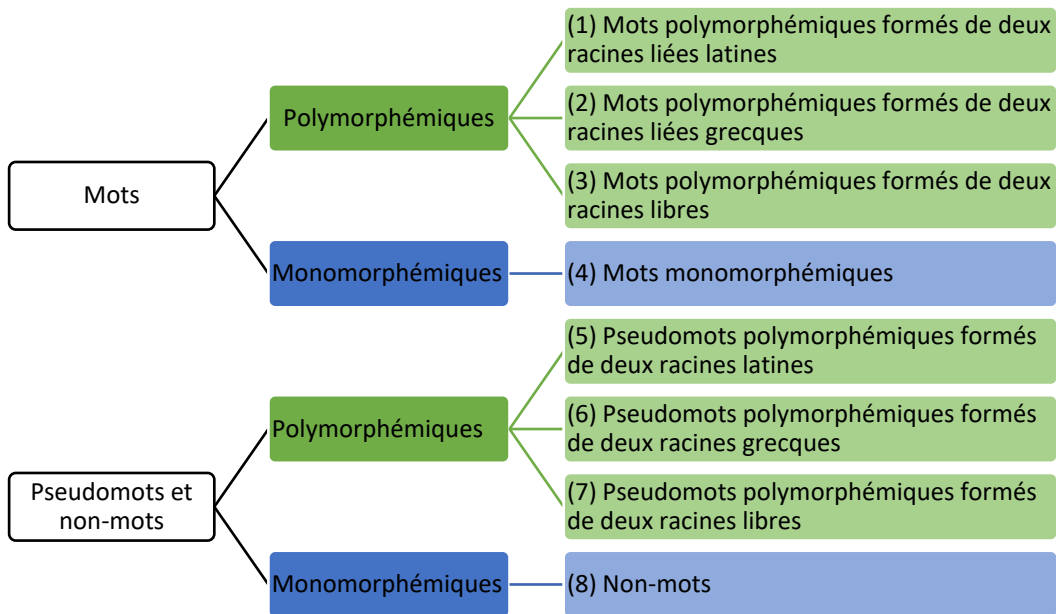
^a Correction Greenhouse-Geisser. ^b Correction Huynh-Feldt.

En résumé, la lexicalité influence grandement les données comportementales de précision et temps de réaction prises lors de la tâche de décision lexicale. En moyenne, les participants réussissent mieux les pseudomots et les non-mots que les mots, mais y répondent plus lentement. Concernant les mesures psychophysiques, la lexicalité influence l'amplitude générée lors de l'identification des mots écrits et sa latence selon les régions topographiques, mais seulement de 200 ms à 350 ms. Dans cette fenêtre temporelle, les mots génèrent une plus grande négativité en région postérieure droite que les pseudomots et les non-mots. Aussi, on y observe un traitement plus rapide pour les pseudomots et les non-mots en région antérieure médiane que pour les mots.

4.3.2 L'effet de la structure morphologique des items

La Figure 4.3 présente les groupements effectués pour vérifier si le traitement des items était influencé par leur structure morphologique. Pour ce faire, tout en conservant la lexicalité dans l'équation, les items polymorphémiques ont été comparés aux mots monomorphémiques et aux non-mots : (1), (2), (3), (5), (6) et (7) vs (4) et (8).

Figure 4.3 Groupements des catégories d'items pour vérifier l'effet de la structure morphologique des items



Les moyennes et les écarts types pour la précision des réponses des participants à la tâche de décision lexicale selon la structure morphologique des items sont présentés dans le Tableau 4.12. Le Tableau 4.13 présente les résultats des ANOVAs à mesures répétées 2×2 (lexicalité \times structure) qui ont été effectuées pour vérifier l'effet de la structure morphologique des items sur la précision.

Tableau 4.12 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon la structure morphologique des items

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Polymorphémiques	64,67	12,75
Monomorphémiques	84,25	8,54
Pseudomots et non-mots		
Polymorphémiques	86,02	8,73
Monomorphémiques (non-mots)	93,10	8,82

Note. $n = 29$.

Tableau 4.13 Résultats de l'ANOVA à mesures répétées pour l'effet de la structure morphologique des items sur la précision (%) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	38,20	1, 28	< 0,001	0,58
Structure	182,84	1, 28	< 0,001	0,87
Lexicalité*Structure	27,52	1, 28	< 0,001	0,50

Note. $n = 29$.

Pour la précision en lecture, les résultats de l'analyse montrent un effet principal de lexicalité et de structure. L'interaction aussi est significative. L'analyse des effets simples permet de constater que la structure monomorphémique mène à un taux de réussite plus élevé que la structure polymorphémique autant pour les mots (84,25 % vs 64,67 %, $p < 0,001$, $d = 4,06$) que les pseudomots et les non-mots (93,10 % vs 86,02 %, $p < 0,001$, $d = 2,10$), mais la différence est plus grande pour les mots.

Pour le temps de réaction à fournir une réponse correcte, les moyennes et écarts types sont présentés dans le Tableau 4.14. Le Tableau 4.15 présente les résultats de l'ANOVA à mesures répétées permettant de vérifier l'effet de la structure morphologique des mots sur le temps de réaction. Les résultats montrent un effet principal de lexicalité et de structure. L'interaction entre la lexicalité et la structure est significative. L'analyse des effets simples permet de constater que les participants fournissent une décision lexicale significativement plus rapidement pour les items qui présentent une structure monomorphémique que pour ceux qui présentent une structure polymorphémique, et ce, au sein des mots (1115,40 ms vs 1352,00 ms, $p < 0,001$, $d = 3,98$) et des pseudomots et des non-mots (1433,63 ms vs 1592,18 ms, $p < 0,001$, $d = 3,29$), avec une différence plus importante au sein des mots.

Tableau 4.14 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon la structure morphologique des items

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Polymorphémiques	1352,00	252,61
Monomorphémiques	1115,40	207,98
Pseudomots et non-mots		
Polymorphémiques	1592,78	311,53
Monomorphémiques (non-mots)	1433,63	324,62

Note. $n = 29$.

Tableau 4.15 Résultats de l'ANOVA à mesures répétées pour l'effet de la structure morphologique des items sur le temps de réaction (ms) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	82,20	1, 28	< 0,001	0,75
Structure	212,26	1, 28	< 0,001	0,88
Lexicalité*Structure	6,91	1, 28	0,014	0,20

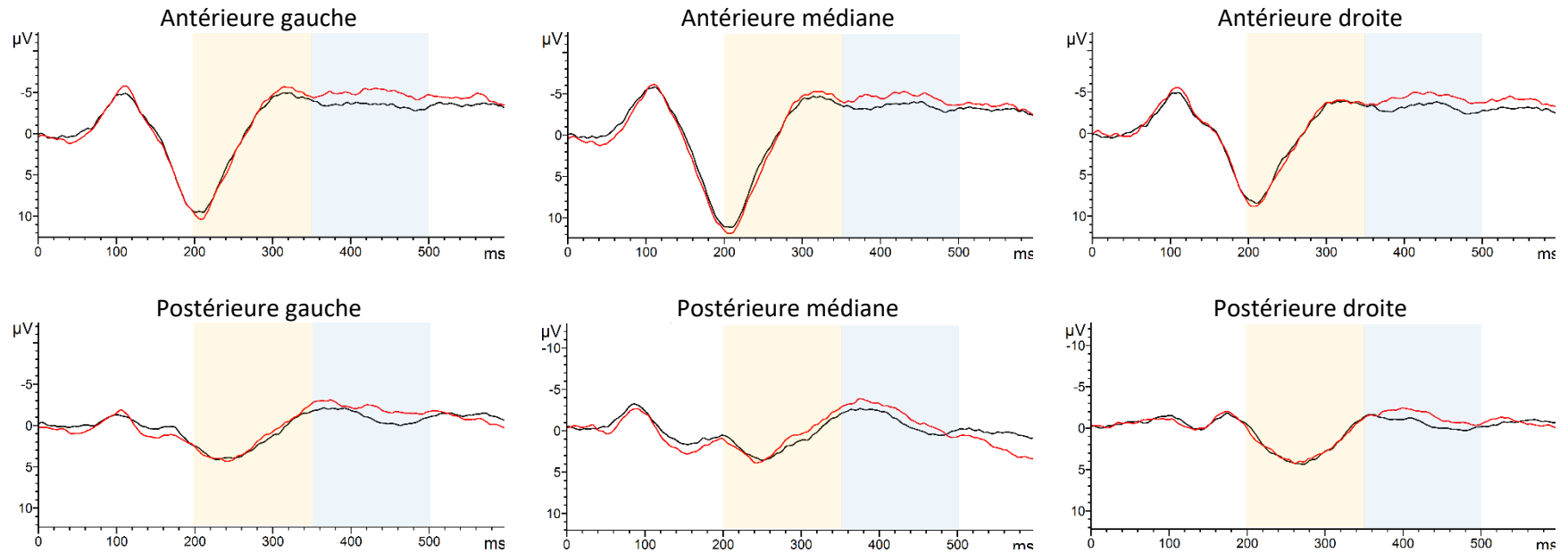
Note. $n = 29$.

Pour continuer, les données électroencéphalographiques recueillies lors de la tâche de décision lexicale sont représentées par les tracés des potentiels évoqués pour chacune des six régions d'intérêt en comparant la structure morphologique des mots (voir Figure 4.4) et des pseudomots et non-mots (Figure 4.5). Pour les mots, le marqueur visuel N1/P2 est bien visible en régions antérieures. Puis, il semble suivi d'un seul pic négatif qui atteint son maximum un peu après 300 ms, sauf pour la région antérieure droite dans laquelle il est possible de détecter deux pics distincts bien que moins prononcés, un vers 300 ms et l'autre vers 450 ms. Autant en régions antérieures que postérieures, les mots monomorphémiques semblent générer une plus grande négativité que les mots polymorphémiques. Pour les pseudomots et les non-mots, le marqueur visuel N1/P2 est aussi présent. Cependant, on distingue davantage deux pics négatifs un peu après 300 ms et 400 ms dans les trois régions antérieures. Enfin, le traitement des non-mots semble entraîner une plus grande négativité que le traitement des pseudomots dès 300 ms dans les trois régions antérieures et dans la région postérieure gauche.

Les moyennes et écarts types de l'amplitude de 200 ms à 350 ms et de 350 ms à 500 ms selon la structure morphologique des items pour chaque région d'intérêt sont présentés (Tableau 4.16). Puis, une ANOVA à mesures répétées $2 \times 2 \times 6$ (lexicalité \times structure \times région) a été effectuée pour mesurer l'effet

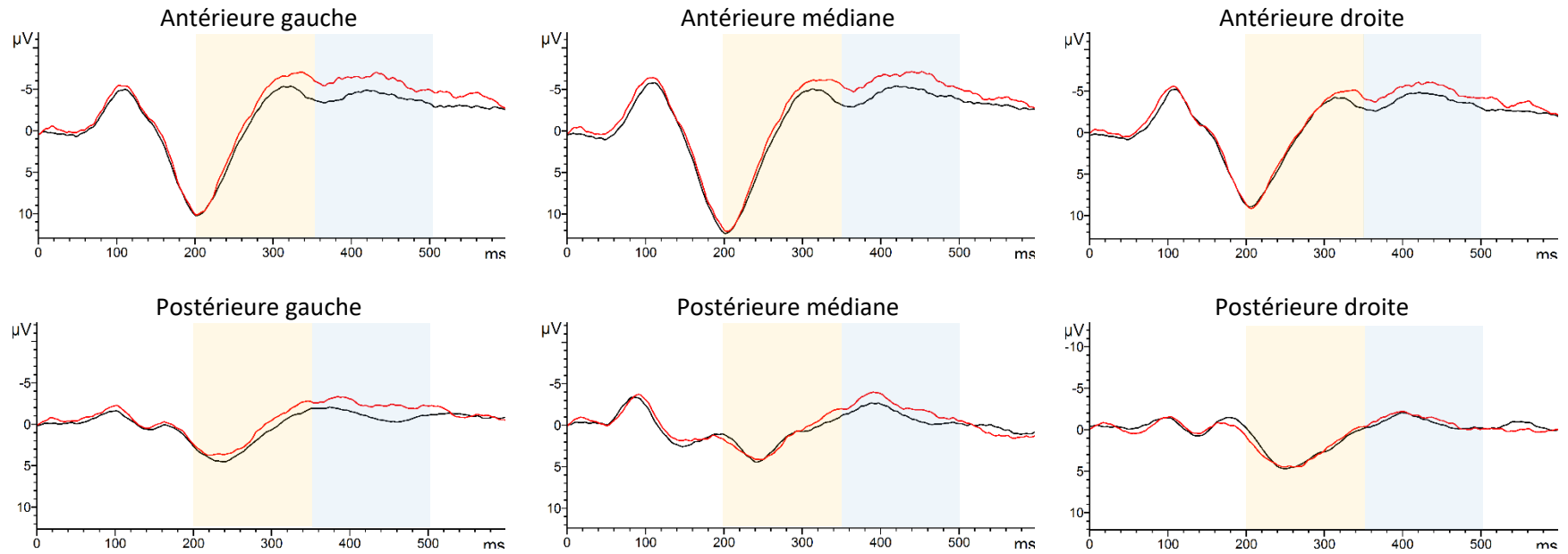
de la structure morphologique des items sur l'amplitude maximale pour chacune des deux fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms (Tableau 4.17).

Figure 4.4 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la structure morphologique des mots



Note. Tracés des potentiels évoqués pour les mots polymorphémiques (ligne noire) et les mots monomorphémiques (ligne rouge) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Figure 4.5 Tracés des potentiels évoqués aux six régions d'intérêt en fonction de la structure morphologique des pseudomots et des non-mots



Note. Tracés des potentiels évoqués pour les pseudomots polymorphémiques (ligne noire) et les non-mots (ligne rouge) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Tableau 4.16 Moyennes et écarts types de l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la structure morphologique des items pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Polymorphémiques				
Antérieure gauche	-6,48	3,88	-6,87	3,46
Antérieure médiane	-6,81	3,58	-7,17	2,91
Antérieure droite	-5,70	2,93	-6,23	2,38
Postérieure gauche	-2,53	2,00	-3,77	2,09
Postérieure médiane	-3,88	2,71	-4,98	2,74
Postérieure droite	-3,27	3,28	-3,49	2,71
Monomorphémiques				
Antérieure gauche	-7,74	3,76	-8,67	3,70
Antérieure médiane	-8,04	4,13	-8,62	3,27
Antérieure droite	-6,17	2,83	-7,64	2,75
Postérieure gauche	-3,87	2,76	-5,62	2,89
Postérieure médiane	-5,84	3,83	-6,86	4,00
Postérieure droite	-3,82	3,49	-5,35	3,18
Pseudomots et non-mots				
Polymorphémiques				
Antérieure gauche	-6,44	3,66	-7,10	3,11
Antérieure médiane	-6,47	3,61	-7,49	2,95
Antérieure droite	-5,35	2,82	-6,48	2,61
Postérieure gauche	-2,76	1,96	-3,55	2,10
Postérieure médiane	-3,33	3,30	-4,53	2,59
Postérieure droite	-2,70	3,01	-3,49	2,74
Monomorphémiques (non-mots)				
Antérieure gauche	-9,21	4,22	-9,71	3,59
Antérieure médiane	-9,30	4,72	-10,14	4,34
Antérieure droite	-7,32	3,84	-8,48	3,31
Postérieure gauche	-4,47	3,29	-5,49	3,00
Postérieure médiane	-4,48	4,38	-6,19	3,77
Postérieure droite	-3,50	3,93	-4,44	2,83

Note. $n = 29$.

Tableau 4.17 Résultats des ANOVAs à mesures répétées pour l'effet de la structure morphologique des items sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	$\eta^2\text{p}$	<i>F</i>	<i>ddl</i>	<i>p</i>	$\eta^2\text{p}$
Lexicalité	0,11	1, 28	0,743	0,00	0,36	1, 28	0,552	0,01
Structure	24,62	1, 28	< 0,001	0,47	47,66	1, 28	< 0,001	0,63
Région	28,12	2,98, 83,46 ^a	< 0,001	0,50	40,07	3,64, 101,90 ^a	< 0,001	0,59
Lexicalité*Structure	1,77	1, 28	0,194	0,06	0,23	1, 28	0,634	0,01
Lexicalité*Région	5,47	5, 140	< 0,001	0,16	4,31	3,45, 96,49 ^a	0,005	0,13
Structure*Région	2,80	5, 140	0,019	0,09	0,81	2,64, 73,86 ^a	0,480	0,03
Lexicalité*Structure*Région	3,07	3,47, 97,15 ^a	0,025	0,10	2,07	3,07, 85,86 ^a	0,109	0,07

Note. $n = 29$.

^a Correction Greenhouse-Geisser.

De 200 ms à 350 ms, pour l'amplitude maximale, des effets principaux de structure et de régions sont observés. L'interaction entre la structure morphologique des items et les régions topographiques d'intérêt est significative. La décomposition des effets simples confirme que dans toutes les régions d'intérêt, sauf la région postérieure droite ($p = 0,055$), les items monomorphémiques entraînent une plus grande négativité que les items polymorphémiques de $p = 0,006$ à $p < 0,001$ (tailles d'effet variant de $d = 1,10$ à $d = 1,94$). L'interaction entre la lexicalité, la structure morphologique des items et les régions topographiques d'intérêt est aussi significative. L'analyse des effets simples confirme que l'effet de structure morphologique des items est plus important pour les pseudomots et les non-mots que pour les mots. Les mots monomorphémiques entraînent une plus grande négativité que les mots polymorphémiques en région postérieure gauche ($-3,87 \mu\text{V}$ vs $-2,53 \mu\text{V}$, $p = 0,005$, $d = 1,12$) et postérieure médiane ($-5,84 \mu\text{V}$ vs $-3,88 \mu\text{V}$, $p = 0,004$, $d = 1,18$). Pour les pseudomots et les non-mots, les non-mots entraînent de plus grandes négativités que les pseudomots dans les régions antérieure gauche ($-9,21 \mu\text{V}$ vs $-6,44 \mu\text{V}$, $p < 0,001$, $d = 1,82$), antérieure médiane ($-9,30 \mu\text{V}$ vs $-6,47 \mu\text{V}$, $p < 0,001$, $d = 1,84$), antérieure droite ($-7,32 \mu\text{V}$ vs $-5,35 \mu\text{V}$, $p = 0,001$, $d = 1,32$) et postérieure gauche ($-4,47 \mu\text{V}$ vs $-2,76 \mu\text{V}$, $p = 0,001$, $d = 1,26$). Dans la fenêtre temporelle de 350 ms à 500 ms, l'amplitude maximale n'est pas influencée par la structure morphologique des items selon les régions d'intérêt désignées.

Pour la latence, les moyennes et les écarts types selon la structure morphologique des items sont rapportés dans le Tableau 4.18. Les résultats des ANOVAs à mesures répétées pour l'effet de la structure morphologique des items sur la latence de 200 ms à 350 ms et de 350 ms à 500 ms sont rapportés dans le Tableau 4.19. De 200 ms à 350 ms, les résultats montrent un effet principal de régions ainsi qu'une interaction entre la structure morphologique des items et les régions topographiques d'intérêt qui

atteint tout juste le seuil de significativité ($p = 0,049$). La décomposition de l'interaction en effets simples permet de constater qu'une différence de grand effet est observable seulement en région postérieure médiane ($p = 0,042$, $d = 0,79$) où l'amplitude maximale est détectée plus tardivement pour les items monomorphémiques que polymorphémiques. De 350 ms à 500 ms, l'interaction entre la structure morphologique des items ainsi que les régions pour la latence de l'amplitude maximale des potentiels évoqués est aussi significative. Cette fois, l'analyse des effets simples montre que l'amplitude maximale est détectée plus tardivement pour les items polymorphémiques (416,26 ms) que monomorphémiques (400,03 ms) en région postérieure médiane ($p = 0,049$, $d = 0,76$).

Tableau 4.18 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon la structure morphologique des items pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Polymorphémiques				
Antérieure gauche	317,76	24,12	420,69	52,40
Antérieure médiane	319,24	25,30	419,72	50,63
Antérieure droite	325,34	19,09	414,79	49,47
Postérieure gauche	313,76	49,85	412,31	47,85
Postérieure médiane	281,03	62,22	412,00	44,62
Postérieure droite	286,24	66,76	392,62	37,13
Monomorphémiques				
Antérieure gauche	313,90	24,14	415,24	45,35
Antérieure médiane	319,62	19,73	406,34	44,94
Antérieure droite	317,28	20,42	427,62	44,84
Postérieure gauche	311,55	47,30	403,48	44,78
Postérieure médiane	304,38	53,81	394,72	32,88
Postérieure droite	284,34	65,73	416,14	40,62
Pseudomots et non-mots				
Polymorphémiques				
Antérieure gauche	317,34	18,40	419,52	43,25
Antérieure médiane	318,10	18,40	420,07	39,97
Antérieure droite	321,66	17,02	415,79	38,93
Postérieure gauche	327,17	35,51	419,52	52,54
Postérieure médiane	290,52	53,72	420,52	44,60
Postérieure droite	277,31	64,50	411,34	38,54
Monomorphémiques (non-mots)				
Antérieure gauche	321,66	19,49	418,76	41,69
Antérieure médiane	320,72	23,61	425,62	39,09
Antérieure droite	317,66	28,34	424,59	42,56
Postérieure gauche	319,62	38,35	425,24	52,35
Postérieure médiane	301,14	53,46	405,34	41,48
Postérieure droite	306,48	55,94	406,21	44,81

Note. $n = 29$.

Tableau 4.19 Résultats des ANOVAs à mesures répétées pour l'effet de la structure morphologique des items sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	1,86	1, 28	0,184	0,06	2,53	1, 28	0,123	0,08
Structure	1,54	1, 28	0,225	0,05	0,04	1, 28	0,843	0,00
Région	7,57	2,48, 69,49 ^a	< 0,001	0,21	1,83	3,01, 84,27 ^a	0,149	0,06
Lexicalité*Structure	0,56	1, 28	0,460	0,02	0,06	1, 28	0,817	0,00
Lexicalité*Région	0,58	2,56, 71,71 ^a	0,605	0,02	0,71	3,55, 99,29 ^a	0,570	0,03
Structure*Région	2,83	2,74, 76,59 ^a	0,049	0,09	2,78	3,73, 104,43 ^a	0,034	0,09
Lexicalité*Structure*Région	2,41	3,04, 85,10 ^a	0,072	0,08	2,15	5, 140	0,063	0,07

Note. *n* = 29.

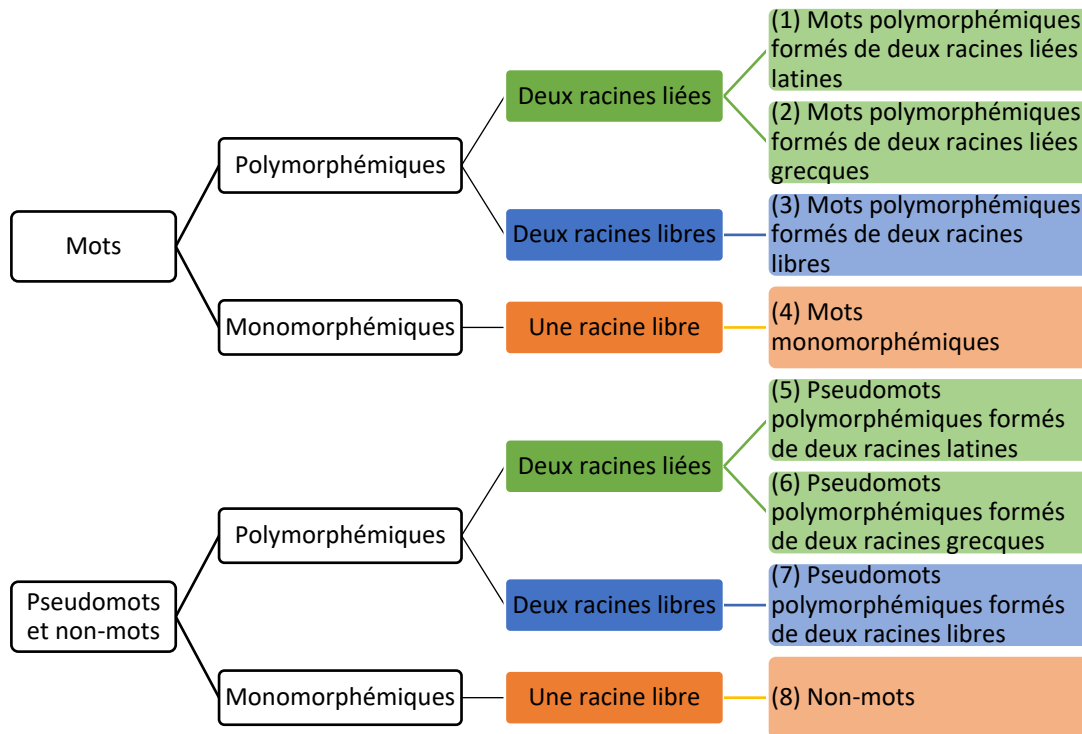
^a Correction Greenhouse-Geisser.

En résumé, les participants fournissent une décision lexicale plus rapidement et avec plus de précision pour les items monomorphémiques que pour les items polymorphémiques, et ce autant pour les mots que pour les pseudomots et les non-mots. Les données électroencéphalographiques ont permis de constater que la structure morphologique des items avait un effet sur l'amplitude maximale dans la fenêtre temporelle de 200 ms à 350 ms. La négativité générée par le traitement des items monomorphémiques est plus importante que celle générée par le traitement des items polymorphémiques, une différence assez généralisée sur le scalp. Plus particulièrement, pour les mots, l'amplitude est distincte selon la structure morphologique en régions postérieures gauche et médiane. Pour les pseudomots et les non-mots, cette différence est plutôt observable en régions antérieures. Cependant, aucun effet de la structure morphologique des items sur l'amplitude n'a été détecté de 350 ms à 500 ms. Pour la latence, la différence selon la structure morphologique est uniquement observable en région postérieure médiane, autant de 200 ms à 350 ms que de 350 ms à 500 ms. Dans la première fenêtre temporelle, le pic d'amplitude maximale est détecté plus tardivement pour les items monomorphémiques que polymorphémiques alors que la situation inverse se produit dans la deuxième fenêtre temporelle : les items polymorphémiques produisent un pic d'amplitude plus tardif que ceux monomorphémiques.

4.3.3 L'effet du type de racine

Afin de vérifier l'effet du type de racine sur le traitement des mots, les items formés de deux racines liées (1, 2, 5 et 6) ont été comparés à ceux formés de deux racines libres (3 et 7) et aux items formés sur la base d'une racine libre (4 et 8), comme illustré à la Figure 4.6.

Figure 4.6 Groupements des catégories d'items pour vérifier l'effet du type de racine



Le Tableau 4.20 présente les moyennes et écarts types de la précision des participants lors de la tâche de décision lexicale selon le type de racine. Pour observer l'effet du type de racine sur la précision, une ANOVA à mesures répétées 2×3 (lexicalité \times type de racine) a été effectuée (voir Tableau 4.21).

Tableau 4.20 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon le type de racine

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Polymorphémiques		
Deux racines liées	65,34	12,39
Deux racines libres	63,33	17,02
Monomorphémiques		
Une racine libre	84,25	8,54
Pseudomots et non-mots		
Polymorphémiques		
Deux racines liées	86,62	10,49
Deux racines libres	90,80	8,58
Monomorphémiques (non-mots)		
Une racine libre	93,10	8,81

Note. $n = 29$.

Tableau 4.21 Résultats de l'ANOVA à mesures répétées pour l'effet du type de racine sur la précision (%) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2_p
Lexicalité	47,97	1, 28	< 0,001	0,63
TypeRacine	63,16	2, 56	< 0,001	0,69
Lexicalité*TypeRacine	19,77	1,68, 47,02 ^a	< 0,001	0,41

Note. *n* = 29.

^a Correction Huynh-Feldt.

Pour la précision, les résultats indiquent un effet principal de la lexicalité et du type de racine. L'interaction entre la lexicalité et le type de racine est aussi significative. L'analyse des effets simples permet de constater un effet différencié du type de racine selon la lexicalité avec une différence plus importante au sein des mots. Pour les mots, aucune différence n'est observée entre les mots polymorphémiques formés de racines liées ($M = 65,35\%$, $ÉT = 2,3\%$) et ceux polymorphémiques formés de racines libres ($M = 63,33\%$, $ÉT = 3,16\%$) quant à la précision des réponses des participants ($p = 0,788$, $d = 0,31$). Or, la précision des participants aux mots monomorphémiques ($M = 84,25\%$, $ÉT = 1,59\%$), donc formés d'une racine libre, est significativement plus élevée que la précision aux deux autres catégories, soit les mots polymorphémiques formés de racines liées ($p < 0,001$, $d = 4,01$) et les mots polymorphémiques formés de racines libres ($p < 0,001$, $d = 2,88$). Pour les pseudomots et les non-mots, les observations sont différentes. Les réponses des participants aux pseudomots polymorphémiques formés de deux racines liées ($M = 83,62\%$, $ÉT = 1,95\%$) sont significativement moins précises que celles aux pseudomots polymorphémiques formés de racines libres ($M = 90,81\%$, $ÉT = 1,59\%$, $p = 0,002$, $d = 1,45$) et que celles aux non-mots ($M = 93,10\%$, $ÉT = 1,64\%$, $p < 0,001$, $d = 2,34$). Cependant, aucune différence significative n'est rapportée entre les pseudomots polymorphémiques formés de racines libres et les non-mots ($p = 0,390$, $d = 0,55$).

En ce qui a trait au temps de réaction à fournir une réponse, les statistiques descriptives (moyennes et écarts types) sont présentées dans le Tableau 4.22. Quant à eux, les résultats de l'ANOVA à mesures répétées permettant de vérifier l'effet du type de racine sur le temps de réaction (Tableau 4.23) sont sensiblement les mêmes que ceux pour la précision, soit un effet principal de la lexicalité, un effet principal du type de racine ainsi qu'une interaction entre ces deux variables. Encore une fois, l'analyse des effets simples montre un effet du type de racine différencié selon la lexicalité plus important au sein des mots. Pour les mots, les participants répondent aussi rapidement ($p = 0,864$, $d = 0,26$) aux mots

polymorphémiques formés de deux racines liées ($M = 1344,15$ ms, $ÉT = 44,74$ ms) qu'aux mots polymorphémiques formés de deux racines libres ($M = 1369,70$ ms, $ÉT = 57,79$ ms). Cependant, les participants répondent significativement plus rapidement aux mots monomorphémiques, donc formés d'une seule racine libre ($M = 1115,40$ ms, $ÉT = 38,62$ ms), qu'aux mots polymorphémiques des deux dernières catégories, soit ceux formés de racines liées ($p < 0,001$, $d = 3,75$) et ceux formés de racines libres ($p < 0,001$, $d = 2,73$). Pour les pseudomots et les non-mots, la tendance se maintient. Aucune différence ($p = 0,136$, $d = 0,77$) n'est observée pour le temps de réaction des participants entre les pseudomots formés de racines liées ($M = 1611,56$ ms, $ÉT = 59,37$ ms) et ceux formés de racines libres ($M = 1553,44$ ms, $ÉT = 59,28$). Cependant, les participants répondent significativement plus rapidement aux non-mots ($M = 1433,63$ ms, $ÉT = 60,28$) qu'aux pseudomots formés de racines liées ($p < 0,001$, $d = 2,79$) ou formés de racines libres ($p < 0,001$, $d = 2,34$).

Tableau 4.22 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon le type de racine

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Polymorphémiques		
Deux racines liées	1344,15	240,94
Deux racines libres	1367,70	311,21
Monomorphémiques		
Une racine libre	1115,40	207,98
Pseudomots et non-mots		
Polymorphémiques		
Deux racines liées	1611,56	319,71
Deux racines libres	1553,43	319,24
Monomorphémiques (non-mots)		
Une racine libre	1433,63	324,62

Note. $n = 29$.

Tableau 4.23 Résultats de l'ANOVA à mesures répétées pour l'effet du type de racine sur le temps de réaction (ms) à la tâche de décision lexicale

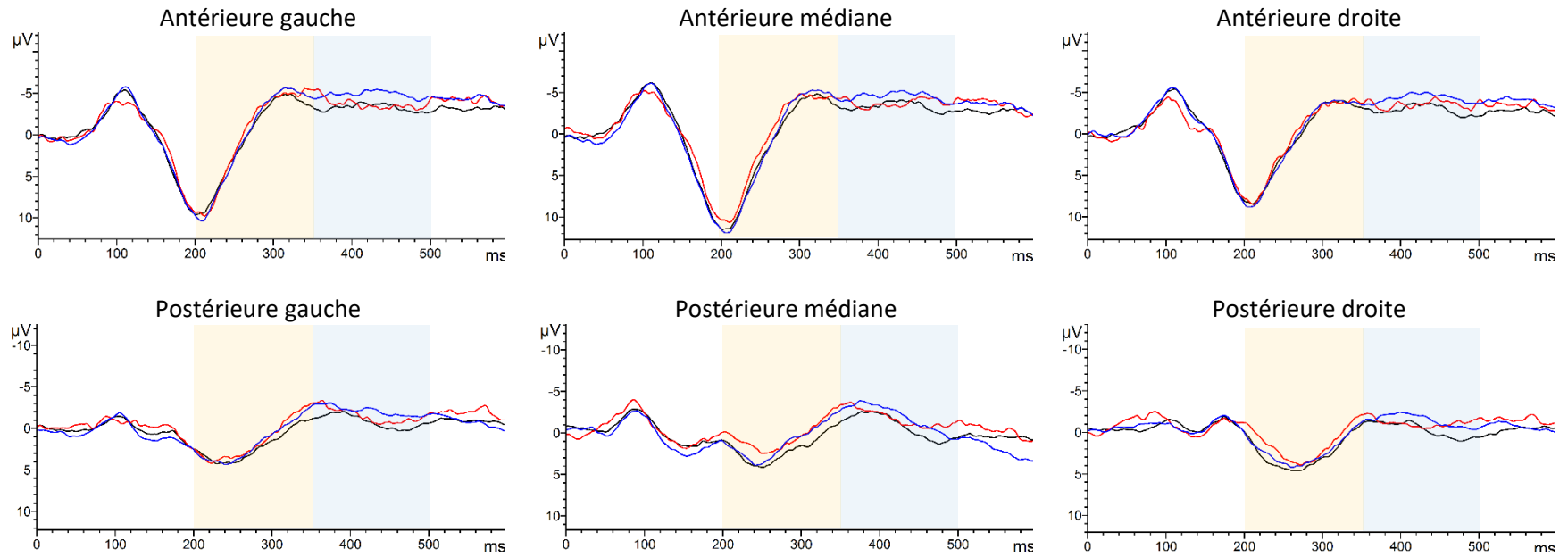
Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2_p
Lexicalité	68,77	1, 28	< 0,001	0,71
TypeRacine	68,31	2, 56	< 0,001	0,71
Lexicalité*TypeRacine	5,87	2, 56	0,005	0,17

Note. $n = 29$.

Les tracés des potentiels évoqués pour chacune des six régions d'intérêt en fonction du type de racine sont présentés à la Figure 4.7 pour les mots et à la Figure 4.8 pour les pseudomots et les non-mots. Pour les mots, l'inspection visuelle permet de constater clairement le marqueur N1/P2 en régions antérieures, avec une négativité qui semble réduite au N1 pour les racines libres. Ce marqueur est suivi d'un pic négatif en régions antérieures vers 320 ms. Pour les racines libres, ce pic atteint plutôt son apogée vers 350 ms en région antérieure gauche seulement. De 350 ms à 500 ms, le traitement des mots monomorphémiques semblent entraîner de plus grandes négativités dans toutes les régions, mais les tracés des potentiels évoqués ne permettent pas d'observer précisément les différences. Pour les pseudomots et les non-mots, le marqueur visuel N1/P2 est suivi de deux pics négatifs en régions antérieures. Le premier atteint sa négativité maximale vers 320 ms. Le second est plus aplati et semble atteindre l'amplitude maximale vers 420 ms à 460 ms. Surtout dans les régions antérieures et la région postérieure gauche, les non-mots semblent produire une plus grande négativité que les pseudomots formés de racines libres ou liées.

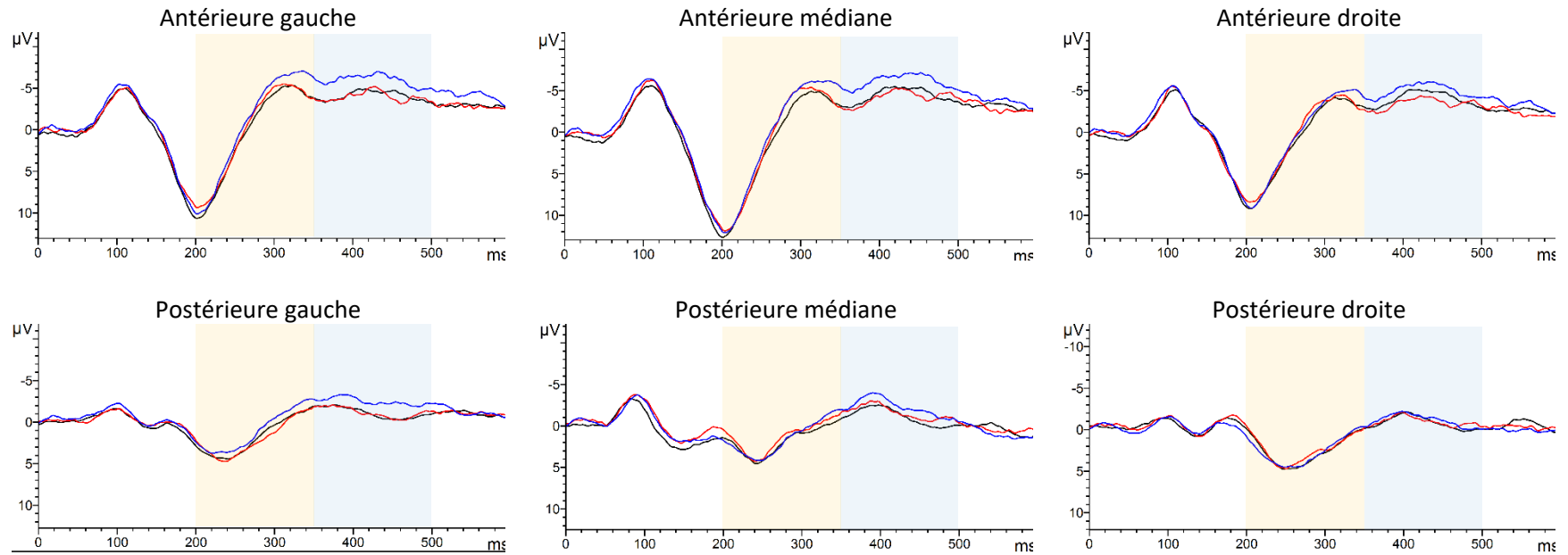
Les moyennes et écarts types pour l'amplitude de 200 ms à 350 ms et de 350 ms à 500 ms selon le type de racine pour chaque région d'intérêt sont présentés (Tableau 4.24). Puis, pour vérifier l'effet du type de racine sur l'amplitude, une ANOVA à mesures répétées $2 \times 3 \times 6$ (lexicalité \times type de racine \times région) a été effectuée pour chacune des fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms (voir Tableau 4.25).

Figure 4.7 Tracés des potentiels évoqués aux six régions d'intérêt en fonction du type de racine des mots



Note. Tracés des potentiels évoqués pour les mots polymorphémiques formés de racines liées (ligne noire), les mots polymorphémiques formés de racines libres (ligne rouge) et les mots monomorphémiques (ligne bleue) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Figure 4.8 Tracés des potentiels évoqués aux six régions d'intérêt en fonction du type de racine des pseudomots et des non-mots



Note. Tracés des potentiels évoqués pour les pseudomots polymorphémiques formés de racines liées (ligne noire), les pseudomots polymorphémiques formés de racines libres (ligne rouge) et les non-mots (ligne bleue) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Tableau 4.24 Moyennes et écarts types de l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon le type de racine pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Deux racines liées				
Antérieure gauche	-6,64	4,05	-6,88	3,44
Antérieure médiane	-7,05	3,86	-7,34	2,91
Antérieure droite	-5,88	3,58	-6,43	2,71
Postérieure gauche	-2,66	2,36	-3,84	2,13
Postérieure médiane	-3,59	2,71	-4,99	2,71
Postérieure droite	-2,98	3,26	-3,49	2,84
Deux racines libres				
Antérieure gauche	-8,31	3,96	-9,36	4,28
Antérieure médiane	-8,61	4,65	-9,32	4,27
Antérieure droite	-7,24	3,41	-8,51	4,05
Postérieure gauche	-4,56	2,60	-5,71	3,67
Postérieure médiane	-7,09	5,95	-7,50	5,80
Postérieure droite	-5,50	4,58	-5,64	3,75
Une racine libre				
Antérieure gauche	-7,74	3,76	-8,67	3,70
Antérieure médiane	-8,04	4,13	-8,62	3,27
Antérieure droite	-6,17	2,83	-7,64	2,75
Postérieure gauche	-3,87	2,76	-5,62	2,89
Postérieure médiane	-5,84	3,83	-6,86	4,00
Postérieure droite	-3,82	3,49	-5,35	3,18
Pseudomots et non-mots				
Deux racines liées				
Antérieure gauche	-6,71	3,61	-7,53	3,05
Antérieure médiane	-6,79	3,69	-8,10	2,85
Antérieure droite	-5,50	2,73	-7,35	2,32
Postérieure gauche	-2,99	2,60	-3,87	2,44
Postérieure médiane	-3,47	3,64	-4,76	2,62
Postérieure droite	-3,04	3,22	-3,64	2,64
Deux racines libres				
Antérieure gauche	-7,41	4,11	-7,70	3,76
Antérieure médiane	-7,66	4,50	-8,07	4,26
Antérieure droite	-6,32	4,03	-6,73	3,67
Postérieure gauche	-3,15	2,21	-4,10	2,13
Postérieure médiane	-4,44	3,97	-5,58	2,76
Postérieure droite	-3,20	3,09	-4,31	3,45
Une racine libre				
Antérieure gauche	-9,21	4,22	-9,71	3,59
Antérieure médiane	-9,30	4,72	-10,14	4,34
Antérieure droite	-7,32	3,84	-8,48	3,31
Postérieure gauche	-4,47	3,29	-5,49	3,00
Postérieure médiane	-4,48	4,38	-6,19	3,77
Postérieure droite	-3,50	3,93	-4,44	2,83

Note. $n = 29$.

Tableau 4.25 Résultats des ANOVAs à mesures répétées pour l'effet du type de racine sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	1,52	1, 28	0,228	0,05	1,41	1, 28	0,245	0,05
TypeRacine	10,07	2, 56	< 0,001	0,26	16,07	2, 56	< 0,001	0,37
Région	27,79	2,77, 77,60 ^a	< 0,001	0,50	40,87	3,72, 104,05 ^a	< 0,001	0,59
Lexicalité*TypeRacine	3,59	1,71, 47,82 ^b	0,042	0,11	4,33	1,63, 45,77 ^b	0,025	0,13
Lexicalité*Région	5,39	3,72, 104,01 ^a	0,001	0,16	3,27	3,33, 93,34 ^a	0,021	0,11
TypeRacine*Région	1,92	5,89, 164,83 ^a	0,081	0,06	0,75	5,00, 140,13 ^a	0,587	0,03
Lexicalité*TypeRacine*Région	1,66	5,34, 149,46 ^a	0,144	0,06	1,30	5,85, 163,81 ^a	0,259	0,05

Note. $n = 29$.

^a Correction Greenhouse-Geisser. ^b Correction Huynh-Feldt.

Dans la fenêtre temporelle de 200 ms à 350 ms, l'ANOVA à mesures répétées $2 \times 3 \times 6$ (lexicalité \times type de racine \times région) a révélé des effets principaux du type de racine et de régions sur l'amplitude, mais aucune interaction significative entre les facteurs de régions topographiques et du type de racine. Les résultats de l'ANOVA à mesures répétées $2 \times 3 \times 6$ (lexicalité \times type de racine \times région) réalisée sur la fenêtre temporelle de 350 ms à 500 ms permet de constater des effets principaux de type de racine et de régions, mais aucune interaction entre ces facteurs.

Pour la latence, les moyennes et les écarts types sont rapportées dans le Tableau 4.26 et les résultats de l'ANOVA à mesures répétées $2 \times 3 \times 6$ (lexicalité \times type de racine \times région) effectuée pour chacune des fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus, sont rapportés dans le Tableau 4.27. Dans la fenêtre temporelle de 200 ms à 350 ms, les résultats montrent un effet principal de régions et un effet moyen détecté pour l'interaction entre le type de racine et les régions topographiques. Cependant, l'analyse des comparaisons appariées ne permet pas de montrer des différences quant au type de racine au sein d'une même région topographique : toutes les comparaisons sont non significatives ($p > 0,05$). Dans la fenêtre temporelle de 350 ms à 500 ms, seul un effet principal marginal de région est détecté.

Tableau 4.26 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon le type de racine pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Deux racines liées				
Antérieure gauche	317,79	22,97	418,79	46,79
Antérieure médiane	317,10	25,69	412,97	47,49
Antérieure droite	321,52	17,43	410,83	43,26
Postérieure gauche	309,14	49,23	414,72	47,47
Postérieure médiane	287,38	58,36	407,21	35,93
Postérieure droite	283,07	65,23	399,83	36,44
Deux racines libres				
Antérieure gauche	321,45	26,76	413,55	53,44
Antérieure médiane	312,66	31,83	418,41	48,28
Antérieure droite	321,03	24,68	423,41	51,78
Postérieure gauche	322,83	37,06	406,76	46,81
Postérieure médiane	284,97	60,69	403,72	47,33
Postérieure droite	272,38	58,97	411,97	55,73
Une racine libre				
Antérieure gauche	313,90	24,14	415,24	45,35
Antérieure médiane	319,62	19,73	406,34	44,94
Antérieure droite	317,28	20,42	427,62	44,84
Postérieure gauche	311,55	47,30	403,48	44,78
Postérieure médiane	304,38	53,81	394,72	32,88
Postérieure droite	284,34	65,73	416,14	40,62
Pseudomots et non-mots				
Deux racines liées				
Antérieure gauche	313,45	23,60	419,03	43,73
Antérieure médiane	318,24	21,22	417,86	39,63
Antérieure droite	320,72	17,63	415,90	43,67
Postérieure gauche	324,21	36,60	416,97	47,10
Postérieure médiane	297,59	52,21	414,90	44,42
Postérieure droite	268,93	64,18	412,66	41,03
Deux racines libres				
Antérieure gauche	314,72	21,76	418,76	47,31
Antérieure médiane	310,28	21,82	438,10	45,16
Antérieure droite	312,34	20,44	433,86	44,42
Postérieure gauche	333,48	19,97	413,38	45,98
Postérieure médiane	288,07	55,38	405,79	39,95
Postérieure droite	276,10	63,04	413,03	43,58
Une racine libre				
Antérieure gauche	321,66	19,49	418,76	41,69
Antérieure médiane	320,72	23,61	425,62	39,09
Antérieure droite	317,66	28,34	424,59	42,56
Postérieure gauche	319,62	38,35	425,24	52,35
Postérieure médiane	301,14	53,46	405,34	41,48
Postérieure droite	306,48	55,94	406,21	44,81

Note. *n* = 29.

Tableau 4.27 Résultats des ANOVAs à mesures répétées pour l'effet du type de racine sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	0,80	1, 28	0,378	0,03	3,09	1, 28	0,090	0,10
TypeRacine	1,91	2, 56	0,157	0,06	0,48	2, 56	0,620	0,02
Région	11,49	2,36, 66,03 ^a	< 0,001	0,29	2,89*	2,74, 76,59 ^a	0,046	0,09
Lexicalité*TypeRacine	0,47	2, 56	0,626	0,02	0,05	1,76, 49,26 ^b	0,930	0,00
Lexicalité*Région	1,27	2,94, 82,19 ^a	0,291	0,04	0,57	5, 140	0,725	0,02
TypeRacine*Région	2,83	5,85, 163,72 ^a	0,013	0,09	1,25	5,92, 165,70 ^a	0,282	0,04
Lexicalité*TypeRacine*Région	1,46	5,45, 152,47 ^a	0,203	0,05	0,98	6,44, 180,29 ^a	0,443	0,03

Note. *n* = 29.

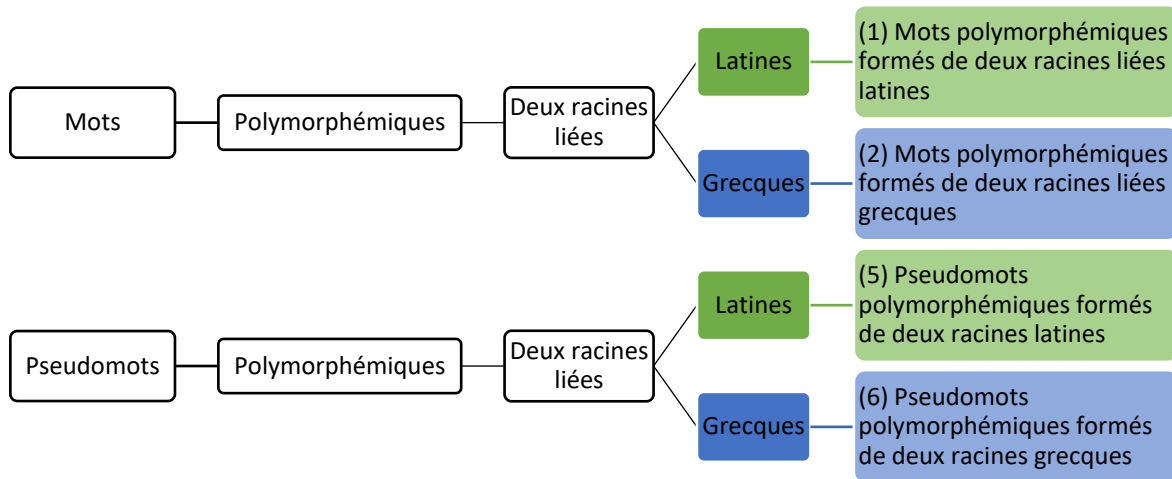
^a Correction Greenhouse-Geisser. ^b Correction Huynh-Feldt.

En résumé, le type de racine n'influence pas la précision des mots polymorphémiques. Les participants commettent autant d'erreurs avec les mots formés de racines liées que ceux formés de racines libres. Pour les pseudomots, ceux formés de racines liées entraînent plus d'erreurs que ceux formés de racines libres, ces derniers étant aussi bien réussis que les non-mots. Concernant le temps de réaction à la tâche de décision lexicale, le type de racine n'influence pas le temps de réaction au-delà de la structure des items : les participants répondent plus lentement aux items polymorphémiques qu'aux items monomorphémiques, et ce, peu importe la lexicalité. Enfin, les résultats des ANOVAs à mesures répétées n'ont pas permis de constater un effet du type de racine au sein des régions d'intérêt.

4.3.4 L'effet de l'origine des racines

L'origine des racines ne s'applique qu'à quatre catégories de mots. Pour comparer l'effet des racines latines à celui des racines grecques, les mots et les pseudomots formés de racines latines, puis de racines grecques, ont été groupés, soit respectivement (1) et (5) vs (2) et (6), comme illustré par la Figure 4.9.

Figure 4.9 Groupements des catégories d'items pour vérifier l'effet de l'origine des racines



Les moyennes et écarts types des mots et des pseudomots formés de racines liées latines et grecques sont présentés dans le Tableau 4.28. Pour vérifier l'effet de l'origine des racines sur la précision des participants lors de la tâche de décision lexicale, une ANOVA à mesures répétées 2 × 2 (lexicalité × origine) a été effectuée avec ces items (voir Tableau 4.29).

Tableau 4.28 Moyennes et écarts types de la précision (%) à la tâche de décision lexicale selon l'origine des racines

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Racines latines	62,41	11,82
Racines grecques	68,28	15,34
Pseudomots		
Racines latines	87,47	10,38
Racines grecques	79,77	12,21

Note. *n* = 29.

Tableau 4.29 Résultats de l'ANOVA à mesures répétées pour l'effet de l'origine des racines sur la précision (%) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	31,75	1, 28	< 0,001	0,53
OrigineRacine	0,55	1, 28	0,464	0,02
Lexicalité*OrigineRacine	22,19	1, 28	< 0,001	0,44

Note. *n* = 29.

Pour la précision, les résultats de l'ANOVA à mesures répétées montrent un effet principal de la lexicalité et une interaction significative entre la lexicalité et l'origine des racines. L'analyse des effets simples permet de constater que l'origine des racines a un effet sur la précision en lecture et que cet effet est plus important au sein des pseudomots. Effectivement, pour les mots, ceux d'origine latine entraînent plus d'erreurs à la tâche de décision lexicale ($M = 62,41$, $ÉT = 2,194$) que ceux d'origine grecque ($M = 68,28$, $ÉT = 2,85$, $p = 0,012$, $d = 1,00$). Pour les pseudomots, le profil inverse se dessine. Les pseudomots d'origine grecque ($M = 79,77$, $ÉT = 2,69$) entraînent significativement plus d'erreurs que ceux d'origine latine ($M = 87,47$, $ÉT = 1,93$, $p < 0,001$, $d = 1,80$).

En ce qui concerne le temps de réaction des participants, les statistiques descriptives sont présentées au Tableau 4.30 et les résultats de l'ANOVA à mesures répétées permettant de vérifier l'effet de l'origine des racines sur le temps de réaction à la tâche de décision lexicale sont présentés au Tableau 4.31. Les résultats permettent de constater des effets principaux de la lexicalité et de l'origine des racines, mais aucune interaction entre ces facteurs. Ainsi, les participants répondent plus rapidement aux items formés de racines latines qu'à ceux formés de racines grecques sans distinction selon la lexicalité.

Tableau 4.30 Moyennes et écarts types du temps de réaction (ms) à la tâche de décision lexicale selon l'origine des racines

Variable	<i>M</i>	<i>ÉT</i>
Mots		
Racines latines	1281,93	236,34
Racines grecques	1406,37	266,86
Pseudomots		
Racines latines	1541,78	320,31
Racines grecques	1681,34	336,23

Note. $n = 29$.

Tableau 4.31 Résultats de l'ANOVA à mesures répétées pour l'effet de l'origine des racines sur le temps de réaction (ms) à la tâche de décision lexicale

Facteur	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	70,05	1, 28	< 0,001	0,71
OrigineRacine	36,96	1, 28	< 0,001	0,57
Lexicalité*OrigineRacine	0,20	1, 28	0,662	0,01

Note. $n = 29$.

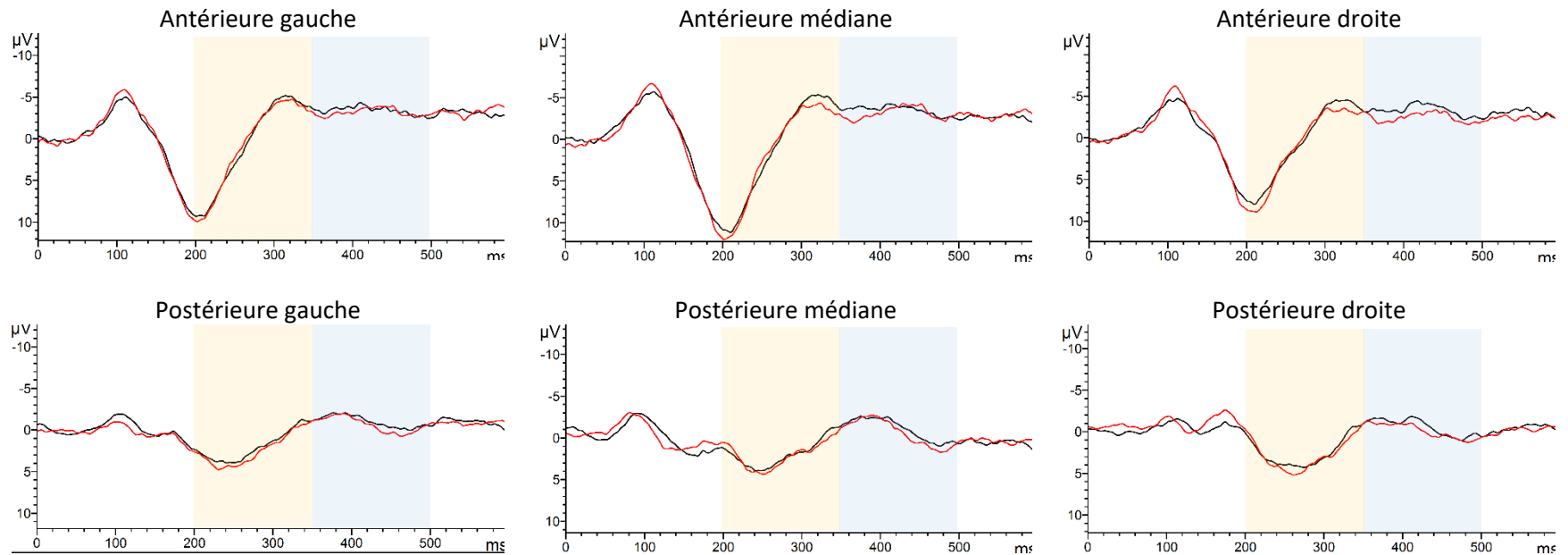
^a Correction Huynh-Feldt.

Les tracés des potentiels évoqués pour chacune des régions topographiques d'intérêt sont présentés à la Figure 4.10 pour les mots et à la Figure 4.11 pour les pseudomots. Pour les mots, ceux formés de racines grecques semblent entraîner des amplitudes plus grandes au marqueur N1/P2 en régions antérieures. Ce marqueur est suivi d'un pic négatif en régions antérieures vers 320 ms. Un deuxième pic négatif est visible surtout en région antérieure droite autour de 420 ms pour les mots formés de racines latines et autour de 460 ms pour les mots formés de racines grecques. En régions antérieure médiane (300 ms à 400 ms) et antérieure droite (300 ms à 500 ms), les mots formés de racines latines semblent entraîner de plus grandes négativités que les mots formés de racines grecques. Pour les pseudomots formés de racines latines et grecques, le marqueur N1/P2 est toujours bien visible dans les régions antérieures, et une plus grande négativité est observée pour les pseudomots formés de racines grecques. Contrairement aux mots, on distingue plus clairement deux pics négatifs qui suivent le N1/P2 dans toutes les régions antérieures, un premier vers 320 ms et un second vers 420 ms. Surtout en régions antérieures, le traitement des pseudomots formés de racines latines semblent entraîner de plus grandes négativités que celui des pseudomots formés de racines grecques.

Les moyennes et écarts types de l'amplitude maximale dans chacune des fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms, sont rapportés dans le Tableau 4.32. L'effet de l'origine des racines sur l'amplitude dans chacune des deux fenêtres temporelles a été mesuré par des ANOVAs à mesures répétées $2 \times 2 \times 6$ (lexicalité \times origine \times région). Les résultats figurent au Tableau 4.33.

Figure 4.10

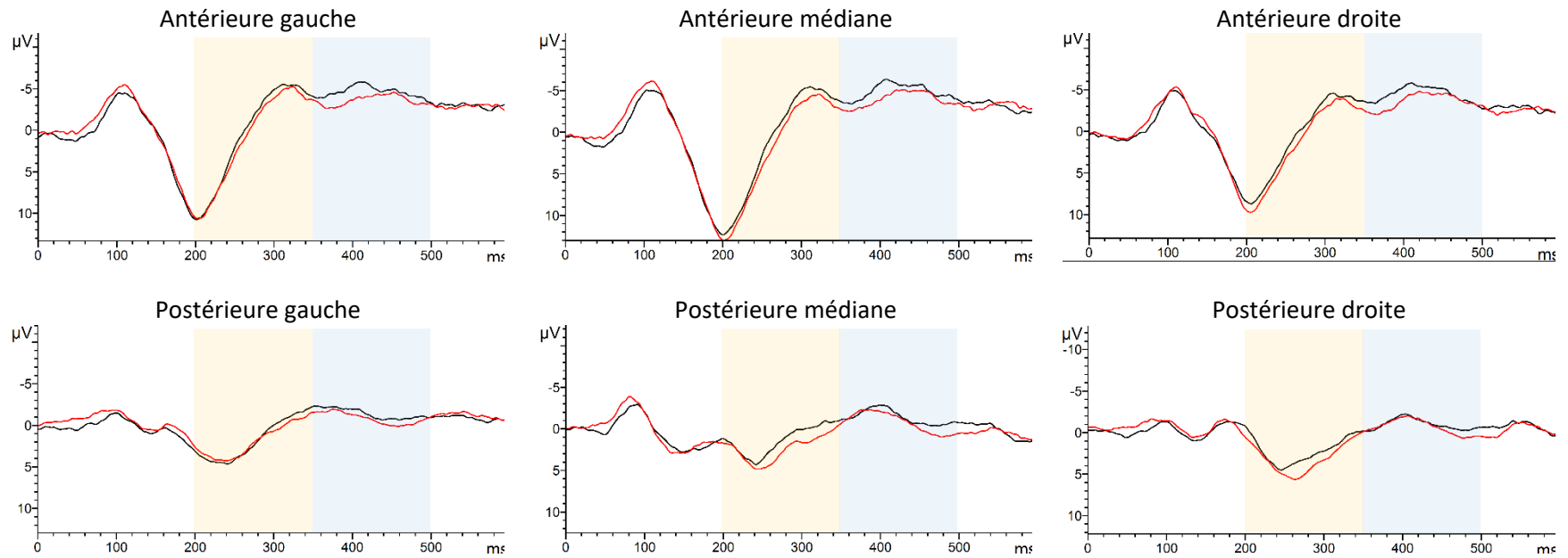
Tracés des potentiels évoqués aux six régions d'intérêt en fonction de l'origine des racines des mots



Note. Tracés des potentiels évoqués pour les mots polymorphémiques formés de racines latines (ligne noire) et les mots polymorphémiques formés de racines grecques (ligne rouge) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Figure 4.11

Tracés des potentiels évoqués aux six régions d'intérêt en fonction de l'origine des racines des pseudomots



Note. Tracés des potentiels évoqués pour les pseudomots polymorphémiques formés de racines latines (ligne noire) et les pseudomots polymorphémiques formés de racines grecques (ligne rouge) pour chacune des six régions d'intérêt. Le fond teinté jaune marque la fenêtre temporelle de 200 ms à 350 ms tandis que le fond teinté bleu marque la fenêtre temporelle de 350 ms à 500 ms.

Tableau 4.32 Moyennes et écarts types de l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon l'origine des racines pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Racines latines				
Antérieure gauche	-7,50	4,27	-8,36	3,90
Antérieure médiane	-8,22	4,15	-8,17	3,50
Antérieure droite	-6,91	4,31	-7,97	3,39
Postérieure gauche	-3,17	2,72	-4,75	1,98
Postérieure médiane	-4,58	3,62	-5,88	2,83
Postérieure droite	-3,71	3,78	-4,46	2,79
Racines grecques				
Antérieure gauche	-7,00	4,35	-7,39	3,75
Antérieure médiane	-7,00	4,63	-7,82	3,04
Antérieure droite	-6,06	3,89	-6,52	3,02
Postérieure gauche	-3,31	2,99	-4,57	2,23
Postérieure médiane	-4,08	4,17	-5,78	3,30
Postérieure droite	-3,51	3,97	-4,01	3,64
Pseudomots				
Racines latines				
Antérieure gauche	-7,30	3,71	-8,60	3,27
Antérieure médiane	-7,73	4,10	-9,27	3,32
Antérieure droite	-6,46	2,91	-8,39	3,04
Postérieure gauche	-3,84	2,54	-4,70	2,51
Postérieure médiane	-4,86	4,21	-5,50	2,72
Postérieure droite	-3,99	3,51	-4,49	3,02
Racines grecques				
Antérieure gauche	-7,17	3,98	-7,83	3,47
Antérieure médiane	-6,77	4,25	-8,48	3,00
Antérieure droite	-5,73	3,46	-7,97	2,93
Postérieure gauche	-3,50	2,91	-4,48	3,25
Postérieure médiane	-3,77	3,76	-5,36	3,40
Postérieure droite	-3,27	3,46	-4,35	2,51

Note. *n* = 29.

Tableau 4.33 Résultats des ANOVAs à mesures répétées pour l'effet de l'origine des racines sur l'amplitude (μV) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	0,03	1, 28	0,872	0,00	1,23	1, 28	0,277	0,04
OrigineRacine	2,94	1, 28	0,097	0,10	4,86	1, 28	0,036	0,15
Région	20,19	2,96, 82,94 ^a	< 0,001	0,42	38,39	3,16, 88,48 ^a	< 0,001	0,58
Lexicalité*OrigineRacine	0,04	1, 28	0,850	0,00	0,20	1, 28	0,657	0,01
Lexicalité*Région	0,68	3,37, 94,48 ^a	0,586	0,03	2,04	3,04, 85,11 ^a	0,113	0,07
OrigineRacine*Région	0,89	2,70, 75,63 ^a	0,444	0,03	0,91	2,69, 75,19 ^a	0,434	0,03
Lexicalité*OrigineRacine*Région	0,32	3,09, 86,53 ^a	0,820	0,01	0,35	4,54, 126,98 ^b	0,864	0,01

Note. $n = 29$.

^a Correction Greenhouse-Geisser. ^b Correction Huynh-Feldt.

De 200 ms à 350 ms, seul un effet principal de région sur l'amplitude des potentiels évoqués est détecté. Aucune interaction avec les régions d'intérêt n'est significative. De 350 ms à 500 ms, des effets principaux de l'origine des racines et de régions sont constatés sur l'amplitude. Les items formés de racines latines entraînent de plus grandes négativités que les items formés de racines grecques. Aucune interaction n'est significative.

En ce qui concerne la latence, les moyennes et écarts types de la latence dans chacune des fenêtres temporelles, soit de 200 ms à 350 ms et de 350 ms à 500 ms, sont rapportés dans le Tableau 4.34. L'effet de l'origine des racines sur la latence dans les deux fenêtres temporelles a été mesuré par des ANOVAs à mesures répétées $2 \times 2 \times 6$ (lexicalité \times origine \times région) dont les résultats figurent au Tableau 4.35. De 200 ms à 350 ms, pour la latence, seul un effet principal de région est constaté. Aucune interaction n'est significative. De 350 ms à 500 ms, aucun effet principal ni interaction atteignent le seuil de significativité.

Tableau 4.34 Moyennes et écarts types de la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale selon l'origine des racines pour chaque région d'intérêt

Variable	200 ms à 350 ms		350 ms à 500 ms	
	<i>M</i>	<i>ÉT</i>	<i>M</i>	<i>ÉT</i>
Mots				
Racines latines				
Antérieure gauche	317,17	23,25	414,69	51,42
Antérieure médiane	316,86	28,39	411,00	46,29
Antérieure droite	319,28	28,88	411,93	44,73
Postérieure gauche	306,52	47,16	413,52	52,13
Postérieure médiane	286,38	59,12	401,21	38,40
Postérieure droite	292,72	60,08	407,14	40,60
Racines grecques				
Antérieure gauche	309,00	26,45	419,10	46,15
Antérieure médiane	309,79	28,09	423,03	45,09
Antérieure droite	319,72	23,08	413,93	47,80
Postérieure gauche	307,66	48,95	413,79	48,45
Postérieure médiane	296,55	55,10	405,17	41,29
Postérieure droite	286,10	62,56	415,07	47,27
Pseudomots				
Racines latines				
Antérieure gauche	314,90	21,6	419,00	42,96
Antérieure médiane	318,21	23,90	429,03	40,88
Antérieure droite	323,28	19,27	421,24	37,28
Postérieure gauche	311,34	44,54	414,21	50,58
Postérieure médiane	291,90	54,45	415,69	39,19
Postérieure droite	269,24	61,43	425,45	45,16
Racines grecques				
Antérieure gauche	311,72	25,42	418,41	42,12
Antérieure médiane	315,28	25,49	420,28	40,38
Antérieure droite	316,31	23,66	422,21	39,96
Postérieure gauche	309,69	46,61	421,83	48,92
Postérieure médiane	292,48	55,57	407,86	38,67
Postérieure droite	284,41	59,56	405,21	38,95

Note. *n* = 29.

Tableau 4.35 Résultats des ANOVAs à mesures répétées pour l'effet de l'origine des racines sur la latence (ms) de 200 ms à 350 ms et de 350 ms à 500 ms à la tâche de décision lexicale

Facteur	200 ms à 350 ms				350 ms à 500 ms			
	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p	<i>F</i>	<i>ddl</i>	<i>p</i>	η^2p
Lexicalité	0,03	1, 28	0,857	0,00	3,96	1, 28	0,056	0,12
OrigineRacine	0,04	1, 28	0,844	0,00	0,00	1, 28	0,971	0,00
Région	8,72	2,78, 77,89 ^a	< 0,001	0,24	1,39	5, 140	0,230	0,05
Lexicalité*OrigineRacine	0,06	1, 28	0,808	0,00	2,20	1, 28	0,149	0,07
Lexicalité*Région	0,78	3,04, 85,01 ^a	0,512	0,03	0,17	3,40, 95,16 ^a	0,934	0,01
OrigineRacine*Région	0,78	2,85, 79,78 ^a	0,500	0,03	0,31	5, 140	0,906	0,01
Lexicalité*OrigineRacine*Région	1,24	3,03, 84,92 ^a	0,301	0,04	1,04	3,36, 94,13 ^a	0,383	0,04

Note. *n* = 29.

^a Correction Greenhouse-Geisser. ^b Correction Huynh-Feldt.

Pour résumer, les résultats à la tâche de décision lexicale permettent de constater que les mots formés de racines grecques sont mieux réussis que les mots formés de racines latines. Pour les pseudomots, c'est l'inverse : les pseudomots formés de racines latines entraînent moins d'erreurs que les pseudomots formés de racines grecques. Aussi, les participants répondent plus rapidement aux mots et aux pseudomots formés de racines latines que grecques, qui, eux, leur demandent plus de temps de traitement. Enfin, les résultats des ANOVAs à mesures répétées n'ont pas permis de détecter d'effets de l'origine des racines sur l'amplitude et la latence aux deux fenêtres temporelles au sein des différentes régions d'intérêt.

4.3.5 Les effets de fréquence du mot entier et des racines

Dans la tâche de décision lexicale, des variables linguistiques ont été manipulées, soit la lexicalité, la structure morphologique des items, le type de racine ainsi que l'origine des racines. D'autres facteurs linguistiques qui n'ont pas été manipulés expérimentalement dans la tâche de décision lexicale sont susceptibles d'intervenir dans les relations à l'étude : la fréquence du mot entier et la fréquence des racines. Des facteurs individuels sont aussi considérés dans le cadre de cette recherche. Ils concernent les données recueillies au test dynamique : la connaissance du vocabulaire, la reconnaissance des racines latines et grecques, l'habileté de décomposition morphologique, la connaissance du sens des racines latines et grecques ainsi que l'habileté d'analyse morphologique. Leur analyse est présentée à la section 4.4.

Afin de vérifier si ces autres facteurs linguistiques et individuels ciblés sont liés aux variables dépendantes (précision, temps de réaction, amplitude de 200 ms à 350 ms et de 350 ms à 500 ms ainsi

que latence de 200 ms à 350 ms et de 350 ms à 500 ms), les régressions multiniveaux sont utilisées. Des régressions logistiques multiniveaux permettent d'évaluer l'effet de ces variables indépendantes sur la précision, tandis que des régressions linéaires multiniveaux mesurent l'effet de ces mêmes variables, mais sur les autres variables dépendantes. Aux modèles de régression, les variables manipulées lors de la tâche de décision lexicale sont ajoutées, soit les catégories d'items ainsi que les six régions topographiques (uniquement pour les variables dépendantes d'amplitude et de latence). Néanmoins, les résultats obtenus avec ces variables dans les modèles de régression ne seront pas présentés puisqu'ils convergent avec les résultats des ANOVAs présentés aux sections précédentes (4.3.1, 4.3.2, 4.3.3 et 4.3.4). De plus, il convient de contrôler pour le niveau scolaire des participants puisque cette variable influençait les résultats sur certaines variables dépendantes de l'étude, soit la précision et l'amplitude de 350 ms à 500 ms (voir section 4.1). Les résultats spécifiques de l'influence du niveau scolaire sont présentés à la prochaine section sur l'influence des facteurs individuels (4.4). Enfin, il est à noter que des modèles de régression différents ont été effectués pour les mots ainsi que les pseudomots et les non-mots comme le facteur de fréquence du mot entier est inexistant pour les pseudomots et les non-mots. Les résultats de l'influence de la fréquence du mot entier et de la fréquence des racines sur le traitement des mots écrits sont présentés dans cette section. Les résultats concernant les facteurs individuels font l'objet de la prochaine section (4.4).

Pour la précision lors de la tâche de décision lexicale, une régression logistique multiniveaux a été effectuée comme la précision est une variable catégorielle à deux niveaux (bonne réponse ou mauvaise réponse). Les résultats du modèle de régression logistique multiniveaux sont présentés au Tableau 4.36. Pour les mots, seule la fréquence du mot entier a un effet significatif sur la précision des participants (rapport de cote = 1,03, $p < 0,001$). Ainsi, si la fréquence des mots augmente d'une valeur de 1, il est 1,03 fois plus probable pour les participants d'obtenir une bonne réponse aux mots de la tâche de décision lexicale, soit une augmentation 0,03 %. Pour bien comprendre l'implication de ces résultats, prenons en exemple deux mots composés formés de racines grecques, le mot *hexagone* qui apparaît 61 fois dans les manuels scolaires des élèves du primaire et le mot *antonyme* qui est plus fréquent, à 128 occurrences (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018). Le mot *antonyme* apparaît donc 67 fois de plus dans les manuels scolaires que le mot *hexagone*. Ainsi, le modèle prédit qu'il est 7,46 fois ($\exp[0,03*67]$) plus probable que les participants fournissent une bonne réponse au mot *antonyme* qu'au mot *hexagone*. Pour les pseudomots et les non-mots, c'est plutôt la fréquence des racines qui intervient dans la relation à l'étude, plus particulièrement la fréquence de la deuxième racine (rapport de cote = 1,000044,

$p < 0,001$). Effectivement, les résultats montrent que si la fréquence de la deuxième racine des mots composés augmente d'une seule valeur, il est 1,000044 fois plus probable pour les participants d'obtenir une bonne réponse aux pseudomots et non-mots de la tâche de décision lexicale, soit une augmentation 0,000044 %. Prenons en exemples les pseudomots **archéoderme* et **tétragraphe*. Ces deux pseudomots sont formés de racines grecques. La racine *derm* apparaît 56 fois dans les manuels scolaires des élèves alors que la racine *graph* apparaît 3 556 fois, une différence de 3 500 occurrences (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018). Le modèle prédit qu'il sera 1,17 fois ($\exp[0,000044 * 3\ 500]$) plus probable de fournir une bonne réponse à la tâche de décision lexicale pour le pseudomot **tétragraphe* que le pseudomot **archéoderme*. Il est à noter que le grand échantillon d'observations pour les analyses de régression permet d'avoir énormément de puissance statistique et ainsi détecter des effets aussi minuscules. Les autres facteurs linguistiques de fréquence ne sont pas suffisamment puissants pour intervenir sur la précision des mots ainsi que des pseudomots et des non-mots.

Tableau 4.36 Résultats du modèle de régression logistique multiniveaux pour la précision (%) à la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Coefficient	Rapport de cote	<i>p</i>	IC 95 %	Coefficient	Rapport de cote	<i>p</i>	IC 95 %
Effets fixes								
Constante	0,48	1,61	0,756	[0,08, 32,22]	8,50	4900,96	0,002	[24,31, 988101,50]
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	-0,97	0,38	<0,001	[0,34, 0,42]	-1,05	0,35	<0,001	[0,29, 0,42]
Polymorphémiques formés de deux racines liées grecques	-0,62	0,54	<0,001	[0,48, 0,60]	-1,63	0,20	<0,001	[0,17, 0,23]
Polymorphémiques formés de deux racines libres	-0,98	0,38	<0,001	[0,33, 0,42]	-0,59	0,56	<0,001	[0,45, 0,69]
Monomorphémiques	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	-0,18	0,84	0,488	[0,50, 1,39]	0,22	1,25	0,627	[0,51, 3,08]
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	0,03	1,03	<0,001	[1,03, 1,03]	-	-	-	-
Fréquence de la 1 ^{re} racine	-0,00	1,00	0,628	[1,00, 1,00]	-0,00	1,00	0,440	[1,00, 1,00]
Fréquence de la 2 ^e racine	-0,00	1,00	0,094	[1,00, 1,00]	0,00	1,00	<0,001	[1,00, 1,00]
Connaissance du vocabulaire	0,05	1,05	0,272	[0,96, 1,16]	-0,01	0,99	0,900	[0,84, 1,17]
Reconnaissance des racines	-0,01	0,99	0,806	[0,93, 1,06]	0,04	1,04	0,482	[0,93, 1,16]
Décomposition morphologique	-0,02	0,98	0,853	[0,83, 1,17]	-0,43	0,65	0,008	[0,48, 0,89]
Connaissance du sens des racines	0,01	1,01	0,423	[0,98, 1,04]	0,02	1,02	0,438	[0,97, 1,07]
Analyse morphologique	0,01	1,01	0,877	[0,93, 1,10]	0,02	1,02	0,830	[0,88, 1,18]
Effets aléatoires								
Variance (constante)	0,32				0,97			
<i>n</i>	29				29			
Observations	20316				20022			

Note. IC 95 % = intervalle de confiance à 95 % pour le rapport de cote.

Pour vérifier l'effet des facteurs linguistiques et individuels sur les variables dépendantes continues (temps de réaction, amplitude de 200 ms à 350 ms et de 350 ms à 500 ms, latence de 200 ms à 350 ms et de 350 ms à 500 ms), la normalité de la distribution des données, postulat à la régression linéaire multiniveaux, a d'abord été vérifiée. Les données relatives au temps de réaction, quoique légèrement asymétriques, ainsi qu'à l'amplitude de 200 ms à 350 ms et de 350 ms à 500 ms étant normales : elles ont donc été utilisées telles quelles dans les analyses subséquentes. Pour les données relatives à la latence aux deux fenêtres temporelles, les bornes inférieures et supérieures étaient surreprésentées (voir Figure 4.12). Cela peut indiquer que les pics d'amplitude maximale des potentiels évoqués qui ont été automatiquement détectés chevauchent les fenêtres temporelles déterminées. De ce fait, les données frontières ont été retirées avant de poursuivre l'analyse des données. Cela correspond à une perte de 19,6 % (8 190 observations) et 12,1 % (5 040 observations) respectivement pour la latence de 200 ms à 350 ms et de 350 ms à 500 ms. Cette modification permet d'obtenir des données qui sont plus centrées autour de la moyenne (voir Figure 4.13).

Figure 4.12

Distribution des données initiales pour la latence de 200 ms à 350 ms et de 350 ms à 500 ms

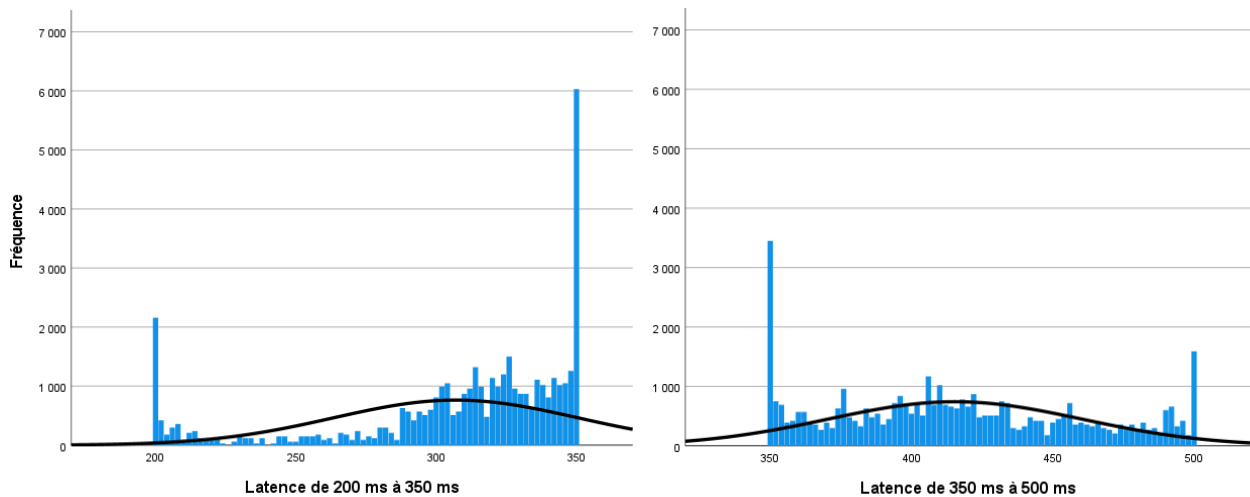
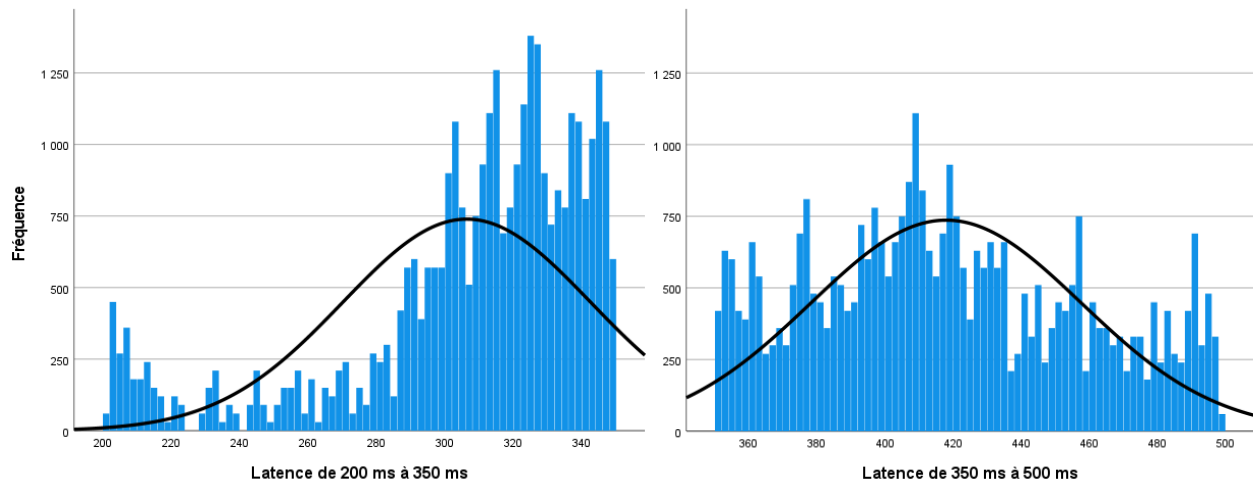


Figure 4.13

Distribution des données filtrées sans les bornes inférieures et supérieures pour la latence de 200 ms à 350 ms et de 350 ms à 500 ms



Enfin, pour répondre à la question de recherche et vérifier si des facteurs linguistiques de fréquence sont liés au temps de réaction (Tableau 4.37), à l'amplitude (Tableau 4.38) et à la latence de 200 ms à 350 ms (Tableau 4.39) ainsi qu'à l'amplitude (Tableau 4.40) et à la latence de 350 ms à 500 ms (Tableau 4.41), des analyses de régression linéaire multiniveaux ont été réalisées pour chacune des variables dépendantes, séparément pour les mots et pour les pseudomots et les non-mots.

Pour le temps de réaction (voir Tableau 4.37), la fréquence du mot a un effet sur le temps de réaction des participants lorsqu'ils répondent à des mots uniquement ($p < 0,001$). Pour les pseudomots et les non-mots, c'est plutôt la fréquence de la première racine qui influence le temps de réaction des participants ($p = 0,006$). Ainsi, les estimations négatives des effets fixes indiquent que les participants sont plus rapides pour fournir une réponse correcte lorsque les mots sont fréquents. Pour les pseudomots et les non-mots, la même situation se produit : plus la fréquence de la première racine est élevée, plus les participants sont rapides à fournir une bonne réponse.

Tableau 4.37 Résultats du modèle de régression linéaire multiniveaux pour le temps de réaction (ms) des participants à fournir une réponse correcte à la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets fixes								
Constante	2164,42	[1038,42, 3290,41]	< 0,001	21,98	2548,49	[1027,51, 4069,46]	0,002	21,99
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	131,71	[112,06, 151,37]	< 0,001	14491,36	114,33	[96,77, 131,89]	< 0,001	18296,12
Polymorphémiques formés de deux racines liées grecques	259,58	[240,86, 278,30]	< 0,001	14491,98	248,87	[231,30, 266,44]	< 0,001	18296,18
Polymorphémiques formés de deux racines libres	233,61	[209,63, 257,59]	< 0,001	14492,45	134,38	[113,82, 154,94]	< 0,001	18296,08
Monomorphémiques	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	138,88	[-52,94, 330,71]	0,147	21,99	152,10	[-106,98, 411,17]	0,236	21,99
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	-1,15	[-1,26, -1,04]	< 0,001	14491,62	-	-	-	-
Fréquence de la 1 ^{re} racine	-0,00	[-0,01, 0,00]	0,772	14491,56	-0,01	[-0,01, -0,00]	0,006	18296,08
Fréquence de la 2 ^e racine	-0,00	[-0,00, 0,00]	0,332	14491,46	-0,00	[-0,00, 0,00]	0,848	18296,07
Connaissance du vocabulaire	-8,90	[-44,43, 26,64]	0,609	21,97	-7,57	[-55,57, 40,43]	0,747	21,99
Reconnaissance des racines	-1,55	[-25,22, 22,11]	0,893	21,98	-9,92	[-41,88, 22,05]	0,527	21,99
Décomposition morphologique	-40,94	[-106,52, 24,63]	0,209	22,00	-47,82	[-136,38, 40,73]	0,275	21,99
Connaissance du sens des racines	-1,97	[-12,31, 8,36]	0,696	21,96	-3,20	[-17,16, 10,77]	0,640	21,99
Analyse morphologique	-3,60	[-35,45, 28,25]	0,817	21,97	3,66	[-39,36, 46,69]	0,861	21,99

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets aléatoires								
Résidus σ^2	165712,16				176310,81			
Variance (constante)	41035,11				75220,30			
<i>n</i>	29				29			
ICC	0,22				0,34			
Observations	14526				18330			

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau 4.38 Résultats du modèle de régression linéaire multiniveaux pour l'amplitude maximale (μV) de 200 ms à 350 ms durant la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets fixes								
Constante	-3,93	[-17,73, 9,86]	0,561	22,00	1,34	[-11,20, 13,89]	0,827	22,00
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	0,23	[0,10, 0,36]	< 0,001	20840,00	0,68	[0,56, 0,80]	< 0,001	20841,00
Polymorphémiques formés de deux racines liées grecques	0,75	[0,63, 0,88]	< 0,001	20840,00	1,35	[1,23, 1,46]	< 0,001	20841,00
Polymorphémiques formés de deux racines libres	-0,97	[-1,12, -0,82]	< 0,001	20840,00	1,02	[0,88, 1,15]	< 0,001	20841,00
Monomorphémiques	-	-	-	-	-	-	-	-
Régions								
Antérieure gauche	-3,50	[-3,65, -3,35]	< 0,001	20840,00	-4,28	[-4,42, -4,14]	< 0,001	20841,00
Antérieure médiane	-3,83	[-3,98, -3,69]	< 0,001	20840,00	-4,37	[-4,51, -4,24]	< 0,001	20841,00
Antérieure droite	-2,45	[-2,61, -2,31]	< 0,001	20840,00	-2,96	[-3,10, -2,83]	< 0,001	20841,00
Postérieure gauche	0,41	[0,26, 0,55]	< 0,001	20840,00	-0,25	[-0,38, -0,11]	< 0,001	20841,00
Postérieure médiane	-1,26	[-1,41, -1,12]	< 0,001	20840,00	-0,90	[-1,03, -0,76]	< 0,001	20841,00
Postérieure droite	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	-1,20	[-3,55, 1,15]	0,302	22,00	-2,21	[-4,35, -0,08]	0,043	22,00
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	0,00	[0,00, 0,00]	1,00	20840,00	-	-	-	-
Fréquence de la 1 ^{re} racine	0,00	[0,00, 0,00]	1,00	20840,00	0,00	[0,00, 0,00]	1,00	20841,00
Fréquence de la 2 ^e racine	0,00	[0,00, 0,00]	1,00	20840,00	0,00	[0,00, 0,00]	1,00	20841,00
Connaissance du vocabulaire	0,06	[-0,38, 0,49]	0,788	22,00	-0,06	[-0,46, 0,34]	0,757	22,00
Reconnaissance des racines	0,07	[-0,22, 0,36]	0,644	22,00	-0,00	[-0,27, 0,26]	0,973	22,00
Décomposition morphologique	-0,09	[-0,89, 0,71]	0,817	22,00	0,13	[-0,60, 0,86]	0,716	22,00
Connaissance du sens des racines	-0,06	[-0,19, 0,07]	0,328	22,00	-0,05	[-0,16, 0,07]	0,397	22,00

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Analyse morphologique	0,09	[-0,30, 0,48]	0,632	22,00	-0,10	[-0,45, 0,26]	0,569	22,00
Effets aléatoires								
Résidus σ^2	9,90				8,56			
Variance (constante)	6,20				5,13			
<i>n</i>	29				29			
ICC	0,29				0,31			
Observations	20880				20880			

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau 4.39 Résultats du modèle de régression linéaire multiniveaux pour la latence (ms) de 200 ms à 350 ms durant la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets fixes								
Constante	338,10	[245,73, 430,46]	< 0,001	21,97	315,62	[216,07, 415,17]	< 0,001	22,00
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	-1,89	[-3,41, -0,37]	0,015	16733,97	-7,39	[-8,67, -6,12]	< 0,001	16762,69
Polymorphémiques formés de deux racines liées grecques	-7,58	[-9,08, -6,09]	< 0,001	16740,74	-11,53	[-12,75, -10,30]	< 0,001	16762,03
Polymorphémiques formés de deux racines libres	-3,88	[-5,66, -2,10]	< 0,001	16732,44	-5,61	[-7,07, -4,14]	< 0,001	16761,71
Monomorphémiques	-	-	-	-	-	-	-	-
Régions								
Antérieure gauche	20,96	[19,17, 22,75]	< 0,001	16734,00	17,54	[15,98, 19,10]	< 0,001	16765,25
Antérieure médiane	20,71	[18,93, 22,49]	< 0,001	16733,85	17,54	[15,97, 19,11]	< 0,001	16765,49
Antérieure droite	24,54	[22,74, 26,34]	< 0,001	16734,09	18,75	[17,18, 20,31]	< 0,001	16765,58
Postérieure gauche	14,92	[13,04, 16,80]	< 0,001	16733,17	13,54	[11,88, 15,20]	< 0,001	16763,53
Postérieure médiane	2,41	[0,52, 4,30]	0,012	16732,19	-9,63	[-11,26, -8,00]	< 0,001	16763,21
Postérieure droite	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	12,07	[-3,67, 27,80]	0,126	21,980	2,62	[-14,34, 19,58]	0,752	22,01
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	0,00	[-0,01, 0,01]	1,00	16729,99	-	-	-	-

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Fréquence de la 1 ^{re} racine	0,00	[-0,00, 0,00]	1,00	16729,99	0,00	[-0,00, 0,00]	1,00	16761,00
Fréquence de la 2 ^e racine	0,00	[-0,00, 0,00]	1,00	16729,99	0,00	[-0,00, 0,00]	1,00	16761,00
Connaissance du vocabulaire	0,39	[-2,52, 3,31]	0,783	21,97	0,00	[-3,14, 3,14]	0,999	22,00
Reconnaissance des racines	0,08	[-1,86, 2,02]	0,934	22,03	-0,78	[-2,88, 1,31]	0,445	22,00
Décomposition morphologique	-0,35	[-5,73, 5,02]	0,893	21,98	3,69	[-2,11, 9,48]	0,201	22,02
Connaissance du sens des racines	-0,02	[-0,87, 0,83]	0,964	21,97	0,00	[-0,91, 0,92]	0,995	21,99
Analyse morphologique	-1,45	[-4,06, 1,17]	0,264	22,00	-1,75	[-4,56, 1,07]	0,212	22,01
Effets aléatoires								
Résidus σ^2	1069,72				792,75			
Variance (constante)	276,59				322,11			
<i>n</i>	29				29			
ICC	0,20				0,26			
Observations	16770				16800			

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau 4.40 Résultats du modèle de régression linéaire multiniveaux pour l'amplitude maximale (μV) de 350 ms à 500 ms durant la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets fixes								
Constante	5,22	[-6,06, 16,49]	0,348	22,00	6,75	[-3,84, 17,34]	0,200	22,00
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	0,44	[0,32, 0,55]	< 0,001	20840,00	0,58	[0,49, 0,68]	< 0,001	20841,00
Polymorphémiques formés de deux racines liées grecques	1,11	[1,00, 1,22]	< 0,001	20840,00	1,00	[0,90, 1,09]	< 0,001	20841,00
Polymorphémiques formés de deux racines libres	-0,55	[-0,68, -0,41]	< 0,001	20840,00	1,33	[1,22, 1,44]	< 0,001	20841,00
Monomorphémiques	-	-	-	-	-	-	-	-
Régions								
Antérieure gauche	-3,58	[-3,71, -3,45]	< 0,001	20840,00	-4,06	[-4,17, -3,95]	< 0,001	20841,00
Antérieure médiane	-3,75	[-3,88, -3,62]	< 0,001	20840,00	-4,59	[-4,71, -4,48]	< 0,001	20841,00
Antérieure droite	-2,80	[-2,92, -2,67]	< 0,001	20840,00	-3,49	[-3,60, -3,38]	< 0,001	20841,00
Postérieure gauche	-0,30	[-0,43, -0,17]	< 0,001	20840,00	-0,29	[-0,41, -0,18]	< 0,001	20841,00
Postérieure médiane	-1,64	[-1,77, -1,51]	< 0,001	20840,00	-1,26	[-1,37, -1,14]	< 0,001	20841,00
Postérieure droite	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	-1,42	[-3,34, 0,50]	0,139	22,00	-1,58	[-3,39, 0,22]	0,083	22,00
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	0,00	[0,00, 0,00]	1,00	20840,00	-	-	-	-
Fréquence de la 1 ^{re} racine	0,00	[0,00, 0,00]	1,00	20840,00	0,00	[0,00, 0,00]	1,00	20841,00
Fréquence de la 2 ^e racine	0,00	[0,00, 0,00]	1,00	20840,00	0,00	[0,00, 0,00]	1,00	20841,00
Connaissance du vocabulaire	0,12	[-0,23, 0,48]	0,483	22,00	-0,06	[-0,40, 0,27]	0,692	22,00
Reconnaissance des racines	0,08	[-0,16, 0,31]	0,503	22,00	-0,01	[-0,24, 0,21]	0,904	22,00
Décomposition morphologique	-0,55	[-1,21, 0,11]	0,096	22,00	-0,29	[-0,91, 0,33]	0,339	22,00
Connaissance du sens des racines	-0,05	[-0,15, 0,05]	0,332	22,00	-0,02	[-0,11, 0,08]	0,732	22,00

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Analyse morphologique	-0,02	[-0,34, 0,30]	0,914	22,00	-0,12	[-0,42, 0,18]	0,405	22,00
Effets aléatoires								
Résidus σ^2	7,48				5,88			
Variance (constante)	4,14				3,65			
<i>n</i>	29				29			
ICC	0,33				0,33			
Observations	20880				20880			

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau 4.41 Résultats du modèle de régression linéaire multiniveaux pour la latence (ms) de 350 ms à 500 ms durant la tâche de décision lexicale

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Effets fixes								
Constante	426,73	[315,97, 537,48]	< 0,001	22,02	434,55	[345,26, 523,85]	< 0,001	22,02
Catégorie d'items								
Polymorphémiques formés de deux racines liées latines	-0,60	[-2,19, 0,99]	0,461	18051,38	2,77	[1,36, 4,17]	< 0,001	18592,80
Polymorphémiques formés de deux racines liées grecques	1,46	[-0,06, 2,98]	0,060	18052,02	-0,65	[-2,02, 0,72]	0,350	18592,39
Polymorphémiques formés de deux racines libres	1,18	[-0,69, 3,06]	0,216	18051,25	3,56	[1,91, 5,21]	< 0,001	18592,18
Monomorphémiques	-	-	-	-	-	-	-	-
Régions								
Antérieure gauche	5,31	[3,48, 7,15]	< 0,001	18051,90	13,10	[11,43, 14,76]	< 0,001	18593,79
Antérieure médiane	5,62	[3,79, 7,45]	< 0,001	18052,14	17,87	[16,19, 19,54]	< 0,001	18592,53
Antérieure droite	9,56	[7,75, 11,37]	< 0,001	18051,42	17,08	[15,43, 18,72]	< 0,001	18592,16
Postérieure gauche	-4,69	[-6,48, -2,90]	< 0,001	18050,63	5,69	[4,04, 7,35]	< 0,001	18592,61
Postérieure médiane	-11,60	[-13,40, -9,79]	< 0,001	18050,58	-4,42	[-6,04, -2,79]	< 0,001	18591,46
Postérieure droite	-	-	-	-	-	-	-	-
Niveau scolaire								
5 ^e année	-9,15	[-28,01, 9,72]	0,326	21,99	2,49	[-12,72, 17,70]	0,738	22,00
6 ^e année	-	-	-	-	-	-	-	-
Fréquence du mot	0,00	[-0,01, 0,01]	1,00	18049,99	-	-	-	-
Fréquence de la 1 ^{re} racine	0,00	[-0,00, 0,00]	1,00	18049,99	0,00	[-0,00, 0,00]	1,00	18591,00

	Mots				Pseudomots et non-mots			
	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>	Estimation	IC 95 %	<i>p</i>	<i>ddl</i>
Fréquence de la 2 ^e racine	0,00	[-0,00, 0,00]	1,00	18049,99	0,00	[-0,00, 0,00]	1,00	18591,00
Connaissance du vocabulaire	-2,66	[-6,15, 0,84]	0,129	22,00	-2,69	[-5,51, 0,13]	0,060	22,00
Reconnaissance des racines	-0,97	[-3,29, 1,36]	0,399	21,99	-1,84	[-3,71, 0,04]	0,055	22,00
Décomposition morphologique	2,32	[-4,13, 8,77]	0,464	22,00	1,22	[-3,98, 6,42]	0,632	22,01
Connaissance du sens des racines	0,44	[-0,57, 1,46]	0,374	21,98	0,62	[-0,20, 1,44]	0,131	21,98
Analyse morphologique	-0,64	[-3,77, 2,49]	0,677	22,02	-0,17	[-2,70, 2,37]	0,891	22,01
Effets aléatoires								
Résidus σ^2	1290,21				1101,82			
Variance (constante)	398,11				258,47			
<i>n</i>	29				29			
ICC	0,22				0,20			
Observations	18090				18630			

Note. IC 95 % = intervalle de confiance à 95 %.

En ce qui concerne les effets des facteurs linguistiques sur l'amplitude et la latence aux deux fenêtres temporelles ciblées, de 200 ms à 350 ms et de 350 ms à 500 ms, aucun facteur au-delà de ceux manipulés dans la tâche expérimentale (sections 4.3.1, 4.3.2, 4.3.3, 4.3.4) n'a révélé d'effets significatifs ($p > 0,05$).

Pour résumer, les résultats permettent de conclure que les indices de fréquence ont un effet sur les données comportementales issues de la tâche de décision lexicale. Plus particulièrement, la fréquence du mot entier a un effet sur la précision et le temps de réaction, un indice de la vitesse de traitement des mots écrits. Plus le mot est fréquent, plus les participants ont de chance de fournir une bonne réponse et plus ils répondent rapidement. Pour le traitement des pseudomots et des non-mots, c'est plutôt la fréquence des racines qui exerce un effet sur la précision et le temps de réaction. La fréquence de la 1^{re} racine influence le temps de réaction des participants : plus la fréquence est élevée, plus la réponse est rapide. La fréquence de la 2^e racine, quant à elle, influence la précision des participants : plus la fréquence est élevée, plus les participants ont de chance de fournir une bonne réponse. Pour terminer, les facteurs linguistiques ciblés dans les régressions multiniveaux ne permettent pas d'expliquer les résultats des variables dépendantes recueillies par des mesures psychophysiques (amplitude et latence de 200 ms à 350 ms et de 350 ms à 500 ms) au-delà des manipulations déjà effectuées lors de la tâche expérimentale de décision lexicale (catégories d'items et régions topographiques).

4.4 L'influence des facteurs individuels sur le traitement des mots écrits

Des facteurs individuels ont aussi été mesurés pour vérifier leur effet sur le traitement des mots écrits, soit la précision, le temps de réaction ainsi que l'amplitude et la latence de 200 ms à 350 ms et de 350 ms à 500 ms. Ces facteurs ont d'ailleurs été inscrits dans les mêmes analyses que celles effectuées pour vérifier l'influence de la fréquence des mots entiers et de celle des racines sur le traitement des mots écrits (section 4.3.5). Les facteurs individuels ciblés dans la présente étude concernent les données recueillies au test dynamique : la connaissance du vocabulaire, la reconnaissance des racines latines et grecques, l'habileté de décomposition morphologique, la connaissance du sens des racines latines et grecques ainsi que l'habileté d'analyse morphologique.

D'abord, contrairement aux résultats du test t (section 4.1), les résultats indiquent que le niveau scolaire des participants n'influence aucune variable dépendante insérée dans les modèles de régression multiniveaux lorsque les autres variables indépendantes sont considérées dans le modèle, ni pour les

mots ($p > 0,05$) ni pour les pseudomots et les non-mots ($p > 0,05$) (Tableau 4.36, Tableau 4.37, Tableau 4.38, Tableau 4.39, Tableau 4.40, Tableau 4.41).

En ce qui concerne la précision en lecture, les résultats du modèle de régression logistique multiniveaux indiquent que le seul facteur individuel à intervenir dans la relation à l'étude est l'habileté à décomposer le mot en morphèmes, et ce, uniquement pour les pseudomots et les non-mots (rapport de cote = 0,65, $p = 0,008$) (Tableau 4.36). Le rapport de cote permet d'inférer que les participants ont 0,65 fois moins de chance de fournir une réponse correcte pour les pseudomots et les non-mots si leur habileté de décomposition morphologique du test augmente d'une valeur de 1, ce qui correspond à une diminution de 0,65 %. On pourrait donc penser que de meilleures habiletés en décomposition morphologique induiraient les participants en erreur lorsqu'ils rencontrent des pseudomots puisque ces items sont morphologiquement plausibles.

Ensuite, les résultats des analyses de régression linéaires multiniveaux montrent qu'aucun facteur individuel mesuré n'influence le temps de réaction (Tableau 4.37), l'amplitude (Tableau 4.38) et la latence de 200 ms à 350 ms (Tableau 4.39) ou encore l'amplitude (Tableau 4.40) et la latence de 350 ms à 500 ms (Tableau 4.41), et ce, autant pour les mots que les pseudomots et les non-mots ($p > 0,05$).

Pour résumer, l'habileté des participants à décomposer les mots et pseudomots polymorphémiques en morphèmes nuit à la précision des participants lors du traitement des pseudomots et des non-mots uniquement. Enfin, les facteurs individuels ciblés dans les régressions multiniveaux ne permettent pas d'expliquer les résultats des variables dépendantes recueillies par des mesures psychophysiques (amplitude et latence de 200 ms à 350 ms et de 350 ms à 500 ms) ni ceux relatifs au temps de réaction des participants.

4.5 Résumé des résultats de recherche

Du point de vue comportemental, chacun des effets des facteurs linguistiques mesurés influence la précision des participants lors de la tâche de décision lexicale. Dans cette tâche, les participants ont répondu plus précisément aux non-mots et aux pseudomots qu'aux mots. La structure morphologique a aussi un effet sur la précision en lecture : les items monomorphémiques entraînent moins d'erreurs que les items polymorphémiques, autant pour les mots que les non-mots et les pseudomots. Le type de racine a aussi un effet sur la précision à la tâche de décision lexicale, mais seulement pour les

pseudomots et les non-mots. En effet, pour les mots, le type de racine n'a pas d'effets au-delà de la structure morphologique. Les mots polymorphémiques sont moins réussis que les mots monomorphémiques, qu'ils soient formés de racines libres ou liées. Néanmoins, pour les non-mots et les pseudomots, ces derniers, lorsqu'ils sont formés de racines liées entraînent le plus d'erreurs de la part des participants. Au sein des items formés de racines liées, l'origine des racines influence aussi la précision, mais différemment pour les mots et les pseudomots. Les mots formés de racines grecques sont mieux réussis que ceux formés de racines latines. Pour les pseudomots, ce sont ceux formés de racines latines qui sont mieux réussis. En ce qui concerne la fréquence, les mots fréquents sont identifiés plus précisément que les mots rares. La fréquence de la deuxième racine, quant à elle, influence le traitement des pseudomots : plus la deuxième racine est fréquente, meilleur est le participant à rejeter le pseudomot durant la tâche de décision lexicale. En ce qui concerne les facteurs individuels, seule l'habileté à décomposer les items en morphèmes influence la précision à la tâche de décision lexicale, mais seulement pour les pseudomots. Effectivement, les élèves qui ont de plus fortes habiletés en décomposition morphologique commettent plus d'erreurs avec les pseudomots à la tâche de décision lexicale que les élèves qui ont de plus faibles habiletés en décomposition morphologique. En résumé, la précision des participants à la tâche de décision lexicale est influencée par l'effet de lexicalité, l'effet de la structure morphologique des items, l'effet du type de racine, l'effet de l'origine des racines, l'effet de la fréquence du mot entier, l'effet de la fréquence de la deuxième racine ainsi que les habiletés de décomposition morphologique des élèves.

L'effet des différentes variables linguistiques et individuelles sur le temps de réaction des participants à la tâche de décision lexicale a aussi été mesuré. D'abord, la lexicalité entraîne un grand effet sur le temps de réaction : les participants répondent plus rapidement aux pseudomots et aux non-mots qu'aux mots. Aussi, l'effet de structure morphologique influence le temps de réaction. Les participants répondent significativement plus rapidement aux mots monomorphémiques et aux non-mots qu'aux items polymorphémiques, qu'ils soient des mots ou des pseudomots. Cependant, aucun effet de type de racine n'a été trouvé. Parmi les items formés de racines latines et grecques, l'origine des racines a un effet sur le temps de réaction: les participants répondent plus rapidement aux items formés de racines latines qu'à ceux formés de racines grecques, autant pour les mots que pour les pseudomots. Enfin, la fréquence du mot entier permet une identification plus rapide des mots tandis que la fréquence de la première racine favorise le temps de réaction des participants aux pseudomots. Aucun facteur individuel n'exerce d'influence sur le temps de réaction des participants. Donc, le temps de réaction des

participants à la tâche de décision lexicale, indice de la vitesse de traitement des mots écrits, est influencé par l'effet de lexicalité, l'effet de la structure morphologique, l'effet de l'origine des racines ainsi que les effets de la fréquence du mot entier de la fréquence de la première racine.

Au regard des données psychophysiques recueillies de 200 ms à 350 ms poststimulus, la lexicalité a un effet sur le traitement des mots écrits. Effectivement, les mots génèrent une plus grande négativité en région postérieure droite que les pseudomots et les non-mots. Aussi, un pic d'amplitude négative est détecté significativement plus tardivement en région antérieure médiane pour les mots que pour les pseudomots et les non-mots. La structure morphologique des items a aussi un effet important sur le traitement des mots écrits. Effectivement, il a été trouvé que la négativité générée était plus grande pour les items monomorphémiques que polymorphémiques. Pour les mots, cette différence est observable en régions postérieures gauche et médiane. Pour les non-mots et les pseudomots, on constate une différence plutôt dans les trois régions d'intérêt antérieures ainsi que la région postérieure gauche. La structure morphologique des items influence aussi la latence de l'amplitude négative maximale de 200 ms à 350 ms: elle est généralement plus tardive pour les items monomorphémiques que polymorphémiques en région postérieure médiane. Aucun effet du type des racines, de l'origine des racines, de fréquence ou des différents facteurs individuels sur l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms n'a été confirmé.

De 350 ms à 500 ms, aucun effet de lexicalité sur l'amplitude ou la latence des potentiels évoqués n'a été détecté. La structure morphologique des items n'a pas non plus d'effet sur l'amplitude des potentiels évoqués, mais le pic d'amplitude maximale est plus tardif pour les items polymorphémiques que monomorphémiques en région postérieure médiane, autant pour les mots que pour les non-mots et les pseudomots. Enfin, l'amplitude et la latence des potentiels évoqués de 350 ms à 500 ms ne sont influencées ni par l'effet du type de racine, ni par l'effet de l'origine des racines, ni par les effets de fréquence, ni par les différents facteurs individuels mesurés.

Les tableaux suivants résument les principaux résultats permettant de vérifier les effets des facteurs linguistiques (lexicalité, structure morphologique des items, type de racine, origine des racines, fréquence des mots et fréquence des racines) et individuels (connaissance du vocabulaire, reconnaissance des racines latines et grecques, décomposition morphologique, connaissance du sens des racines latines et grecques et analyse morphologique) sur la précision (Tableau 4.42), sur le temps de

réaction (Tableau 4.43), sur l'amplitude des potentiels évoqués de 200 ms à 350 ms (Tableau 4.44) et de 350 ms à 500 ms (Tableau 4.45) ainsi que sur la latence des potentiels évoqués de 200 ms à 350 ms (Tableau 4.46) et de 350 ms à 500 ms (Tableau 4.47).

Tableau 4.42 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la précision des participants à la tâche de décision lexicale

Variables indépendantes		Résultats sur la précision					
Facteurs linguistiques							
Lexicalité		Mots			<	Pseudomots et non-mots	
Structure morphologique	Mono	>	Poly		Mono	>	Poly
Type de racine	Libres	=	Liées		Libres	>	Liées
Origine des racines	Latines	<	Grecques		Latines	>	Grecques
Fréquence du mot	Fréquent	>	Rare				
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	=	Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	>	Rare
Facteurs individuels ^a							

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur la précision (à l'exception des habiletés de décomposition morphologique qui mènent à une moins bonne précision pour les pseudomots et non-mots), ceux-ci ne sont pas répétés dans ce tableau.

Tableau 4.43 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur le temps de réaction des participants à la tâche de décision lexicale

Variables indépendantes		Résultats sur le temps de réaction					
Facteurs linguistiques							
Lexicalité		Mots			<	Pseudomots et non-mots	
Structure morphologique	Mono	<	Poly		Mono	<	Poly
Type de racine	Libres	=	Liées		Libres	=	Liées
Origine des racines	Latines	<	Grecques		Latines	<	Grecques
Fréquence du mot	Fréquent	<	Rare				
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	<	Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	=	Rare
Facteurs individuels ^a							

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur le temps de réaction, ceux-ci ne sont pas répétés dans ce tableau.

Tableau 4.44 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur l'amplitude des potentiels évoqués de 200 ms à 350 ms à la tâche de décision lexicale

Variables indépendantes		Résultats sur l'amplitude de 200 ms à 350 ms				
Facteurs linguistiques						
Lexicalité		Mots		> ^b	Pseudomots et non-mots	
Structure morphologique	Mono	> ^c	Poly		Mono	> ^d Poly
Type de racine	Libres	=	Liées		Libres	= Liées
Origine des racines	Latines	=	Grecques		Latines	= Grecques
Fréquence du mot	Fréquent	=	Rare			
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	= Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	= Rare
Facteurs individuels ^a						

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur l'amplitude des potentiels évoqués de 200 ms à 350 ms, ceux-ci ne sont pas répétés dans ce tableau. ^b Région postérieure droite.

^c Régions postérieure gauche et postérieure médiane. ^d Régions antérieure gauche, antérieure médiane, antérieure droite et postérieure gauche.

Tableau 4.45 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur l'amplitude des potentiels évoqués de 350 ms à 500 ms à la tâche de décision lexicale

Variables indépendantes		Résultats sur l'amplitude de 350 ms à 500 ms				
Facteurs linguistiques						
Lexicalité		Mots		=	Pseudomots et non-mots	
Structure morphologique	Mono	=	Poly		Mono	= Poly
Type de racine	Libres	=	Liées		Libres	= Liées
Origine des racines	Latines	>	Grecques		Latines	> Grecques
Fréquence du mot	Fréquent	=	Rare			
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	= Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	= Rare
Facteurs individuels ^a						

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur l'amplitude des potentiels évoqués de 350 ms à 500 ms, ceux-ci ne sont pas répétés dans ce tableau.

Tableau 4.46 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la latence des potentiels évoqués de 200 ms à 350 ms à la tâche de décision lexicale

Variables indépendantes		Résultats sur la latence de 200 ms à 350 ms				
Facteurs linguistiques						
Lexicalité		Mots		> ^b	Pseudomots et non-mots	
Structure morphologique	Mono	> ^c	Poly		Mono	> ^c Poly
Type de racine	Libres	=	Liées		Libres	= Liées
Origine des racines	Latines	=	Grecques		Latines	= Grecques
Fréquence du mot	Fréquent	=	Rare			
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	= Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	= Rare
Facteurs individuels ^a						

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur la latence des potentiels évoqués de 200 ms à 350 ms, ceux-ci ne sont pas répétés dans ce tableau. ^b Région antérieure médiane.

^c Région postérieure médiane.

Tableau 4.47 Résumé des principaux résultats de l'étude concernant les effets des facteurs linguistiques et individuels sur la latence des potentiels évoqués de 350 ms à 500 ms à la tâche de décision lexicale

Variables indépendantes		Résultats sur la latence de 350 ms à 500 ms				
Facteurs linguistiques						
Lexicalité		Mots		=	Pseudomots et non-mots	
Structure morphologique	Mono	< ^b	Poly		Mono	< ^b Poly
Type de racine	Libres	=	Liées		Libres	= Liées
Origine des racines	Latines	=	Grecques		Latines	= Grecques
Fréquence du mot	Fréquent	=	Rare			
Fréquence de la 1 ^{re} racine	Fréquente	=	Rare		Fréquente	= Rare
Fréquence de la 2 ^e racine	Fréquente	=	Rare		Fréquente	= Rare
Facteurs individuels ^a						

Note. Mono = monomorphémiques; Poly = polymorphémiques.

^a Comme les résultats montrent que les facteurs individuels n'exercent aucune influence sur la latence des potentiels évoqués de 350 ms à 500 ms, ceux-ci ne sont pas répétés dans ce tableau. ^b Région postérieure médiane.

CHAPITRE 5

DISCUSSION

La présente thèse doctorale a pour objectif général de vérifier comment les élèves du 3^e cycle du primaire traitent les mots polymorphémiques du vocabulaire disciplinaire formés de racines latines et grecques : les décomposent-ils en morphèmes avant l'accès lexical ou accèdent-ils à leur sens dans leur globalité? Plus spécifiquement, la présente étude permet de mesurer l'effet des facteurs linguistiques (lexicalité, structure morphologique des items, type de racine, origine des racines, fréquence du mot, fréquence des racines) sur la précision, le temps de réaction ainsi que l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus lors de l'identification des mots écrits (5.1). Les résultats indiquent des effets de la lexicalité, de la structure morphologique des items, du type de racine et de l'origine des racines sur la précision et le temps de réaction à la tâche de décision lexicale. Pour l'amplitude et la latence, principalement de 200 ms à 350 ms, seuls des effets de lexicalité et de structure morphologique ont été détectés. La précision et le temps de réaction lors de l'identification des mots étaient influencés par la fréquence des mots, alors que, lors de l'identification de pseudomots, ces deux variables étaient plutôt influencées par la fréquence des racines. Par la suite, l'effet des facteurs individuels (connaissance du vocabulaire, reconnaissance des racines latines et grecques dans un mot, sens des racines latines et grecques, habileté à décomposer les mots polymorphémiques sur la base de leurs morphèmes et analyse morphologique) sur la précision, le temps de réaction ainsi que l'amplitude et la latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus lors de l'identification des mots écrits a été mesuré (5.2). Les résultats montrent que la décomposition morphologique est la seule variable individuelle qui influence négativement la précision en lecture des pseudomots et non-mots. Enfin, ce chapitre se termine par une proposition d'inclure le traitement des mots composés formés de racines latines et grecques dans les modèles de traitement des mots polymorphémiques (5.3).

5.1 L'influence des facteurs linguistiques sur le traitement des mots écrits

Dans cette section, pour répondre à la première question de recherche, les résultats relatifs à l'effet de la lexicalité (mots vs pseudomots et non-mots) (5.1.1), à la structure morphologique des items (polymorphémiques vs monomorphémiques) (5.1.2), au type de racine (libre vs liée) (5.1.3), à l'origine des racines (latine vs grecque) (5.1.4) ainsi qu'à la fréquence des mots entiers et des racines (5.1.5) lors de l'identification des mots écrits sur le temps de réaction et la précision ainsi que sur l'amplitude et la

latence des potentiels évoqués de 200 ms à 350 ms et de 350 ms à 500 ms lors d'une tâche de décision lexicale seront interprétés successivement.

5.1.1 L'effet de la lexicalité sur le traitement des mots écrits

À la lumière des recherches antérieures, l'hypothèse émise voulait qu'un effet de la lexicalité soit observé dans les résultats de la présente étude. Il était attendu que les mots soient traités plus rapidement et avec plus de précision que les pseudomots et les non-mots. Aussi, une négativité plus grande pour les pseudomots et les non-mots était prévue, autant dans la fenêtre temporelle de 200 ms à 350 ms que de 350 ms à 500 ms. Ces résultats auraient marqué respectivement une tentative de décomposition morpho-orthographique précoce des pseudomots polymorphémiques puis l'incapacité à accéder au sens du pseudomot ou du non-mot présenté.

Les résultats des analyses de variance à mesures répétées montrent des effets de la lexicalité sur les données comportementales et psychophysiologiques extraites de la tâche de décision lexicale. Des différences de traitement ont été observées entre les mots et les non-mots, incluant les pseudomots, quant à la précision, le temps de réaction ainsi que l'amplitude et la latence dès 200 ms poststimulus. Plus précisément, les résultats suggèrent que les élèves du 3^e cycle du primaire réussissent mieux les pseudomots et les non-mots que les mots. Ainsi, lors de la présentation d'un pseudomot ou d'un non-mot, ils répondent plus souvent que ces suites de lettres ne forment pas des mots en français que l'inverse. En contrepartie, lorsqu'on présente un mot aux élèves, ils tendent à faire plus d'erreurs en répondant que ce mot n'existe pas en français. Concernant le temps de réaction, les participants sont plus rapides avec les mots, lorsqu'ils les identifient correctement, qu'avec les pseudomots ou les non-mots. De ce fait, les élèves du 3^e cycle du primaire commettent moins d'erreurs à la tâche de décision lexicale avec les pseudomots et les non-mots, mais leur traitement est plus lent que les mots, et ce, même en contrôlant pour leur main dominante. Pour les résultats des mesures psychophysiologiques, les résultats ont révélé un effet de lexicalité pour l'amplitude et la latence de 200 ms à 350 ms : les mots entraînent une plus grande négativité en région postérieure droite que les pseudomots et les non-mots et la latence est significativement plus tardive pour les mots que les pseudomots et les non-mots. Ce traitement prélexical pourrait être associé à la décomposition morpho-orthographique des mots avant l'accès lexical (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008). Néanmoins, aucun effet de la lexicalité n'a été détecté significatif pour l'amplitude et la latence dans la fenêtre temporelle de

350 ms à 500 ms, ce qui veut dire qu'il n'y aurait pas de différence observée quant à l'accès lexical entre les mots choisis pour la tâche de décision lexicale ainsi que les pseudomots et les non-mots.

Concernant l'effet de lexicalité et son influence sur la précision des participants à la tâche de décision lexicale, les résultats corroborent ceux de recherches antérieures en italien et en néerlandais auprès de participants adultes qui ont observé que les pseudomots et les non-mots étaient mieux réussis que les mots (El Yagoubi *et al.*, 2008; Kuperman *et al.*, 2009). En contrepartie, Coch et ses collaborateurs (2013; 2020), qui travaillent sur des items anglais formés de racines liées (p. ex. *discern*, **disject*), n'avaient trouvé aucun effet de lexicalité dans leur première recherche (2013). Leur deuxième recherche (Coch *et al.*, 2020) indique toutefois un effet de lexicalité sur la précision, mais en faveur des mots. Dans leur recherche, les participants adultes étaient plus précis pour identifier les mots que les pseudomots et les non-mots. La divergence des résultats entre les résultats de la présente étude ainsi que ceux des recherches antérieures (Coch *et al.*, 2013; Coch *et al.*, 2020) peut être expliquée par le fait que la plupart des mots de la tâche de décision lexicale de la présente étude étaient rares, par conséquent susceptibles d'être inconnus des élèves. C'est aussi une explication proposée par Kuperman et ses collaborateurs (2009) dans le cadre de leur propre étude. D'ailleurs, les résultats de la présente étude montrent que les élèves réussissent mieux à identifier les mots plus fréquents que rares. Les fréquences moyennes des quatre catégories de mots sont à moins de 50 occurrences, même que plusieurs mots affichent une fréquence nulle. C'est donc dire que les élèves rencontrent ces mots moins de 50 fois dans leurs manuels scolaires durant toute leur scolarité au primaire. Pour les mots formés de racines latines et grecques, les fréquences diminuent respectivement à 16 et à 18 occurrences. Par conséquent, puisque les mots sont rares, les élèves peineraient à les reconnaître lors de la tâche de décision lexicale et répondraient, à tort, que le mot présenté n'existe pas. Pour les pseudomots et les non-mots, l'absence d'occurrence étant déjà un critère intrinsèque à cette catégorie, la décision lexicale des élèves pencherait aussi plus souvent vers la non-reconnaissance de l'item présenté, ce qui, ici, entrainerait une bonne réponse. Ainsi, la fréquence des mots expliquerait que les élèves commettent plus d'erreurs avec les mots qu'avec les pseudomots et les non-mots. Il est aussi important de considérer que les participants connaissent probablement moins de mots que les participants adultes des recherches antérieures, ce qui peut expliquer le taux d'erreurs plus élevé.

Pour continuer, le temps de réaction des participants à fournir une réponse correcte à la tâche de décision lexicale est influencé par l'effet de la lexicalité. L'hypothèse de départ est ainsi confirmée : les

participants fournissent plus rapidement une décision lexicale pour les mots que pour les pseudomots et les non-mots. Ces résultats corroborent les résultats d'études antérieures qui ont observé des temps de réponse plus longs pour les non-mots, notamment pour les pseudomots, que pour les mots (Coch *et al.*, 2013; Coch *et al.*, 2020; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014; Kuperman *et al.*, 2009). Comme il sera discuté plus loin (section 5.2), les résultats de la présente étude ont aussi montré que le traitement des mots était plus rapide plus les mots étaient fréquents. Les mots étant, par définition, plus fréquents que les pseudomots et les non-mots, ils sont traités plus rapidement par les participants. Ces résultats pourraient être expliqués grâce aux modèles hybrides (p. ex. Caramazza *et al.*, 1985) qui prévoient une étape supplémentaire de décomposition morphologique pour les items polymorphémiques transparents avant d'accéder au sens du mot entier. Si le mot est fréquent, il serait traité par la voie globale qui ne nécessite aucune décomposition morphologique et qui est donc plus rapide. On pourrait donc penser que les élèves du 3^e cycle du primaire décomposent les items polymorphémiques avant de rendre leur décision lexicale, plus particulièrement les pseudomots puisqu'il est impossible pour les élèves d'accéder à la voie globale pour ce type d'item. Cela entraînerait un traitement plus lent en ajoutant une étape intermédiaire à l'accès lexical. Comme la fréquence des mots est plus élevée que celle des pseudomots et des non-mots, il est possible que plusieurs mots empruntent plutôt la voie globale, ce qui entraînerait un traitement plus rapide.

Aussi, à la lumière des résultats des recherches antérieures (El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014), il a été surprenant de constater que les résultats de la présente étude suggèrent une négativité plus importante pour les mots que pour les pseudomots et les non-mots en région postérieure droite de 200 ms à 350 ms poststimulus, ce qui pourrait signifier que les mots sont décomposés avant l'accès lexical contrairement aux pseudomots et aux non-mots. Ces résultats ne permettent pas de corroborer ceux des études antérieures qui avaient plutôt trouvé une négativité plus grande pour les non-mots et les pseudomots que pour les mots, et ce, dans toutes les régions d'intérêt ciblées de 200 ms à 350 ms (El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014), suggérant que les pseudomots et les non-mots étaient décomposés en morphèmes avant l'accès lexical. Comme les résultats positionnent un effet de lexicalité en région postérieure droite de 200 ms à 350 ms, cela suscite des questionnements quant au marqueur de potentiel évoqué qui pourrait y être attribué, donc sur l'interprétation de cet effet. S'il s'agissait du N250 témoignant de la décomposition morpho-orthographique des mots écrits, l'effet aurait plutôt été attendu en régions antérieures centrale et gauche. Aussi, certains auteurs segmentent le marqueur N400 témoignant de l'activation du sens d'un mot et sensible aux anomalies sémantiques. Effectivement, la

fenêtre temporelle de 250 ms à 450 ms pourrait être associée à un N400 précoce (Beyersmann *et al.*, 2014), d'autant plus que la plus grande négativité de ce marqueur est habituellement observée en région postérieure médiane, ce qui est très près de ce que nos résultats rapportent. Ainsi, il n'est pas clair si l'effet de lexicalité détecté de 200 ms à 350 ms chez les participants est associé au marqueur N250 de décomposition morpho-orthographique ou au marqueur N400 précoce d'accès lexical. Comme les pseudomots de la présente étude sont formés des mêmes morphèmes que les mots, il aurait été intéressant de vérifier l'effet de la lexicalité séparément pour les pseudomots et les non-mots. Effectivement, il est possible que les pseudomots soient décomposés au même titre que les mots, mais que l'absence de décomposition qui pourrait être observée pour les non-mots ne contenant aucun morphème vienne influencer de façon importante la moyenne d'amplitude lorsque les pseudomots et les non-mots sont observés ensemble.

Le modèle d'activation parallèle lexicomorphologique (MAPL) [ma traduction de *Augmented Addressed Morphology* (Caramazza *et al.*, 1985)] pourrait fournir une piste d'explication valable pour l'effet de lexicalité détecté de 200 ms à 350 ms. Effectivement, comme les mots qui ont été présentés durant la tâche de décision lexicale sont relativement rares, la voie de décomposition morphologique serait plus rapide que la voie globale, qui elle, serait inefficace dû à la rareté du mot. De ce fait, on peut émettre l'hypothèse que les élèves, ayant déjà vu les mots polymorphémiques proposés ou les morphèmes constituants, décomposent les mots en morphèmes afin d'accéder au sens de ces mots. En contrepartie, il est possible que les pseudomots et les non-mots, particulièrement les non-mots qui ne sont pas formés de morphèmes, soient identifiés correctement à la tâche de décision lexicale sans recourir à la décomposition morphologique. Ce processus cognitif générerait une négativité plus grande au marqueur N250 pour les mots que pour les pseudomots et les non-mots, comme c'est le cas dans les résultats de la présente étude. Dans tous les cas, de futures recherches devraient être menées en tentant de mieux départager ces deux fenêtres temporelles contigües.

De plus, aucun effet de lexicalité n'a été détecté sur l'amplitude et la latence de 350 ms à 500 ms, fenêtre temporelle communément attribuée au marqueur de potentiel évoqué N400 sensible aux anomalies sémantiques, donc au traitement des pseudomots et des non-mots (Kutas et Federmeier, 2011). Il aurait aussi été attendu à ce que les pseudomots et les non-mots produisent une négativité plus grande que les mots de 350 ms à 500 ms (Coch *et al.*, 2013; El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014). Or, aucune différence significative n'a été observée entre les mots ainsi que les pseudomots et les

non-mots malgré une taille d'effet moyenne, $\eta^2p = 0,07$, même si les tracés des potentiels évoqués aux six régions topographiques d'intérêt semblent montrer un pic d'amplitude plus élevé pour les pseudomots et les non-mots que pour les mots en régions antérieures (Figure 4.2). Plusieurs facteurs pourraient expliquer cette absence de significativité (et les résultats similaires dans les sections subséquentes pour les données psychophysologiques): le manque de puissance statistique; une grande variabilité entre les participants; le moyennage de plusieurs électrodes en régions d'intérêt qui aurait pu effacer des effets plus localisés; ou encore l'utilisation des mauvaises fenêtres temporelles qui n'a pas permis de détecter d'effets aux moments ciblés. Le fait que les pseudomots et les non-mots aient été analysés dans une même catégorie peut aussi expliquer les résultats. En effet, en plus d'être sensible aux anomalies sémantiques et ainsi présenter une plus grande négativité lors du traitement des pseudomots et des non-mots (Coch *et al.*, 2013; McKinnon *et al.*, 2003), le marqueur N400 est aussi sensible à la fréquence des mots, générant de plus grandes négativités lors de la lecture de mots rares (Luck, 2014; Vergara-Martínez *et al.*, 2009). De plus, il a aussi été reconnu que le marqueur N400 serait un indice de composition sémantique qui surviendrait après la décomposition morpho-orthographique (Coch *et al.*, 2013). Ainsi, les deux catégories d'items, mots ainsi que pseudomots et non-mots, pourraient générer des négativités importantes similaires, mais pour des raisons différentes, ce qui empêche de distinguer des différences entre ces deux amplitudes.

Il convient de rappeler que les recherches antérieures ont toutes été menées auprès de participants adultes non-francophones : aucune n'a été menée auprès d'enfants dont les compétences langagières sont encore en développement. De ce fait, il est difficile de savoir si les résultats contradictoires observées entre ces études et la nôtre sont liés au matériel de collecte de données utilisé, à l'âge des participants, à une différence dans le décours temporel des processus cognitifs observés ou encore à la langue des participants.

Pour résumer, l'identification des mots écrits par les lecteurs intermédiaires de 3^e cycle du primaire est influencée par la lexicalité dès 200 ms poststimulus avec une négativité plus importante pour les mots que pour les pseudomots et les non-mots qui pourrait témoigner d'une décomposition morpho-orthographique des mots. La négativité moins importante pour les pseudomots et les non-mots dans cette fenêtre temporelle pourrait être attribuable au traitement des non-mots. Il est possible que les pseudomots soient aussi décomposés lors de l'identification des mots, mais que les effets spécifiques à ce type d'items aient potentiellement été effacés par l'agrégation aux non-mots. D'ailleurs, le

traitement des mots est plus rapide, mais moins précis, que le traitement des pseudomots et des non-mots. Cela laisse envisager un traitement plus long pour les pseudomots et les non-mots qui pourrait être modulé par le traitement des pseudomots.

5.1.2 L'effet de la structure morphologique des items sur le traitement des mots écrits

Concernant l'effet de la structure morphologique des mots, les hypothèses soulevées quant aux données comportementales de précision et de temps de réaction étaient nuancées puisque les résultats des recherches antérieures étaient contradictoires. Basée sur les résultats d'El Yagoubi et ses collaborateurs (2008), la première hypothèse supposait que les mots monomorphémiques seraient traités plus rapidement et plus précisément que les mots polymorphémiques. Or, partant des résultats de Fiorentino et ses collègues (2014) et de Fiorentino et Poeppel (2007), nous avons émis une seconde hypothèse qui postulait que les mots polymorphémiques seraient traités plus rapidement et plus précisément que les mots monomorphémiques. Les résultats de la présente étude révèlent que les élèves du 3^e cycle du primaire réussissent mieux à identifier les items monomorphémiques que les items polymorphémiques, qu'ils soient ou non des mots qui existent en français. Aussi, non seulement les items polymorphémiques sont moins bien réussis, mais ils sont aussi traités plus lentement que les items de contrôle, soit les mots monomorphémiques et les non-mots. Nos résultats corroborent les résultats d'El Yagoubi et ses collaborateurs (2008), ce qui nous mène à rejeter la seconde hypothèse. De surcroît, les hypothèses de départ quant aux données psychophysiques sont réfutées. À la lumière des travaux d'El Yagoubi et ses collaborateurs (2008) et de Fiorentino et ses collaborateurs (2014), il était attendu que les mots polymorphémiques produisent une plus grande négativité que les mots monomorphémiques de 200 ms à 350 ms, ce qui aurait témoigné d'une décomposition morpho-orthographique pour les mots polymorphémiques. Or, les résultats permettent de constater que les items de contrôle (mots monomorphémiques et non-mots) entraînent une plus grande négativité que les items polymorphémiques composés de 200 ms à 350 ms, fenêtre temporelle habituellement associée à la décomposition morpho-orthographique prélexicale. Pour les mots, cette différence est localisée en régions postérieures gauche et médiane. Pour les pseudomots et les non-mots, cette différence est plutôt observée dans les trois régions antérieures (gauche, médiane et droite) et en région postérieure gauche. Aussi, l'hypothèse d'une interaction entre les variables de lexicalité et de structure avait été émise en énonçant que les pseudomots polymorphémiques généreraient une négativité plus grande de 350 ms à 500 ms que les non-mots monomorphémiques (Fiorentino *et al.*, 2014). Cet effet aurait laissé présager que les lecteurs perçoivent l'anomalie sémantique lors de la lecture de pseudomots et de non-

mots et que celle-ci est plus importante lors de la lecture de pseudomots compte tenu de leur caractère plausible. Toutefois, les résultats obtenus réfutent cette hypothèse puisqu'aucun effet de structure sur l'amplitude n'a été détecté dans la fenêtre temporelle de 350 ms à 500 ms poststimulus. Cependant, pour la latence de l'amplitude maximale en région postérieure médiane, les résultats montrent une différence entre les items de contrôle et les items polymorphémiques dans les deux fenêtres temporelles. Dans la fenêtre temporelle de 200 ms à 350 ms, l'amplitude maximale serait plus précoce pour les items polymorphémiques que les items de contrôle alors que, dans la fenêtre suivante, de 350 ms à 500 ms, les items de contrôle qui seraient traités avant les items polymorphémiques. Ces résultats sont interprétés plus loin.

D'abord, les résultats de la présente étude ont permis de détecter que la précision des participants à la tâche de décision lexicale était influencée par la structure morphologique des mots, les items de contrôle étant mieux réussis que les items polymorphémiques. Ces résultats vont dans le même sens que ceux d'El Yagoubi et ses collaborateurs (2008) ayant découvert que les mots italiens non composés (p. ex. *cocodrillo*, *crocodile* en français) étaient mieux réussis par les participants adultes que les mots italiens composés (p. ex. *astronave*, *astronef* en français). Sachant que les résultats de la présente étude quant à la précision ont été influencés par la fréquence des mots (voir section 5.2), il est possible que la fréquence des items présentés ait affecté l'effet de structure observé. Effectivement, étant donné les nombreux critères établis pour sélectionner les mots de la tâche de décision lexicale, il était impossible de sélectionner des mots à fréquence similaire dans toutes les catégories. À cet effet, les mots monomorphémiques sont plus fréquents ($M = 45,67$, $ÉT = 71,83$) que les mots composés des trois autres catégories : formés de racines liées latines ($M = 16,00$, $ÉT = 30,92$), formés de racines liées grecques ($M = 18,20$, $ÉT = 37,08$), formés de racines libres ($M = 31,70$, $ÉT = 69,66$). Pour Fiorentino et Poeppel (2007), les résultats sont différents : les items composés et les non-mots sont mieux réussis que les items monomorphémiques. Ces différences avec les résultats de la présente étude pourraient être expliquées par le fait que le procédé de formation des mots le plus productif peut différer selon les langues étudiées. Par exemple, la composition est plus fréquente en anglais qu'en français (Casalis *et al.*, 2015), le français étant plutôt marqué par la dérivation (Casalis *et al.*, 2015). De plus, les comparaisons de la présente étude ne prévoyaient pas de distinction entre les mots monomorphémiques et les non-mots. Il est donc impossible de savoir, avec l'analyse effectuée, si les mots monomorphémiques et les non-mots sont traités différemment.

Le temps de réaction des participants est aussi influencé par l'effet de la structure morphologique des mots, les items monomorphémiques étant identifiés plus rapidement que les items polymorphémiques. El Yagoubi et ses collègues (2008) en sont arrivés à la même conclusion : en italien, les participants adultes répondent plus rapidement aux items non composés (p. ex. *cocodrillo*, *crocodile* en français) qu'aux items composés (p. ex. *astronave*, *astronef* en français). Cependant, les résultats des recherches de Fiorentino et Poeppel (2007) ainsi que Fiorentino et ses collègues (2014) sont différents de ceux de la présente étude. Selon ces auteurs, les mots composés sont ceux qui seraient traités les plus rapidement suivis par les mots monomorphémiques et les non-mots (Fiorentino *et al.*, 2014; Fiorentino et Poeppel, 2007). Néanmoins, ayant aussi observé le traitement des pseudomots composés, Fiorentino et ses collègues (2014) ont trouvé que cette catégorie d'items était celle qui était traitée la plus lentement. Comme la composition est le procédé de formation principal des mots en anglais (Casalis *et al.*, 2015), il n'est pas étonnant de constater que le traitement de ces items est plus rapide. Effectivement, à la lumière des modèles hybrides (p. ex. Caramazza *et al.*, 1985), il est possible que les items composés présentés soient plus fréquents et familiers pour les participants et qu'ils soient plutôt traités par la voie globale que par la voie de décomposition morphologique, alors que les pseudomots, ayant des racines connues seraient décomposés.

Pour l'effet de structure morphologique, à la lumière des recherches antérieures, il était attendu que les items polymorphémiques entraînent une négativité plus grande que celle des mots monomorphémiques ou des non-mots dans la fenêtre temporelle de 200 ms à 350 ms, témoignant d'une décomposition morpho-orthographique des items polymorphémiques composés (El Yagoubi *et al.*, 2008; Fiorentino *et al.*, 2014). Les résultats obtenus dans la présente étude montrent plutôt l'inverse, soit une plus grande négativité pour les items monomorphémiques. Dans la fenêtre temporelle de 350 ms à 500 ms, conjointement à l'effet de lexicalité, il aurait été attendu que les pseudomots entraînent une plus grande négativité que les non-mots comme les résultats de Fiorentino et ses collaborateurs (2014) lors de l'identification de mots composés en anglais auprès de lecteurs adultes. Les résultats sur la latence de ces pics d'amplitude offrent une piste d'hypothèse pour expliquer ces résultats, mais qui demeurerait à être prouvée empiriquement. En effet, en région postérieure médiane, il est possible d'observer que le pic d'amplitude maximal généré par les items polymorphémiques apparaît significativement plus tôt que celui des items monomorphémiques dans la première fenêtre temporelle, de 200 ms à 350 ms. Puis, de 350 ms à 500 ms, c'est plutôt le pic d'amplitude maximal des items monomorphémiques qui apparaît plus tôt, suivi par celui des items polymorphémiques. En raison de la contiguïté des fenêtres temporelles,

il est possible que les résultats au sein de chacune des fenêtres temporelles témoignent de processus cognitifs distincts pour chacune des catégories d'items observées.

Pour les items monomorphémiques, les résultats ont déjà permis de constater un traitement significativement plus rapide que pour les items polymorphémiques. En s'appuyant sur les modèles hybrides (p. ex. Caramazza *et al.*, 1985), les mots monomorphémiques seraient traités par la voie globale et ne nécessiteraient aucune décomposition morphologique. De ce fait, on n'observerait qu'un seul pic d'amplitude vers 400 ms témoignant de l'activation sémantique du mot, mais aucun dans la fenêtre temporelle de 200 ms à 350 ms associée à la décomposition morpho-orthographique. Il est probable que le pic d'amplitude des mots monomorphémiques chevauche deux fenêtres temporelles. C'est ce qui expliquerait que la latence du pic d'amplitude maximal pour les items monomorphémiques est plus tardive dans la première fenêtre temporelle et arrive de façon précoce dans la deuxième. Pour les items monomorphémiques, le marqueur observé pourrait être un N400 précoce qui reflèterait l'activation du sens du mot (Beyersmann *et al.*, 2014), raison pour laquelle on obtiendrait un effet de structure morphologique aussi puissant et aussi tôt pour les items monomorphémiques.

Pour les items polymorphémiques, un autre cas de figure pourrait se dessiner. Il est possible que le traitement observé témoigne d'une décomposition morpho-orthographique rapide associée au marqueur N250 (Leminen *et al.*, 2018), suivi par un indice de composition sémantique associé au marqueur N400 (Coch *et al.*, 2013). Dans leur recherche, Coch et ses collègues (2013) ont attribué au marqueur N400 l'effet de composition sémantique plutôt que de décomposition morpho-orthographique comme le traitement des items anglais formés de racines libres (p. ex. *earring*) et liées (p. ex. *discern*) était distinct. En effet, dans leur étude, la négativité au N400 était significativement plus grande pour les items formés de racines libres que pour ceux formés de racines liées (Coch *et al.*, 2013). D'autres chercheurs ont aussi suggéré un traitement similaire (Brooks et Cid de Garcia, 2015; Fiorentino *et al.*, 2014), soit un traitement décompositionnel suivi d'une (re)composition sémantique, pour les items polymorphémiques. Ces interprétations sont toutes compatibles avec les modèles hybrides (p. ex. Caramazza *et al.*, 1985) qui prévoient une étape de décomposition pour les items polymorphémiques avant l'accès lexical. Une tâche d'amorçage masqué permettant de vérifier l'apport de la sémantique dans le traitement des mots écrits pourrait permettre de confirmer cette hypothèse.

En résumé, les mots polymorphémiques sont moins bien réussis à la tâche de décision lexicale que les items de contrôle (mots monomorphémiques et non-mots). Aussi, les mots polymorphémiques sont traités plus lentement que les items de contrôle, probablement parce qu'ils sont décomposés avant l'accès lexical. Malgré l'effet observé pour les données comportementales, les résultats aux mesures psychophysiques ne permettent pas d'arriver d'emblée au même constat. En effet, la division des fenêtres temporelles telle qu'elle a été réalisée dans le cadre de cette étude ne semble pas adaptée au traitement lexical des lecteurs intermédiaires, traitement plus lent que les lecteurs experts. De ce fait, les résultats observés sont décalés par rapport aux fenêtres temporelles habituelles. En s'appuyant sur la latence des pics d'amplitude maximale observée lors de la lecture des items de la tâche de décision lexicale, il est possible de penser que les mots monomorphémiques et les non-mots ne seraient pas décomposés avant l'accès lexical qui, lui, surviendrait plus rapidement pour les items de contrôle que pour les items polymorphémiques. Les items polymorphémiques, quant à eux, seraient d'abord décomposés en morphèmes puis recomposés pour accéder au sens du mot entier. Les modèles hybrides (p. ex. Caramazza *et al.*, 1985) appuieraient ces résultats.

5.1.3 L'effet du type de racine sur le traitement des mots écrits

Compte tenu du peu d'études portant sur les mots composés formés de racines liées, il a été difficile d'émettre des hypothèses robustes sur l'effet du type de racine sur le traitement des mots écrits. Toutefois, sur la base des résultats des études antérieures de Coch et ses collaborateurs (2013, 2020), il était envisagé que les participants répondent plus rapidement et plus précisément aux items formés de racines liées qu'à ceux formés de racines libres. Néanmoins, il convient de rappeler que, dans ces études, les racines liées forment des mots dérivés plutôt que composés. Ainsi, il était attendu que cet effet soit observé dans les résultats de la présente étude si l'effet était indépendant du type de racine sans influence du type de procédé de formation des mots. Concernant les mesures psychophysiques, l'étude de Coch et ses collègues (2013) a permis d'émettre l'hypothèse voulant que les mots composés formés de racines libres entraîneraient une plus grande négativité vers 400 ms poststimulus que les mots formés de racines liées si l'effet du type de racine était indépendant du type de procédé de formation des mots, ce qui aurait pu témoigner d'une étape de composition sémantique plus importante pour les mots composés de racines libres plutôt que liées. À la suite de l'analyse des résultats, aucune de ces hypothèses n'a pu être confirmée.

Un effet du type de racine a toutefois été constaté sur les mesures comportementales de précision et de temps de réaction. Pour les mots polymorphémiques, aucune différence dans le traitement des mots formés de racines libres ou liées n'est observée chez les élèves de 5^e et de 6^e année du primaire. Pour les pseudomots et les non-mots, les participants réussissent mieux les items composés formés de racines libres que de racines liées. Ainsi, lorsque les mots composés formés de racines liées latines ou grecques existent en français, la précision des élèves est similaire à celle des mots composés formés de racines libres. En contrepartie, quand la suite de lettres présentée formée de deux racines liées latines ou grecques n'est pas un mot qui existe en français, la précision est moins bonne que pour les suites de lettres formées de deux racines libres, probablement parce que le pseudomot est plausible (p. ex. le pseudomot **quintilingue* pourrait désigner une personne qui parle cinq langues). En ce qui concerne le temps de réaction des participants à fournir une réponse correcte, celle-ci est similaire pour tous les items polymorphémiques composés (mots et pseudomots), sans distinction quant au type de racine.

Les résultats des études antérieures sont différents. D'abord, l'étude de McKinnon et ses collègues (2003) a montré que leurs participants adultes anglophones étaient plus précis lors de la lecture de non-mots plutôt que lors de la lecture des autres catégories d'items : mots dérivés formés de racines liées, mots dérivés ou composés formés de racines libres, et pseudomots formés de racines liées. Pour Coch et ses collaborateurs (2013), une différence était observable entre le traitement des items de contrôle (mots monomorphémiques et non-mots), des items formés de racines liées et des items formés de racines libres sans distinction quant à la lexicalité. En effet, les participants fournissaient des réponses plus rapides et précises à la tâche de décision lexicale pour les items de contrôle que pour les items formés de racines liées. La précision de réponse des participants était meilleure pour les items formés de racines liées que pour ceux formés de racines libres. Cependant, aucune différence de temps de réaction n'était observable entre ces deux catégories d'items, ce qui corrobore les résultats de la présente étude. Lors de leur deuxième recherche, Coch et ses collaborateurs (2020) ont obtenu des différences à la fois selon le type de racine et selon la lexicalité. Lors de l'identification des mots, leurs participants anglophones réussissaient mieux et plus rapidement les items monomorphémiques que les items dérivés formés de racines liées. Les participants répondaient aussi plus rapidement et plus précisément aux items formés de racines liées qu'à ceux formés de racines libres. Quant à l'identification des pseudomots et des non-mots, les participants répondaient plus précisément et rapidement aux items de la catégorie de contrôle, les non-mots. Cependant, cette fois-ci, les participants performaient mieux et plus rapidement aux pseudomots formés de racines libres qu'aux pseudomots formés de racines liées (Coch *et al.*, 2020), ce

qui rejoint les résultats obtenus dans la présente étude en termes de précision, mais pas en termes de temps de réaction.

Concernant les mesures psychophysiques, aucun effet du type de racine n'a été trouvé sur l'amplitude et la latence aux deux fenêtres temporelles observées dans la présente étude. À notre connaissance, aucune étude n'a observé l'effet du type de racine dans une intervalle similaire à celle de la présente étude, de 200 ms à 350 ms poststimulus. Or, dans la fenêtre temporelle habituellement associée au N400, de 350 ms à 500 ms, les résultats de McKinnon et ses collègues (2003) rejoignent les résultats obtenus dans la présente étude. Les auteurs n'ont trouvé aucune différence d'amplitude associée au N400 entre les mots formés de racines libres, les mots formés de racines liées et les pseudomots formés de racines liées. Seuls les non-mots produisaient une plus grande négativité que les autres catégories d'items. À l'inverse, Coch et ses cochercheurs (2013) ont trouvé des différences significatives quant à l'amplitude, et ce, pour chacune des catégories d'items, sans distinction quant à la lexicalité. Leurs items formés de racines libres généraient de plus grandes négativités lors du traitement des mots écrits, suivis par les items de contrôle et, enfin, par les items formés de racines liées. Comme les auteurs ont aussi trouvé que la lecture de pseudomots et de non-mots entraînait une plus grande négativité que la lecture de mots bien qu'ils partagent les mêmes racines, ils associent le marqueur N400 à la composition sémantique plutôt qu'à la décomposition morphologique (Coch *et al.*, 2013). Néanmoins, ces résultats ne signifient pas nécessairement que les mots polymorphémiques ne sont pas décomposés avant l'accès lexical, mais simplement que ce ne serait pas le marqueur N400 qui serait associé à ce processus cognitif, ou encore qu'il n'y aurait pas de différence notable entre le traitement des mots formés de ces deux types de racine. Aussi, les résultats des études antérieures ayant observé l'effet du type de racine (Coch *et al.*, 2013; Coch *et al.*, 2020; McKinnon *et al.*, 2003) sont à interpréter avec prudence. Lors du contrôle des items, les racines liées ont été intégrées à des mots dérivés plutôt qu'à des mots composés. On ne sait donc pas si l'effet observé par les auteurs relève du procédé de formation des mots ou du type de racine. À notre connaissance, aucune recherche n'a comparé des items composés formés de racines libres et liées, ce qui empêche la comparaison efficace des résultats de la présente recherche avec ceux d'études antérieures.

Pour résumer, les résultats de la présente étude indiquent que le type de racine n'influence pas le traitement des mots polymorphémiques composés au-delà de la structure morphologique. Ainsi, les lecteurs identifieraient différemment les mots selon leur structure morphologique comme cela a été

discuté à la section 5.1.2, et ce, sans distinction selon le type de racine. En fait, seule la précision des pseudomots et des non-mots est affectée par le type de racine. Étant donné leur caractère plausible pour un lecteur intermédiaire conscient qu'il ne connaît pas tous les mots du français, mais qui reconnaît certaines racines, les pseudomots formés de racines liées portaient davantage à confusion que les pseudomots formés de racines libres ou les non-mots.

5.1.4 L'effet de l'origine des racines sur le traitement des mots écrits

Pour continuer, les hypothèses à visée exploratoire sur l'origine des racines concernaient exclusivement les mesures comportementales de précision et de temps de réaction. Étant donné la transparence sémantique plus importante des racines grecques par rapport aux racines latines (Templeton, 1983) et l'effet de cette transparence sur le traitement des mots composés (Davis *et al.*, 2019), il était attendu que les participants répondent plus rapidement et précisément aux items formés de racines grecques qu'à ceux formés de racines latines. Néanmoins, aucune donnée empirique ne permettait d'appuyer ces hypothèses ni de fournir des hypothèses concernant les résultats aux mesures psychophysiques. Au terme des analyses, il est possible de conclure que les hypothèses de départ ont été généralement réfutées et que des nuances sont à apporter.

Les résultats des ANOVAs à mesures répétées montrent un effet de l'origine des racines sur les mesures comportementales de précision et de temps de réaction lors de la tâche de décision lexicale. Aucun effet n'a été constaté sur les mesures psychophysiques d'amplitude et de latence lors du traitement de ces mots. Les résultats de l'ANOVA à mesures répétées pour la précision lors de la tâche de décision lexicale révèlent que les participants traitent différemment les mots et les pseudomots des deux origines. Les mots formés de racines grecques sont mieux réussis que les mots formés de racines latines. Il s'agit d'ailleurs du seul résultat quant à l'effet de l'origine des racines qui confirme l'hypothèse de départ. Pour les pseudomots, le profil inverse se dessine. Les participants identifient mieux les pseudomots formés de racines latines comme n'étant pas des mots qui existent en français que les pseudomots formés de racines grecques. C'est donc dire que les participants croient, à tort, que plusieurs pseudomots formés de racines grecques sont des mots qui existent en français. Les résultats montrent aussi que les items formés de racines latines sont traités plus rapidement que ceux formés de racines grecques, sans distinction entre les mots et les pseudomots.

Selon Templeton (1983), les racines latines sont moins transparentes que les racines grecques. Comme la transparence sémantique des morphèmes est une variable qui influence le traitement des mots polymorphémiques composés (Davis *et al.*, 2019), les résultats obtenus confirment l'hypothèse de départ à l'effet que les mots formés de racines latines sont moins bien réussis que ceux formés de racines grecques. Les résultats concordent aussi avec ceux d'une étude précédente ayant montré que les élèves québécois de 6^e année du primaire connaissent moins le sens des racines latines que celui des racines grecques (Whissell-Turner *et al.*, 2022). La transparence sémantique des racines grecques et latines pourrait aussi expliquer les résultats relatifs au traitement des pseudomots. Il est possible que les élèves du 3^e cycle du primaire connaissent mieux les racines grecques que latines, ce qui entraînerait de faux positifs de la part des élèves lorsqu'ils rencontrent des pseudomots formés de racines grecques. Les élèves de ce niveau scolaire, conscients que leur vocabulaire est moins développé que les adultes, pourraient penser que le pseudomot formé de racines grecques est plausible en français, surtout s'ils connaissent chacune des racines, mais que le mot entier leur est tout simplement inconnu (Henderson et Chard, 1980). Dans cette optique, on pourrait penser que les représentations lexicales des racines latines sont moins précises que celles des racines grecques pour les élèves de 3^e cycle du primaire.

Toujours sur la base de la transparence sémantique des racines (Templeton, 1983), l'hypothèse émise était que les participants répondraient plus rapidement aux items formés de racines grecques qu'à ceux formés de racines latines puisqu'ils seraient mieux connus des élèves. Or, les résultats de la présente étude infirment l'hypothèse de départ. Néanmoins, les modèles hybrides comme le modèle d'activation parallèle lexicomorphologique (Caramazza *et al.*, 1985) offre un cadre pour expliquer les résultats obtenus. Les racines latines étant plus opaques, les items qui en sont formés emprunteraient la voie directe durant le traitement morphographique. Au contraire, les items formés de racines grecques, plus transparentes, pourraient emprunter la voie de décomposition morphologique où ils seront décomposés en morphèmes avant l'accès lexical. La voie directe étant plus rapide puisqu'elle ne contient aucune étape transitoire, le traitement des items formés de racines latines serait donc plus rapide que celui des items formés de racines grecques (Caramazza *et al.*, 1985). Cependant, ces résultats ne sont pas observables par les mesures psychophysiologiques d'amplitude et de latence recueillies lors de la présente étude.

En somme, des différences entre les items formés de racines latines et ceux formés de racines grecques sont observables au niveau comportemental, soit par rapport à la précision et au temps de réaction des

participants lors de la tâche de décision lexicale. Les mots formés de racines grecques sont mieux reconnus des élèves que ceux formés de racines latines. Au contraire, les pseudomots formés de racines grecques entraînent des erreurs chez les élèves qui les considèrent davantage comme des mots qui existent en français comparativement aux pseudomots formés de racines latines. Aussi, les participants répondent généralement plus rapidement aux items formés de racines latines durant la tâche de décision lexicale qu'aux items formés de racines grecques qui, eux, entraînent un temps de traitement plus long. Cependant, au niveau psychophysique, aucune différence de traitement n'est significative. Les items des deux origines seraient traités de façon similaire. Les différences observées ne s'expliqueraient donc pas par l'activation cérébrale, mais pourraient s'expliquer par d'autres variables qui n'ont pas été manipulées directement dans la tâche de décision lexicale comme la fréquence, la productivité ou la transparence sémantique des racines latines et grecques. D'autres études devront être menées pour vérifier cette hypothèse, notamment en mesurant directement la transparence sémantique des racines plutôt qu'en se basant uniquement sur l'origine des racines dans le but de déterminer leur transparence.

5.1.5 Les effets de fréquence du mot entier et des racines sur le traitement des mots écrits

À la lumière des études antérieures (Duñabeitia *et al.*, 2007; Goodwin *et al.*, 2013; Kuperman *et al.*, 2009; Vergara-Martínez *et al.*, 2009), l'hypothèse formulée était que la fréquence des mots et celle des racines influenceraient positivement le temps de réaction et la précision des participants lors de la tâche de décision lexicale. Ainsi, il était attendu que plus les mots étaient fréquents ou formés de racines fréquentes, plus les participants allaient y répondre précisément et rapidement. Malgré certaines nuances à apporter, cette hypothèse a été confirmée. Quant aux mesures psychophysiques, peu d'hypothèses avaient été avancées. Sur la base des travaux de Vergara-Martínez et ses collègues (2009), différentes interprétations ont été envisagées en fonction des résultats de la présente étude. S'il advenait que la fréquence de chacune des racines influence différemment l'amplitude de l'activité électrique maximale générée aux deux fenêtres temporelles, il aurait été possible de conclure que le traitement de ces racines faisait appel à différents processus cognitifs. Dans le cas contraire, le traitement des racines ferait appel à des processus cognitifs similaires, qu'elles soient en première ou en deuxième position d'un mot composé. C'est cette dernière hypothèse qui a été retenue.

Les résultats de la régression logistique multiniveaux sur la précision des élèves lors de la tâche de décision lexicale ont révélé que, pour les mots, seule la fréquence du mot entier favorisait la précision.

Plus le mot est fréquent, plus les élèves de 5^e et de 6^e année du primaire sont susceptibles de fournir une réponse juste à la tâche de décision lexicale. Pour les pseudomots et les non-mots, parmi les facteurs linguistiques, c'est plutôt la fréquence de la deuxième racine qui influence ce résultat. Plus la deuxième racine du pseudomot est fréquente, meilleurs sont les élèves de 5^e et de 6^e année à identifier le pseudomot comme n'étant pas un mot existant en français durant la tâche de décision lexicale. Il est possible que ce soit la fréquence de la deuxième racine qui ait une plus grande influence que celle de la première racine puisque la lecture du pseudomot en entier est nécessaire pour porter un jugement lexical. Un élève ne pourrait pas porter son jugement lexical uniquement à la suite de la lecture de la première racine.

Pour le temps de réaction à fournir une réponse correcte à la tâche de décision lexicale, des variables linguistiques influencent la relation étudiée. Les résultats de la régression linéaire multiniveaux montrent que le temps de réaction aux mots est influencé par la fréquence du mot entier, alors que le temps de réaction aux pseudomots et aux non-mots est influencé par la fréquence de la première racine. Ainsi, les participants identifient plus rapidement les mots plus ils sont fréquents. De plus, durant la tâche de décision lexicale, les participants sont plus à même d'identifier que la suite de lettres présentée n'est pas un mot qui existe en français plus la fréquence de la première racine est élevée. Dans ce cas-ci, contrairement à la précision des pseudomots, c'est plutôt la fréquence de la première racine qui influence le temps de réaction à la tâche de décision lexicale. Il serait pertinent de présumer qu'un participant pourrait émettre son jugement lexical avant même d'avoir terminé la lecture de l'item : s'il ne reconnaît pas la première partie du mot, il ne connaît nécessairement pas le mot entier.

Les résultats dénotent aussi l'importance de la fréquence du mot entier dans l'identification des mots écrits. Plus le mot est fréquent, mieux il est reconnu et plus il est lu rapidement par les élèves du 3^e cycle du primaire. Ces résultats concordent avec les résultats des recherches antérieures qui identifient la fréquence du mot entier comme variable linguistique prédictive de la complexité du vocabulaire, autant en français (Gala *et al.*, 2014) qu'en anglais (Cervetti *et al.*, 2015; Hiebert *et al.*, 2019). Dans le même ordre d'idées, la recherche de Goodwin et ses collègues (2013) a aussi permis d'identifier la fréquence des mots et la fréquence des racines comme des facteurs linguistiques prédisant la précision en lecture des mots polymorphémiques formés de racines latines ou grecques auprès d'adolescents anglophones de 7^e et de 8^e année. Bien que le contexte linguistique différent de la présente étude, le groupe d'âge des participants et le choix des items est similaire. Pour les lecteurs experts adultes, les indices de

fréquence se sont aussi révélés des facteurs facilitateurs dans l'identification des mots écrits en néerlandais (Kuperman *et al.*, 2009) et en basque (Duñabeitia *et al.*, 2007; Vergara-Martínez *et al.*, 2009).

Cependant, les résultats de la présente étude contredisent légèrement une recherche antérieure. Effectivement, ni la fréquence du mot entier ni celle des racines n'influençait la performance des élèves de 6^e année du primaire lors d'un test de connaissance des racines latines et grecques (Whissell-Turner *et al.*, 2022). Cependant, des différences majeures sont notables entre les deux études. D'abord, dans la précédente étude, les chercheuses ne s'intéressaient pas à l'identification des mots écrits, mais à la connaissance des racines latines et grecques. De plus, la fréquence des mots entiers avait été mesurée à l'aide de la base de données Manulex (Lété *et al.*, 2004), une base de données qui répertorie les mots des manuels scolaires français de lecture du cours préparatoire au cours élémentaire 2. Dans la présente étude, une base de données mieux adaptée à la réalité des élèves québécois du primaire a été utilisée, soit le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018) qui répertorie les mots des manuels scolaires québécois dans différentes disciplines scolaires de la première année à la sixième année du primaire. Les mots formés de racines latines et grecques issus du vocabulaire disciplinaire sont plus susceptibles d'être répertoriés dans le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018) plutôt que dans Manulex (Lété *et al.*, 2004). Enfin, alors que le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018) a été utilisé à nouveau pour recueillir les fréquences des racines dans la présente étude, la base de données Polymots (LIF-TALEP, 2012) avait été utilisée dans la précédente étude. Déjà restreinte en termes de nombre de mots et du choix des mots (peu de mots du vocabulaire disciplinaire) (Gala et Rey, 2008), cette ressource fournit un indice de productivité de la racine parmi les mots disponibles dans la base de données. Cet indice s'apparente à la taille de la famille morphologique plutôt qu'à la fréquence de la racine.

Enfin, les résultats des régressions linéaires multiniveaux n'ont pas permis de détecter de facteurs linguistiques de fréquence (mots ou racines) exerçant une influence sur les données psychophysologiques recueillies, soit l'amplitude de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus ainsi que la latence à ces mêmes fenêtres temporelles. L'absence de l'influence de la fréquence des mots sur les résultats des mesures psychophysologiques peut être due au manque de variabilité de l'indice de fréquence. En effet, les mots sélectionnés dans chacune des catégories étaient tous relativement rares (voir Annexe D), particulièrement les mots formés de racines latines et grecques

comme c'est d'ailleurs une caractéristique des mots issus du vocabulaire disciplinaire. Pour mieux explorer cette relation, d'autres études devraient être menées en portant une attention particulière à la sélection des stimulus selon leur fréquence, et ce, en amont de la tâche de décision lexicale. Pour la fréquence des racines, les résultats de l'étude de Vergara-Martínez (2009) avaient montré que la fréquence des morphèmes avait un effet sur l'amplitude à différentes fenêtres temporelles de 100 ms à 650 ms poststimulus chez les lecteurs adultes basques, ce que les auteurs interprétaient comme un traitement différencié de chacune des racines faisant appel à des processus cognitifs différents. Dans la présente étude, la fréquence des racines n'était pas liée à l'amplitude générée lors de l'identification des mots écrits. Ainsi, nos résultats suggèrent que le traitement de la première racine et le traitement de la deuxième racine font appel à des processus cognitifs similaires. Cependant, cette variable n'a pas été manipulée lors de la conception de la tâche de décision lexicale. Si la fréquence des premières et des deuxièmes racines pour choisir les stimulus à la tâche de décision lexicale était contrôlée en amont, il serait possible que les résultats obtenus diffèrent.

En résumé, la fréquence du mot entier et la fréquence de chacune des racines d'un mot composé influencent l'identification des mots écrits. Les participants fournissent plus rapidement et plus précisément une réponse à la tâche de décision lexicale pour les mots fréquents que pour les mots rares. Bien qu'elle soit un facteur déterminant dans l'identification des mots écrits, la fréquence du mot entier à elle seule ne saurait expliquer le traitement des pseudomots qui représenteraient des mots inconnus pour des élèves du 3^e cycle du primaire lors de leurs lectures. Étant donné que les pseudomots ont une fréquence nulle, la fréquence des racines joue un rôle plus important dans le traitement de ces items. Comme l'indiquent Bowers et Kirby (2010), la connaissance de la racine est nécessaire pour accéder au sens des mots qui ont une faible fréquence. Ainsi, les participants rejettent plus rapidement un pseudomot qui commence par une racine fréquente que rare. De plus, ce rejet a plus de chance d'être juste plus la deuxième racine est fréquente. Ces résultats suggèrent la prise en compte des deux racines dans le traitement des mots polymorphémiques composés.

5.2 L'influence des facteurs individuels sur le traitement des mots écrits

À la suite de la recension des écrits antérieurs, il a été possible de formuler une hypothèse quant à l'influence des facteurs individuels sur le traitement des mots écrits polymorphémiques. À l'instar des résultats de la recherche de Goodwin et de ses collaborateurs (2013), il était attendu que les élèves ayant de meilleures habiletés morphologiques au test dynamique traitent les mots polymorphémiques

avec plus de précision et plus rapidement que leurs pairs. Aussi, l'hypothèse voulant que les élèves ayant de meilleures habiletés morphologiques décomposeraient davantage les mots polymorphémiques qu'ils lisent a été émise, ce qui se serait traduit par une plus grande négativité de 200 ms à 350 ms, fenêtre temporelle associée à la décomposition morpho-orthographique avant l'accès lexical (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008).

Avant de vérifier l'influence des habiletés morphologiques sur le traitement des mots écrits, les différences de traitement selon le niveau scolaire ont été vérifiées comme les participants étaient de deux niveaux scolaires différents : 5^e année et 6^e année du primaire. Bien que des tests *t* révélaient des différences entre les élèves de 5^e année et de 6^e année quant à la précision à la tâche de décision lexicale ainsi qu'à l'amplitude générée par l'activité cérébrale de 350 ms à 500 ms poststimulus, le niveau scolaire n'explique aucune de ces relations dans les régressions multiniveaux lorsqu'on insère les autres variables indépendantes. Cela pourrait s'expliquer par le fait que les tests *t* ont comparé les moyennes entre les niveaux scolaires en combinant les mots et les pseudomots et non-mots alors que des régressions multiniveaux distinctes ont été faites pour les mots ainsi que pour les pseudomots et les non-mots. Aussi, cette absence de significativité du niveau scolaire dans les analyses de régression multiniveaux indique que, statistiquement, le niveau scolaire (tout comme les indices de fréquence et les autres variables individuelles non significatives) ne contribue pas suffisamment à l'amplitude du potentiel évoqué de 350 ms à 500 ms au-delà des variables linguistiques manipulées (structure morphologique des items, type de racine, origine des racines) qui ont aussi été intégrées dans la régression. Il n'y aurait donc aucune différence significative entre les élèves de l'échantillon quant au traitement des mots écrits sur la base de leur niveau scolaire.

Ensuite, parmi les cinq facteurs individuels mesurés lors du test dynamique (connaissance du vocabulaire, reconnaissance des racines, décomposition morphologique, connaissance du sens des racines latines et grecques et analyse morphologique), aucun n'influence le traitement des mots polymorphémiques écrits. Ces résultats infirment l'hypothèse de départ. Néanmoins, lors de l'identification des pseudomots et des non-mots, la décomposition morphologique influence négativement la précision à la tâche de décision lexicale. Ainsi, plus les participants ont de bonnes habiletés à décomposer morphologiquement les mots et les pseudomots du test dynamique, plus ils commettent d'erreurs avec les pseudomots et les non-mots dans la tâche de décision lexicale. C'est donc dire que les participants qui décomposent efficacement les mots et les pseudomots selon leur structure morphologique ont plus tendance à

identifier les pseudomots (et peut-être les non-mots) comme étant des mots susceptibles d'exister en français. Les élèves du 3^e cycle du primaire savent qu'ils ne connaissent pas tous les mots du français, mais ils reconnaissent certaines racines. Ainsi, il est possible qu'un élève qui reconnaît une ou plusieurs unités de sens dans un pseudomot le considère comme un mot qui pourrait exister en français, ce qui influencerait sa précision à la tâche de décision lexicale. Aucun autre facteur individuel n'influence significativement le traitement des mots écrits sélectionnés pour la tâche de décision lexicale, que ce soit la précision, le temps de réaction ou encore l'amplitude ou la latence aux deux fenêtres temporelles ciblées.

Parmi les deux facteurs individuels mesurés, la compréhension en lecture et la conscience morphologique, Goodwin et ses collègues (2013) ont trouvé que la conscience morphologique prédisait la précision en lecture des mots formés de racines latines et grecques chez des adolescents anglophones de 7^e et de 8^e année. Aucune donnée n'est disponible pour la lecture de pseudomots. La présente étude n'a pas permis de trouver de relation entre la précision en lecture des mots et les habiletés morphologiques des participants. Néanmoins, plus les participants avaient de bonnes habiletés en décomposition morphologique, plus ils commettaient d'erreurs à la tâche de décision lexicale lors de la lecture de pseudomots et de non-mots. Ainsi, les élèves reconnaîtraient donc chacune des racines du mot, y attribueraient un sens et, en jumelant le sens des deux racines pour accéder au sens du pseudomot en entier, ils émettraient un jugement lexical positif. Probablement qu'ils arrivent à la conclusion que le concept analysé morphologiquement est tout à fait plausible en français, même s'ils ne connaissent pas le mot (p. ex. **quintilingue* qui signifierait « qui parle cinq langues »). Malgré les différences de résultats, les deux recherches mènent au même constat : les élèves de cet âge s'appuient sur leurs habiletés morphologiques pour identifier des mots formés de racines latines et grecques. Cependant, la nature du lien entre les habiletés morphologiques et le traitement morphographique est encore incertain. D'autres études sont nécessaires pour approfondir cette relation, notamment en mesurant différentes habiletés de la conscience morphologique chez les participants.

Aussi, contrairement à l'hypothèse de départ, il n'a pas été montré que les élèves ayant de meilleures habiletés morphologiques décomposaient les mots polymorphémiques avant l'accès lexical. L'absence de relation significative entre les mesures psychophysiques de la tâche de décision lexicale ainsi que les facteurs individuels recueillis lors du test dynamique laisse croire que le test dynamique, dans sa forme actuelle, ne mesure pas les connaissances impliquées dans la tâche de décision lexicale. On

pourrait aussi se demander si, comme le traitement morphographique est un processus cognitif inconscient, la connaissance explicite des racines par les élèves est nécessaire pour mener à la décomposition morpho-orthographique du mot ou si ce processus s'enclenche automatiquement par exposition à la langue. Alors que les données psychophysiologiques recueillies lors de la tâche de décision lexicale mesurent l'activation cérébrale durant la lecture de 60 mots polymorphémiques composés formés de racines latines et grecques, soit 60 racines latines et 60 racines grecques, le test dynamique, quant à lui, ne permet de mesurer la connaissance que du tiers de ces racines, soit 20 racines latines et 20 racines grecques. Cette différence peut être suffisamment importante pour ne pas être en mesure de détecter d'effets significatifs. Il est possible que les résultats aient été différents si la connaissance de toutes les racines latines et grecques avait pu être mesurée. Cette absence de résultats significatifs pourrait aussi être expliquée par la nature de la tâche, le test dynamique, ou encore la petite taille de l'échantillon. D'autres recherches sont nécessaires pour mieux comprendre la relation entre les données psychophysiologiques et les habiletés morphologiques.

Enfin, les résultats du test dynamique permettent de constater que les élèves du 3^e cycle du primaire connaissent peu le sens des racines latines et grecques présentées. Au total, le test comptait 40 racines latines et grecques à raison de deux par item. Un score maximal de trois points par racine était attribué lors d'une connaissance précise du sens de la racine. Les résultats indiquent que les élèves de cet âge connaissent partiellement les racines latines et grecques présentes dans leurs manuels scolaires (45,83 %). Les résultats d'une étude antérieure confirment ce pronostic : les élèves québécois de 6^e année du primaire connaissent 56,58 % des racines latines et grecques (Whissell-Turner et Fejzo, 2021). Il est à noter que les résultats de la présente étude sont inférieurs à ceux de l'étude antérieure, notamment parce que les participants n'avaient pas accès à des choix de réponse pour déterminer le sens des racines latines et grecques. Ainsi, dans l'étude de Whissell-Turner et Fejzo (2021), une partie des réponses correctes données par les participants peut être attribuable au hasard.

5.3 Les résultats à la lumière des modèles de l'identification des mots écrits polymorphémiques

Les résultats de la présente étude permettent de compléter les modèles actuels de l'identification des mots écrits polymorphémiques qui ne considèrent pas explicitement la composition morphologique et le développement des connaissances chez des lecteurs toujours en apprentissage. D'un point de vue développemental, les résultats indiquent que les élèves du 3^e cycle du primaire traitent les morphèmes lors de l'identification des mots écrits polymorphémiques. Ainsi, selon le modèle développemental à

double fondation de Seymour (1999, 2007), les élèves de cet âge établissent déjà leur système morphographique. Ainsi, cela confirme que l'identification des mots écrits chez des élèves de cet âge pourrait être expliquée par un ou des modèles d'identification des mots écrits qui considèrent la morphologie dans le traitement des mots.

D'abord, les résultats ne semblent pas appuyer les modèles supralexicaux (Giraudo, 2005; Giraudo et Grainger, 2000, 2001, 2003; Grainger *et al.*, 1991) qui soutiennent que l'activation des représentations morphologiques des mots, notamment les familles morphologiques, serait une étape intermédiaire entre le traitement de la forme orthographique entière du mot et l'accès lexical. Les résultats de la présente étude indiquent qu'en région postérieure médiane de 200 ms à 350 ms, l'activation cérébrale associée au traitement des mots monomorphémiques et des non-mots est détectée plus tard que l'activation cérébrale associée au traitement des mots et des pseudomots polymorphémiques. Cependant, comme cette fenêtre temporelle est associée au marqueur N250 témoignant du traitement prélexical des mots polymorphémiques basé sur la forme morpho-orthographique du stimulus (Leminen *et al.*, 2018; Morris *et al.*, 2007; Morris *et al.*, 2008), aucune différence n'aurait dû être constatée quant à la latence de ce marqueur entre les items polymorphémiques et les items de contrôle pour permettre d'appuyer les modèles supralexicaux. Cependant, cela reste à confirmer empiriquement à l'aide du paradigme d'amorçage.

Pour continuer, il est difficile d'associer les résultats obtenus aux modèles connexionnistes distribués (Gonnerman et Plaut, 2000; Gonnerman *et al.*, 2007; Plaut et Gonnerman, 2000; Seidenberg et Gonnerman, 2000) qui prévoient une séquence de traitement identique pour tous les mots écrits. En effet, ces modèles prévoient que tous les mots sont activés dans le lexique mental sur la base des représentations orthographiques, phonologiques et sémantiques et non sur la base des représentations morphologiques. Ainsi, il serait difficile d'expliquer la raison pour laquelle, dans la tâche de décision lexicale, des différences de traitement sont observables entre les mots monomorphémiques et les mots polymorphémiques quant à la précision, au temps de réaction, à l'amplitude et à la latence de 200 ms à 350 ms et de 350 ms à 500 ms. Il est possible que les mots monomorphémiques soient activés plus rapidement dans le lexique mental puisqu'ils convergent avec moins de mots que les mots polymorphémiques qui, eux, convergent nécessairement avec tous les mots de leurs familles morphologiques et étymologiques (p. ex. *hydrographie* : *hydrographique, hydrater, hydratation, réhydrater, graphique, orthographe*, etc.). Il en va de même pour les non-mots qui devraient être traités

rapidement selon les modèles connexionnistes puisque la convergence entre les représentations orthographiques, phonologiques et sémantiques échoue. Cependant, qu'en est-il des pseudomots? Dans cette optique, le traitement des pseudomots devrait être beaucoup plus lent que celui des non-mots puisque les pseudomots convergeraient, au même titre que les mots composés, avec de multiples représentations orthographiques et phonologiques. Certaines représentations sémantiques pourraient même être récupérées sans qu'il y ait convergence totale. Les résultats de la présente étude ont d'ailleurs montré que les pseudomots formés de racines libres et liées entraînaient des temps de réponse plus lents que les non-mots. Or, comme aucun effet d'amplitude ou de latence aux deux fenêtres temporelles n'a été observé selon le type de racine, il semble que les mots et les pseudomots polymorphémiques composés formés de racines libres et liées soient traités de façon similaire au niveau cognitif. D'autres recherches devraient être menées, notamment en amorçage masqué, pour vérifier si l'accès au sens des mots composés est influencé par l'accès au sens de ses racines, c'est-à-dire si l'amorce *six* permet d'accéder plus facilement au mot *hexagone* ainsi qu'au pseudomot **hexachrome*. Une telle recherche serait particulièrement intéressante dans la mesure où malgré le chevauchement sémantique, le chevauchement phonologique ou orthographique entre l'amorce et la cible est réduit.

Cependant, des questionnements peuvent être soulevés quant aux résultats obtenus par rapport aux effets de fréquence. Selon les modèles connexionnistes distribués (Gonnerman et Plaut, 2000; Gonnerman *et al.*, 2007; Plaut et Gonnerman, 2000; Seidenberg et Gonnerman, 2000), comment la fréquence des mots et celle des racines interviendraient dans l'identification des mots écrits? Il est possible que plus un mot soit fréquent, plus rapidement seront activées les différentes dimensions du mot afin d'accéder à son sens, ce qui concorderait avec les résultats de la présente étude quant à la fréquence des mots. Pour les pseudomots, les résultats indiquent que ce serait plutôt la fréquence des racines, notamment la première racine qui favoriserait le temps de réaction lors de la tâche de décision lexicale. Pour fournir un exemple, le pseudomot **univole* est formé d'une première racine très fréquente : la racine *uni* a une fréquence de 6 618 occurrences dans le Lexique Scientifique Transdisciplinaire Junior (Tremblay *et al.*, 2023; Tremblay *et al.*, 2018). C'est aussi une racine très productive : elle forme une vingtaine de mots différents présents dans les manuels scolaires des élèves dont *unité, unique, réunir, union, uniforme, unisson*, etc. Ainsi, malgré sa fréquence qui pourrait faciliter le temps de réaction, la productivité de la racine selon les modèles connexionnistes distribués serait plutôt une entrave à ce temps de réaction puisqu'un plus grand nombre de représentations orthographiques, phonologiques et sémantiques seraient activées avant de pouvoir traiter ce

pseudomot pour l'exclure lors d'une tâche de décision lexicale, ce qui irait à l'encontre des résultats obtenus. Toutefois, d'autres recherches qui permettraient d'évaluer les effets de la fréquence et de la productivité des racines devraient être menées pour éclaircir cette situation.

Enfin, les modèles hybrides (p. ex. Caramazza *et al.*, 1985; Chialant et Caramazza, 1995; Frauenfelder et Schreuder, 1992; Schreuder et Baayen, 1995) qui prévoient deux voies de traitement des mots écrits, une voie de traitement du mot entier et une voie de décomposition morphologique, permettent d'expliquer les résultats obtenus dans la présente étude. En effet, les élèves traitent plus rapidement les items de contrôle (mots monomorphémiques et non-mots) que les items polymorphémiques. Les items de contrôle emprunteraient donc la voie globale alors que les items polymorphémiques emprunteraient la voie de décomposition morphologique. Ce résultat permet de constater que les modèles hybrides (p. ex. Caramazza *et al.*, 1985) expliquant le traitement des mots fléchis et dérivés permettent aussi d'expliquer le traitement des mots composés, sans distinction quant au type de racine (libre ou liée) ou à l'origine des racines (latines ou grecques). Comme prévu par les modèles hybrides, les facteurs de fréquence des mots et des racines dans l'étude se sont aussi révélés significatifs lors de la précision et du temps de réaction des participants à la tâche de décision lexicale. S'appuyer sur le traitement des pseudomots est un autre indice de la présence de la voie de décomposition morphologique dans le traitement des mots écrits. En effet, les élèves qui arrivent adéquatement à décomposer un mot en morphèmes réalisent la même procédure lors de la lecture de pseudomots. Cela les mène à croire, à tort, que les pseudomots transparents sémantiquement qui leur sont présentés sont des mots en français. Cela porte à croire que les élèves toujours en apprentissage du vocabulaire, donc avec un vocabulaire possiblement moins étendu que les lecteurs experts, s'appuient sur leurs habiletés morphologiques pour accéder au sens des mots inconnus. Aussi, comme le lexique mental des élèves de cet âge est probablement moins vaste que celui des lecteurs experts, on pourrait penser que les élèves surutilisent la voie de décomposition morphologique pour tenter d'accéder au sens des mots. Cela pourrait expliquer les temps de réponse plus longs ($M = 1422,70$ ms) que des lecteurs experts pour qui les temps de réponse sont plutôt autour de 1000 ms (Luck, 2014), puisque l'emprunt de la voie globale serait plus rapide que la voie de décomposition morphologique. Une autre explication possible aux temps de réponse plus longs de la part des élèves du 3^e cycle du primaire par rapport aux lecteurs experts pourrait concerner leur mémoire de travail qui est encore développement et le manque d'automatisation des processus cognitifs d'identification des mots écrits.

En conclusion, les résultats obtenus dans la présente étude pourraient être compatibles avec les modèles connexionnistes distribués ainsi que les modèles hybrides. Cependant, des questionnements subsistent toujours quant à la convergence des résultats avec les modèles connexionnistes distribués, notamment quant à la fréquence des racines ainsi qu'à l'accès au sens des racines, qui devraient être abordés par d'autres recherches. Pour le moment, les modèles hybrides sont ceux qui expliquent le mieux les résultats obtenus en attribuant un traitement différent pour les items monomorphémiques et pour les items polymorphémiques. De plus, ces modèles permettent un traitement des mots écrits favorisé par la fréquence des mots et accordent une place importante à la fréquence des racines dans le traitement des pseudomots. Enfin, les modèles hybrides permettent aussi d'expliquer les résultats en lien avec les habiletés de décomposition morphologique des élèves du 3^e cycle du primaire qui mène à de faux positifs lors de la tâche de décision lexicale. Les élèves utiliseraient la voie de décomposition morphologique en présence d'un pseudomot inconnu formé de deux racines et transparent sémantiquement afin de tenter de déduire le sens de ce pseudomot.

Étant la première étude à porter sur le traitement morphographique des mots composés formés de racines latines et grecques en contexte francophone, les résultats ont permis de vérifier les variables qui influencent les processus cognitifs observés. Effectivement, il a été possible de constater une différence dans le traitement des mots selon leur structure morphologique : les items polymorphémiques sont moins bien réussis et sont traités plus lentement que les items monomorphémiques qui, eux, sont accompagnés d'une activation cérébrale plus grande dès 200 ms. Les résultats ont aussi permis de renseigner sur le fait que le type de racine ainsi que l'origine des racines n'avaient un effet que sur les mesures comportementales de précision et de temps de réaction à la tâche de décision lexicale. Aucune différence n'est observable dans l'activité cérébrale générée lors de l'identification des mots écrits. Les élèves du 3^e cycle du primaire identifient plus facilement les items formés de racines grecques que latines, ce qui les mène à produire des erreurs avec les pseudomots formés de racines grecques. D'ailleurs, alors que le traitement lexical est similaire pour les mots formés de racines libres et ceux formés de racines liées, les élèves produisent plus d'erreurs avec les pseudomots formés de racines liées puisqu'ils les identifient à tort comme des mots qui existent en français. Ces résultats permettent d'émettre l'hypothèse que les élèves de 3^e cycle du primaire sont sensibles aux composantes morphologiques des mots et qu'ils tentent de déduire le sens des mots inconnus lors de l'identification des mots écrits lorsque la structure morphologique et la transparence sémantique le permettent. C'est d'ailleurs pour cette raison que les élèves qui présentent de meilleures habiletés de décomposition

morphologique produisent plus d'erreurs lors de l'identification des pseudomots : ils décomposent ces pseudomots et leur attribuent un sens tout à fait probable (p. ex. **quintilingue* = « qui parle cinq langues »). De plus, les résultats de la présente étude permettent aussi de mettre l'accent sur la fréquence du mot entier, mais aussi sur la fréquence des racines lors de l'identification des mots écrits, notamment pour les lecteurs toujours en apprentissage qui rencontreraient une grande quantité de mots inconnus lors de leurs lectures. Les recherches futures devraient considérer ces variables dans le choix des items pour leurs tâches expérimentales.

La combinaison des deux types de mesures, comportementales et psychophysiologiques, éclaire la compréhension du traitement morphographique en permettant d'observer des processus cognitifs inconscients et qui ne sont pas observables par les mesures comportementales, mais l'inverse est aussi vrai. D'un côté, les processus de décomposition morpho-orthographique et d'accès lexical se mesurent par la prise de mesures psychophysiologiques, ici l'activation cérébrale aux fenêtres temporelles correspondantes. Cependant, comme les recherches en neurosciences éducationnelles dans le domaine des connaissances morphologiques sont encore récentes et que l'objectif de l'apprentissage est de favoriser la performance et la réussite des élèves, la collecte de données comportementales est tout aussi importante. Ces données comportementales permettent de mieux interpréter et nuancer les données psychophysiologiques recueillies. Ainsi, d'un autre côté, la combinaison des mesures permet de déterminer les résultats qui relèvent directement du traitement des mots écrits et ceux qui seraient déterminés par des facteurs externes. Par exemple, des différences comportementales ont été observées quant à l'origine des racines, mais aucune différence n'a été trouvée au niveau cérébral. Cette absence de significativité pourrait indiquer que les mots formés de racines latines et grecques sont traités de façon similaire au niveau cognitif bien que les réponses fournies par les élèves diffèrent, ce qui peut être influencé par d'autres facteurs comme la fréquence ou la transparence sémantique. L'utilisation unique des mesures comportementales de précision et de temps de réaction aurait pu mener vers de fausses interprétations avançant un traitement différencié des items formés de racines latines et grecques. Cependant, ces résultats doivent être interprétés avec prudence puisque l'absence de différences entre les potentiels évoqués des catégories d'items peut aussi être expliquée par plusieurs autres facteurs (p. ex. taille de l'échantillon, nombre d'items retenus, électrodes moyennées par régions d'intérêts, etc.).

CONCLUSION

Les textes présentés aux élèves du 3^e cycle du primaire pour l'enseignement des disciplines scolaires sont remplis de mots formés de racines latines et grecques. Des modèles théoriques ainsi que des études empiriques tentent d'expliquer le traitement morphographique de ces mots chez les lecteurs experts. Cependant, aucun modèle ni aucune étude ne permet de renseigner spécifiquement sur le traitement des mots composés formés de racines latines et grecques chez des élèves du 3^e cycle du primaire qui développent toujours leurs connaissances et habiletés d'identification des mots écrits. De plus, le statut particulier des mots formés de racines latines et grecques nécessite une attention particulière puisqu'ils sont majoritairement des mots composés, donc formés de deux racines porteuses du sens principal du mot. Qui plus est, elles sont liées, donc non autonomes lexicalement. Dans ce contexte, les effets de facteurs linguistiques (la lexicalité, la structure morphologique, le type de racine, l'origine des racines, la fréquence des mots et la fréquence des racines) sur le traitement morphographique (la précision, le temps de réaction ainsi que l'amplitude et la latence de 200 ms à 350 ms et de 350 ms à 500 ms poststimulus) chez des élèves du 3^e cycle du primaire durant une tâche de décision lexicale ont été examinés. De plus, nous avons vérifié si des facteurs individuels liés à la connaissance du vocabulaire et à différentes habiletés morphologiques influençaient le traitement de ces mots composés formés de racines latines et grecques. Pour ce faire, différents cadres théoriques issus de domaines de recherche connexes comme la linguistique, la psycholinguistique, la didactique des langues ainsi que les neurosciences éducationnelles ont servi d'appui. Le caractère innovant de la recherche réside dans l'utilisation combinée de données issues des mesures comportementales et psychophysiologiques. L'apport des données psychophysiologiques est inestimable dans la mesure où elle permet de renseigner sur des processus cognitifs, comme le traitement morphographique, qui ne peuvent pas être mesurés uniquement à l'aide de mesures comportementales puisque l'élève en est inconscient.

Du point de vue scientifique, les résultats ont permis d'alimenter les modèles de traitement morphographique disponibles, notamment le modèle d'activation parallèle lexicomorphologique (Caramazza *et al.*, 1985), afin de permettre de mieux comprendre le traitement morphographique des mots composés formés de racines latines et grecques, mais aussi des pseudomots. Par exemple, l'effet de la fréquence du mot entier sur la précision et le temps de réaction des participants à la tâche de décision lexicale a corroboré les résultats d'études antérieures. Les résultats de la présente étude permettent aussi d'inférer le traitement de mots inconnus par les élèves du primaire, des mots

représentés dans la tâche expérimentale par les pseudomots. En effet, une contribution de la présente étude concerne les effets de fréquence des racines lors du traitement des pseudomots : la fréquence de la 1^{re} racine permet un temps de réaction plus rapide alors que la fréquence de la 2^e racine entraîne une meilleure précision à la tâche de décision lexicale. Dans une visée didactique, les résultats invitent à considérer l'enseignement des composantes morphologiques comme une avenue prometteuse pour l'enseignement du vocabulaire disciplinaire des élèves du primaire, mais aussi pour soutenir les élèves dans leur compréhension de textes dans différentes disciplines scolaires et dans l'accès au sens des mots inconnus. En s'appuyant sur les résultats obtenus, des didacticiens pourraient concevoir des programmes d'intervention sur le vocabulaire disciplinaire qui respectent le développement des élèves du 3^e cycle du primaire notamment en lien avec le développement des connaissances morphologiques. Une vérification des effets de la mise en œuvre d'un tel programme dans les salles de classe sur les compétences en littératie pourrait ensuite être menée.

De plus, la présente étude a permis de clarifier le flou terminologique qui entourait le concept de racines latines et grecques. Il s'agit d'une contribution importante dans le domaine de la morphologie. De plus, le travail d'analyse des propriétés psycholinguistiques des items de la tâche de décision lexicale est une autre contribution de la présente étude. En effet, peu de ressources permettent à ce jour de recenser les racines latines et grecques du vocabulaire français dans le respect des définitions présentées. Dans le cadre de cette thèse, il a fallu calculer manuellement les indices de fréquence des racines, les indices de la taille des familles morphologiques (mot entier, 1^{re} racine, 2^e racine), ainsi que le cycle de la première occurrence du mot et des racines dans les manuels scolaires des élèves du primaire. Cela témoigne d'un besoin important de recension pour le milieu scientifique, mais aussi pour le milieu scolaire. À ce jour, aucune liste ne permet de répertorier les racines latines et grecques les plus fréquentes et productives dans les manuels scolaires des élèves ni leurs propriétés psycholinguistiques (p. ex. nombre de lettres, nombre de graphèmes, nombre de syllabes, fréquence, taille de la famille morphologique, fréquence de la famille morphologique). Ce projet devrait faire l'objet d'une future recherche afin de guider les chercheurs dans la conception de listes de mots expérimentales ou de dispositifs didactiques ainsi que les personnes enseignantes dans l'enseignement de ces racines latines et grecques.

D'un point de vue méthodologique et didactique, la présente étude a permis de concevoir un test dynamique pour mesurer la connaissance du vocabulaire disciplinaire, le sens des racines latines et grecques ainsi que les habiletés morphologiques d'élèves du 3^e cycle du primaire. Comme il n'était pas

attendu que les participants aient une connaissance approfondie du vocabulaire disciplinaire, du sens de chacune des racines et qu'ils aient maîtrisés toutes les habiletés morphologiques, le test dynamique était un instrument de mesure à privilégier pour renseigner sur les connaissances des participants qui sont toujours en développement (Burton et Watkins, 2007; Poehner et Lantolf, 2010). En contrepartie, les tests comportementaux généralement utilisés en éducation, où une question est posée aux élèves qui fournissent une réponse orale ou écrite, utilisent la réponse de l'élève pour renseigner sur des connaissances ou habiletés statiques. L'utilisation de tests dynamiques permet de visualiser l'apprentissage sur un continuum et comme un processus itératif plutôt que séquentiel : la maîtrise totale d'une habileté n'est pas nécessaire pour entamer des apprentissages essentiels à une autre habileté. La conception de ce test dynamique est un apport méthodologique important puisque ce type de test est encore peu utilisé dans le domaine de l'éducation malgré son apport didactique indéniable. De futures recherches gagneraient à concevoir et à valider des tests dynamiques pour mesurer diverses habiletés cognitives auprès des élèves du primaire et ainsi renseigner les milieux scientifiques et scolaires sur le développement des connaissances et habiletés des élèves. D'ailleurs, le test dynamique de la présente étude pourrait être amélioré de différentes façons. D'abord, pour les besoins de l'étude, l'accent a été mis davantage sur le sens du vocabulaire disciplinaire et des racines. Une version améliorée du test dynamique pourrait considérer tous les aspects de la connaissance du vocabulaire : la forme, le sens et l'usage (Nation, 2001). De plus, afin d'éviter la surgénéralisation de la stratégie d'analyse morphologique par les élèves, il serait préférable d'ajouter des mots de contrôle aux items du test dynamique. Ces mots présenteraient une suite orthographique rappelant les racines latines et grecques, et ce, même si les racines latines et grecques ont peu d'homographes (p. ex. : *agripper* comme homographe à la racine *agri*).

Du point de vue éducationnel, les résultats de la recherche permettent de constater que les élèves du 3^e cycle du primaire sont sensibles aux informations morphologiques des mots et des pseudomots pour accéder à leur sens, y compris les racines latines et grecques. Dans cette optique, si les élèves traitent des morphèmes relativement opaques et rares telles que les racines latines et grecques, on peut penser que les morphèmes plus fréquents, comme les préfixes (p. ex. *re-*, *in-*) et les suffixes dérivationnels (p. ex. *-age*, *-ion*), seraient aussi traités par des élèves de cet âge. De futures recherches devraient comparer le traitement de ces morphèmes liés. En se basant sur les résultats de la présente étude, les enseignantes et enseignants pourraient consacrer du temps à l'enseignement de la morphologie, plus précisément des racines latines et grecques les plus fréquentes et productives pour chacune des matières scolaires. En

mathématiques, une activité rapide pourrait être mise en place par les personnes enseignantes en associant les nombres en grec aux nombres arabes connus des élèves. Ainsi, il est possible que les élèves qui connaissent les racines numériques grecques se rappellent plus aisément le nom des différents polygones (p. ex. *hexagone*, *heptagone*, etc.) et, par le fait même, des caractéristiques essentielles associées, soit leur nombre d'angles et de côtés. L'objectif ne serait pas d'enseigner les langues latines et grecques, mais bien d'attribuer un sens à ces morphèmes d'origine latine et grecque fréquents et productifs dans les manuels scolaires des élèves au même titre que les préfixes et suffixes fréquents, comme il est déjà recommandé dans la progression des apprentissages du français langue d'enseignement au primaire (Ministère de l'Éducation, du Loisir et du Sport, 2009). L'enseignement des racines latines et grecques pourrait être envisagé de façon interdisciplinaire, notamment au primaire où la personne titulaire est responsable de l'enseignement de la majorité des matières scolaires.

Des limites sont à considérer dans l'interprétation des résultats de la présente étude. D'abord, la petite taille de l'échantillon, les critères de sélection des participants ainsi que le choix des items empêchent la généralisation des résultats obtenus. En effet, les critères de sélection des participants favorisent une représentation plus homogène de la population. Les participants issus de milieux défavorisés ou présentant des difficultés d'apprentissage, pour ne nommer que ces exemples, ne sont pas représentés. De plus, l'échantillon est très petit ($n = 29$). Cela a d'ailleurs un effet considérable sur la puissance statistique des ANOVAs à mesures répétées effectuées compte tenu du nombre de variables manipulées. À l'exception de la précision à la tâche de décision lexicale, les autres données (temps de réaction, amplitude et latence des potentiels évoqués) étaient analysées en fonction des bonnes réponses fournies par les participants. Étant donné que les participants sont en situation d'apprentissage, leur vocabulaire est moins développé que celui des lecteurs experts. Ils produisent notamment un grand nombre d'erreurs. Pour certaines conditions, seuls 23 % des items étaient répondus correctement (voir Tableau 3.7). À cela, des artefacts contaminant les segments corrects (p. ex. mouvements variés, clignement des yeux) étaient inévitables, surtout auprès d'enfants. Ces segments ont été retirés, entraînant une perte supplémentaire moyenne de 20,16 %. Dans des cas extrêmes, ce pourcentage est monté jusqu'à 94,12 % pour certaines conditions chez certains participants, nous obligeant à ne conserver qu'un seul item par condition. La quantité d'items pouvant être conservée pour les analyses était donc considérablement réduite, ce qui a fortement réduit la puissance statistique. De ce fait, il n'a été possible de détecter de façon significative que des effets moyens et grands. Les petits effets n'étaient pas détectés significativement. Lors des analyses sur les potentiels évoqués, plusieurs variables

peuvent influencer la puissance statistique comme la taille de l'échantillon, le nombre d'items retenus ou encore la taille de l'effet. Cependant, le nombre d'items retenus serait une variable qui aurait le plus d'effet sur la puissance statistique des résultats, même lorsque le nombre de participants est réduit (Boudewyn *et al.*, 2018). Des auteurs ont proposé un seuil correspondant au nombre d'items idéal pour détecter des résultats significatifs. Par exemple, en contexte de traitement de phrases, le N400 ne nécessiterait qu'une quinzaine d'items par condition alors que le P600 en nécessiterait jusqu'à 40 en fonction de la taille de l'échantillon (Kim *et al.*, 2024). Pour des potentiels évoqués qui produisent typiquement de petits effets (p. ex. le potentiel de préparation latéralisé, *lateralized readiness potential* [LRP]), 45 items retenus par participant seraient nécessaires pour assurer une forte fiabilité interne et détecter des effets petits et moyens (Boudewyn *et al.*, 2018). Ainsi, le nombre d'items retenus par condition suffisant pour mener à des effets significatifs varie selon les potentiels évoqués mesurés. Dans le cadre de cette étude, il a été impossible de trouver plus de 30 items par condition qui respectaient les critères établis. Ce choix a définitivement eu un impact sur les résultats de l'étude puisque le nombre d'items retenus était insuffisant pour détecter des petits effets. De plus, pour les analyses de régression multiniveaux (logistique et linéaire), malgré une puissance statistique suffisante, le nombre d'observations ($n = 29$) était en deçà des recommandations optimales d'avoir au moins 50 observations au deuxième niveau de la régression multiniveaux (Sommet et Morselli, 2017). Cependant, plusieurs obstacles ont été rencontrés lors du recrutement des participants. D'abord, le nombre élevé de critères de sélection a réduit considérablement l'échantillon. Aussi, il a été difficile de rejoindre les participants en ne recrutant pas directement dans les écoles. De plus, la collecte de données a eu lieu à la fin de l'année scolaire et durant l'été. La fin de l'année scolaire est marquée par une surcharge d'évaluations pour les élèves, particulièrement pour les élèves de 6^e année du primaire qui doivent réaliser les examens ministériels. L'été, c'est la période des vacances pour plusieurs familles qui sont indisponibles. Il ne faut pas non plus négliger le fait que la collecte de données a eu lieu en période post-COVID au laboratoire de recherche NeuroLab à l'UQAM, au centre-ville de Montréal. Plusieurs familles à l'extérieur de Montréal étaient réticentes à prévoir un voyage dans la métropole où on dénombrait toujours plusieurs cas de la COVID-19, surtout lorsque la proximité avec les assistantes de recherche était inévitable (malgré le port du masque). Enfin, l'électroencéphalographie est un instrument de mesure encore peu connu et peu utilisé en éducation, ce qui peut effrayer certains jeunes participants, et même certains parents. Pour de futures recherches avec des participants de cet âge, il serait favorable de prévoir un plus grand échantillon, mais surtout de déterminer des critères de sélection des items qui

permettent d'augmenter le nombre d'items par condition afin d'accroître la puissance statistique des analyses.

De plus, les résultats obtenus doivent être interprétés au regard des mots utilisés dans les mesures expérimentales, sans généraliser à tous les mots formés de racines latines et grecques dans les manuels scolaires des élèves québécois du 3^e cycle du primaire. Effectivement, de nombreux critères ont orienté le choix des stimulus pour la tâche de décision lexicale afin d'isoler chacune des variables indépendantes (lexicalité, structure morphologique, type de racine, origine des racines). De ce fait, certains items qui auraient pu s'avérer pertinents pour observer le traitement morphographique des mots composés formés de racines latines et grecques chez des élèves du 3^e cycle du primaire n'ont pas pu être considérés. Par exemple, le mot *décennie* n'a pas pu être utilisé bien qu'il soit formé de deux racines latines puisqu'il contient aussi le suffixe *-ie*. Aussi, les mots qui sont formés d'une racine liée d'origine latine ou grecque et d'une racine libre étaient retirés de la sélection, comme le mot *décimètre* pourtant bien fréquent dans les manuels scolaires de mathématiques au primaire. Dans le même ordre d'idées, les mots formés d'une racine liée de chaque origine (p. ex. *patriarche*) devaient aussi être exclus. Néanmoins, comme aucune différence significative n'a été détectée au niveau des mesures psychophysiques dans le traitement des mots formés de racines libres ou liées ainsi que ceux d'origine latine ou grecque, de futures recherches pourraient utiliser une sélection plus étendue de mots du vocabulaire disciplinaire, en intégrant plutôt la transparence sémantique dans les modèles d'analyses statistiques.

Aussi, dans le choix des stimulus pour la tâche de décision lexicale, il n'a pas non plus été possible de contrôler pour la fréquence des mots ni pour la fréquence des racines. La quantité élevée de critères réduisait déjà la sélection des stimulus possibles. Comme la fréquence des mots et celle des racines influencent l'identification des mots écrits, particulièrement la précision et le temps de réaction à la tâche de décision lexicale, il serait intéressant de contrôler ces variables en amont dans une future tâche de décision lexicale, d'autant plus que les résultats de l'étude permettent de réduire les variables indépendantes à manipuler (type de racine et origine des racines). À priori, un meilleur contrôle de la fréquence des mots et de celle des racines pourrait permettre d'émettre des hypothèses plus robustes sur le traitement des mots formés de racines latines et grecques.

En outre, un élément indispensable à discuter dans le cadre de cette recherche est la validité écologique qui est une forme de validité externe témoignant de la représentativité des résultats obtenus en laboratoire au regard de ceux pouvant être observés au quotidien (Brewer et Crano, 2014). Étant donné les contraintes matérielles liées à l'utilisation de l'électroencéphalographie, la tâche de décision lexicale a été réalisée en laboratoire et non dans le contexte de la classe, soit le contexte le plus écologique pour l'élève. Bien que le lieu de passation de la tâche expérimentale en se basant uniquement sur la ressemblance physique à l'environnement quotidien du participant soit un facteur de validité écologique, la validation écologique considère aussi le réalisme et l'authenticité des processus psychologiques (ou cognitifs) observés (Brewer et Crano, 2014). De ce fait, la tâche effectuée en laboratoire est une tâche authentique de lecture de mots susceptibles d'être rencontrés par des élèves du 3^e cycle du primaire qui a servi à observer des processus cognitifs relatifs au traitement morphographique et naturellement présents durant l'identification des mots écrits isolés chez des élèves de cet âge. De plus, il aurait été impossible d'observer aussi précisément les mêmes processus cognitifs si les mots avaient été intégrés dans des phrases. D'autres processus seraient intervenus dans la relation à l'étude, dont le traitement morphosyntaxique. Aussi, les choix méthodologiques ont été sélectionnés de façon à augmenter le plus possible la validité écologique de la recherche. Par exemple, la présentation des items en noir sur fond blanc durant la tâche de décision lexicale a été privilégiée plutôt que son contraste puisque c'est ce format qui est principalement utilisé dans les livres et les manuels scolaires.

Une autre limite importante de l'étude réside dans la sélection des fenêtres temporelles des potentiels évoqués observés. Nous souhaitons observer des processus cognitifs très rapprochés dans le temps : la décomposition morpho-orthographique vers 250 ms poststimulus et l'accès lexical vers 400 ms poststimulus. La latence des deux fenêtres temporelles a été sélectionnée sur la base des choix méthodologiques et des résultats d'études antérieures menées auprès d'adultes lecteurs experts (p. ex. Davis *et al.*, 2019; El Yagoubi *et al.*, 2008; Holcomb et Grainger, 2006; McKinnon *et al.*, 2003). Il est possible que, pour des lecteurs intermédiaires, les processus cognitifs soient plus tardifs et qu'il soit nécessaire de décaler les fenêtres d'observation. D'ailleurs, l'analyse visuelle des tracés des potentiels évoqués (Figures 4.2, 4.4, 4.5, 4.7, 4.8, 4.10, 4.11) confirme cette hypothèse : les effets sont plutôt détectés à partir de 300 ms. L'utilisation de fenêtres temporelles qui contiennent peu ou pas d'effets peut avoir effacé certains résultats. De plus, les fenêtres temporelles sélectionnées étaient contigües (200 à 350 ms et 350 à 500 ms). Sans que cela soit nécessairement problématique, cette contigüité peut

entraîner un chevauchement de l'activation cérébrale dans les deux fenêtres temporelles pour certains items chez certains participants. Comme la signature électroencéphalographique de chaque individu est unique, il est possible que pour certains items aucune activation cérébrale associée à la décomposition morpho-orthographique n'ait été détectée, mais que le pic maximal d'un N400 ait été détecté de 200 à 350 ms. Cela aurait été interprété à tort comme une décomposition morpho-orthographique et non comme reflétant les processus d'accès lexical. Bien que les analyses de potentiels évoqués soient réalisées à partir de la moyenne de nombreux stimulus pour une même catégorie et pour plusieurs participants et que cette moyenne est peu susceptible d'être influencée par quelques-uns de ces cas problématiques, une prévalence importante de ces situations pourrait avoir une incidence majeure sur les données. Ainsi, pour mieux identifier les fenêtres temporelles d'observation des potentiels évoqués et pour identifier plus précisément les effets des variables sur les potentiels évoqués, il pourrait être pertinent d'envisager une sélection manuelle des pics d'activation cérébrale plutôt qu'une sélection automatique ancrée dans des fenêtres temporelles. De plus, comme cela a déjà été mentionné, des tâches d'amorçage masqué devraient être privilégiées pour déclencher les éventuels processus de décomposition morpho-orthographique chez les élèves.

Enfin, une limite intrinsèque à la tâche de décision lexicale est le non-respect du principe de Hillyard qui devrait être appliqué dans les tâches de potentiels évoqués (Luck, 2014). Dans le cadre de la présente recherche, nous avons tenté de contrôler le plus possible différentes variables psycholinguistiques de nos stimulus. Cependant, il a été impossible de respecter le principe de Hillyard. Afin de réduire les confusions sensorielles attribuables aux listes expérimentales, ce principe suggère que les séquences de stimulus devraient demeurer constantes et seule la consigne de la tâche devrait varier d'une condition expérimentale à l'autre. Avec des expérimentations sur la langue, comme les tâches de décision lexicale, il est impossible de respecter ce principe (Luck, 2014). Même si ces confusions sensorielles affectent surtout les potentiels évoqués précoces, des effets significatifs peuvent aussi être détectés jusqu'à 300 ms (Luck, 2014). Par conséquent, dans de futures recherches, des précautions devraient être prises pour réduire les biais associés au non-respect du principe de Hillyard. Pour les tâches expérimentales sur les mots, Luck (2014) suggère d'ajouter une condition de contrôle où les mêmes mots que ceux des conditions expérimentales seraient présentés à l'envers : une absence d'effets dans cette condition confirmerait qu'il n'y a pas de confusion sensorielle. L'utilisation d'une tâche de décision lexicale jumelée au paradigme d'amorçage pourrait aussi être une option permettant de s'assurer que tous les stimuli sont présentés dans chaque condition expérimentale (Royle et Steinhauer, 2023). Enfin, lorsque le non-

respect du principe de Hillyard est inévitable, des analyses statistiques comme les modèles linéaires mixtes intégrant les effets des items comme facteurs aléatoires pourraient réduire les possibles biais (Royle et Steinhauer, 2023).

Les résultats de cette présente étude soulèvent de nouvelles questions et nous mènent à envisager de futurs projets de recherche. D'abord, les recherches antérieures qui ont porté sur le traitement morphographique des mots composés ont été réalisées auprès de participants adultes dont la plupart étaient anglophones alors que notre étude a été menée auprès d'enfants francophones. Cette différence majeure nous amène à réfléchir à une future recherche qui répliquerait la présente étude auprès d'une population adulte. Ainsi, cette recherche pourrait poursuivre deux principaux objectifs. D'abord, la comparaison des résultats de la présente étude réalisée auprès d'enfants avec ceux recueillis auprès d'adultes pourrait renseigner sur le développement du traitement morphographique des mots composés formés de racines latines et grecques chez les francophones, du 3^e cycle du primaire à l'université. Aussi, il serait intéressant de comparer ces nouvelles données collectées auprès d'une population adulte francophone aux résultats des études antérieures menées auprès d'une population adulte anglophone. Ainsi, on pourrait vérifier si la langue parlée influence le traitement morphographique des mots composés auprès de populations semblables.

Aussi, compte tenu de l'influence de la fréquence des mots et de celle des racines dans le traitement morphographique, il serait intéressant de considérer les indices de fréquence dès la sélection des items. De futures recherches pourraient porter sur les effets de la fréquence des mots et de celle des racines sur le traitement des items formés de racines latines et grecques, comme la précision et le temps de réaction, auprès d'élèves du 3^e cycle du primaire. Quatre conditions expérimentales, ainsi que leurs équivalents en pseudomots, pourraient être comparées lors d'une tâche de décision lexicale : 1) des mots fréquents composés de racines latines ou grecques fréquentes (p. ex. *téléphone*); 2) des mots rares composés de racines latines ou grecques fréquentes (p. ex. *manuscrit*); 3) des mots fréquents composés de racines latines ou grecques rares (p. ex. *pétrole*); ainsi que 4) des mots rares composés de racines latines ou grecques rares (p. ex. *ptérodactyle*). Sur la base des résultats de la présente étude, il est possible d'émettre l'hypothèse que les mots fréquents seraient identifiés plus rapidement et plus précisément que les mots rares. La fréquence des racines influencerait l'identification des pseudomots. Il est aussi possible que cet indice de fréquence influence l'identification des mots écrits rares puisqu'ils sont susceptibles d'être inconnus des élèves du primaire et, donc, traités comme des pseudomots.

De plus, le fait que les résultats de la présente recherche et ceux de Whissell-Turner et Fejzo (2021) soutiennent que les élèves du 3^e cycle du primaire connaissent environ 50 % des racines latines et grecques présentes dans leurs manuels scolaires nous mène à envisager une future recherche. En effet, la méconnaissance des racines latines et grecques peut être une raison qui explique le taux élevé d'erreurs à la tâche de décision lexicale ($M = 21,12\%$). Si les élèves du 3^e cycle du primaire connaissent les racines latines et grecques des mots et des pseudomots présentés, il est possible que les résultats à la tâche de décision lexicale soit meilleurs et qu'une plus grande quantité d'items soient retenus, ce qui pourrait exercer une influence sur les résultats des potentiels évoqués. Aussi, mieux connaître la structure morphologique des mots formés de racines latines et grecques peut influencer leur identification. Ainsi, il serait intéressant de mener une recherche qui présente un devis quasi-expérimental avec prétests, intervention morphologique sur les racines latines et grecques et posttests afin de mesurer la différence dans le traitement morphographique des mots formés de racines latines et grecques à la suite d'un enseignement.

Comme discuté au chapitre 5, l'amorçage est un paradigme qui devrait être utilisé dans des recherches subséquentes pour mieux distinguer le traitement morphographique des traitements orthographique et sémantique. Des recherches antérieures ont déjà permis de constater que la présentation de l'amorce dérivée (p. ex. Andrews et Lo, 2013; Diependaele *et al.*, 2009; Forster et Azuma, 2000) ou fléchie (p. ex. Royle *et al.*, 2019) facilitait l'activation de la cible, soit la racine d'un mot. Cependant, avec ce type de mots, l'amorce et la cible partagent des représentations orthographiques et sémantiques. Avec les racines latines et grecques, il serait possible de minimiser ces chevauchements toujours en vérifiant si la présentation de la racine du mot (ou de son sens) en amorce facilite le traitement de la cible. Par exemple, nous pourrions vérifier si la présentation des amorces *hexa* et *six* permettent d'accéder de la même façon au mot *hexagone* ainsi qu'au pseudomot **hexachrome*. Alors que la racine *hexa* partage des représentations orthographiques, phonologiques et sémantiques avec la cible, le sens de la racine, *six*, ne partage que des représentations sémantiques. Une telle recherche permettrait de mieux définir le rôle des racines et de leur sens dans le traitement des mots polymorphémiques.

En conclusion, l'interdisciplinarité de cette recherche a mené à l'utilisation de mesures comportementales et psychophysiologiques complémentaires pour répondre aux questions de recherche. Les mesures comportementales de temps de réaction et de précision à la tâche de décision lexicale ont permis de renseigner sur la conséquence de la mise en place des nombreux processus

d'identification des mots écrits. L'analyse des potentiels évoqués, quant à elle, a permis de mettre la loupe sur le déroulé temporel de cesdits processus en temps réel. La présente recherche a ainsi permis d'éclairer sur la façon dont les élèves du 3^e cycle du primaire traitent les mots composés de racines latines et grecques et d'identifier les facteurs qui influencent ce traitement. Cette recherche mène au constat que les élèves du 3^e cycle du primaire traitent les racines latines et grecques dès 200 ms lors de l'identification des mots écrits et que ces processus sont tributaires de différents facteurs dont la structure morphologique et la fréquence des mots et des racines. Cette recherche est la première d'une lignée d'autres recherches qui porteront sur l'identification des mots écrits formés de racines latines et grecques en contexte francophone.

ANNEXE A

CERTIFICAT D'APPROBATION ÉTHIQUE



No. de certificat : 2022-4717

Date : 2022-03-31

CERTIFICAT D'APPROBATION ÉTHIQUE

Le Comité d'éthique de la recherche pour les projets étudiants impliquant des êtres humains (CERPE plurifacultaire) a examiné le projet de recherche suivant et le juge conforme aux pratiques habituelles ainsi qu'aux normes établies par la *Politique No 54 sur l'éthique de la recherche avec des êtres humains*(2020) de l'UQAM.

- Titre du projet : **Le traitement morphographique des mots composés formés de racines latines et grecques chez les élèves du 3e cycle du primaire**
- Nom de l'étudiant : **Kathleen Whissell-Turner**
- Programme d'études : **Doctorat en éducation**
- Direction(s) de recherche : **Anila Fejzo; Julien Mercier**

Modalités d'application

Toute modification au protocole de recherche en cours de même que tout événement ou renseignement pouvant affecter l'intégrité de la recherche doivent être communiqués rapidement au comité.

La suspension ou la cessation du protocole, temporaire ou définitive, doit être communiquée au comité dans les meilleurs délais.

Le présent certificat est valide pour une durée d'un an à partir de la date d'émission. Au terme de ce délai, un rapport d'avancement de projet doit être soumis au comité, en guise de rapport final si le projet est réalisé en moins d'un an, et en guise de rapport annuel pour le projet se poursuivant sur plus d'une année au plus tard un mois avant la date d'échéance (**2023-03-31**) de votre certificat. Dans ce dernier cas, le rapport annuel permettra au comité de se prononcer sur le renouvellement du certificat d'approbation éthique.

Élise Ducharme

Pour **Raoul Graf**, M.A., Ph.D.

Président CERPÉ plurifacultaire et Professeur titulaire, département de marketing

A handwritten signature in black ink, appearing to be 'R. Graf'.

ANNEXE B
STIMULUS POUR LA TÂCHE DE DÉCISION LEXICALE

Tableau B.1 Liste des 8 mots de la tâche de décision lexicale comme items de pratique

Mots			
Polymorphémiques			Monomorphémiques
2 racines liées latines	2 racines liées grecques	2 racines libres	1 racine libre
interroge midi	astronaute bibliothèque	beaucoup chauvesouris	caribou girafe

Tableau B.2 Liste des 8 pseudomots et non-mots de la tâche de décision lexicale comme items de pratique

Pseudomots			Non-mots
2 racines liées latines	2 racines liées grecques	2 racines libres	Syllabes
interspect miroge	astrothèque biblionaute	beausouris chavecoup	carirafe gibou

Note : La transformation des mots polymorphémiques formés de deux racines latines liées *interroge* et *midi* aurait dû donner **interdi* et **miroge*. Cependant, le pseudomot **interdi* est beaucoup trop près orthographiquement du mot *interdit* qui existe en français. De ce fait, la racine *di* dans cet exemple a été changée par une autre racine latine liée, *spect*, susceptible d'avoir été rencontrée par les élèves de cet âge et qui n'est pas présente dans les stimulus du test.

Tableau B.3 Liste des 120 mots de la tâche de décision lexicale

Mots			
Polymorphémiques			Monomorphémiques
2 racines liées latines	2 racines liées grecques	2 racines libres	1 racine libre
ambidextre	aérodrome	abribus	artichaut
apicole	antonyme	aussitôt	bouteille
aqueduc	archéologue	autrefois	cactus
auditorium	autonome	bassecour	caoutchouc
bénévole	chronophage	bien-être	castor
binocle	décathlon	bonsoir	chevreuil
équinoxe	démocrate	contresens	chimpanzé
funambule	épiderme	courtepointe	crocodile
homicide	hétérogène	doubleclac	éléphant
ignifuge	hexagone	entredoux	goéland
introverti	hippocampe	essuie-tout	langouste
majordome	hydrophobe	faire-part	luzerne
manuscrit	hypocrite	gentilhomme	macaroni
multicolore	isotherme	hautbois	macédoine
nauffrage	métropole	mainforte	nénufar
omnivore	microscope	malpropre	opossum
ovipare	misogyne	marchepied	orchestre
pédicure	monochrome	millepatte	perdrix
péninsule	octopode	passe-temps	perruque
pétrole	ornithorynque	plateforme	pingouin
plurilingue	orthodoxe	porteclé	pupitre
posthume	pédiatre	pourboire	rutabaga
privilège	philosophe	rondpoint	safari
quintuple	photographe	sagefemme	salamandre
sacrifice	polyglotte	soucoupe	salami
somnifère	psychopathe	superhéros	sésame
transpire	ptérodactyle	terreplein	tambour
trimestre	pyromane	tirebouchon	tartare
univoque	rhinocéros	tournevis	toboggan
verdict	tétraèdre	vinaigre	tulipe

Tableau B.4 Liste des 120 pseudomots et non-mots de la tâche de décision lexicale

Pseudomots			Non-mots
2 racines liées latines	2 racines liées grecques	2 racines libres	Syllabes
ambinoxe	aéropathe	abritout	artibour
apidome	anticrite	aussifois	bousame
aquascrit	archéoderme	autretôt	cacri
audiduc	autotherme	basseforme	caoutteille
bénédextre	chroniatre	bienpart	casphant
bicole	décadrome	bonclit	chegouin
équimestre	démologue	contrefemme	chimbaga
funifère	épirynque	courtepatte	crodoine
homispire	hétérophage	doublepied	éléchaut
ignivoque	hexachrome	entrepointe	goéssum
introfice	hippodoxe	essuietemps	lanchestre
majinsule	hydrodactyle	fairecoupe	lumandre
manucolore	hypoèdre	gentilplein	macapanzé
multivore	isocampe	hautbouchon	macétor
naupare	métrogène	mainboire	nénudrix
omnifuge	microcéros	malaigre	opouchouc
ovicure	misonyme	marchevis	orlipe
pedambule	monogone	millepoint	pecodile
pénhume	octoglotte	passelé	perlami
pétrocle	orninome	platecour	pintare
pluriple	orthophobe	portesens	puroni
postfrage	pédopole	pourhomme	rutaggan
prividict	philomane	rondpropre	safazerne
quintilingue	photogyne	sageforte	salarruque
sacricide	polyscope	souhéros	savreuil
somnirium	psychocrate	supersoir	séland
transole	ptérosophe	terrebois	tampitre
triverti	pyrathlon	tiredeux	tartus
univole	rhinopode	tournebus	tobofar
vérilège	tétragraphe	vinêtre	tugouste

ANNEXE C

LISTES FRANCOPHONES DE VOCABULAIRE

Liste	Clientèle cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
Manulex (Lété <i>et al.</i> , 2004)	Primaire (CP, CE1, CE2-CM2)	<ul style="list-style-type: none"> - 54 manuels scolaires (1,9 million de mots) - 48 900 formes orthographiques - 23 900 lemmes 	<ul style="list-style-type: none"> - Cycle (CP/CE1/CE2-CM2) - Nombre de lettres - Classe grammaticale - Fréquence brute - Indice de dispersion - Fréquence d'usage pour 1 million de mots 	<p>(+) Présence des formes orthographiques ainsi que des lemmes</p> <p>(-) La liste s'arrête en CM2, ce qui équivaut à la 5e année du primaire au Québec</p> <p>(-) Vocabulaire français : certains termes ne sont pas utilisés en contexte québécois et vice versa</p> <p>(-) Ne tient pas compte des locutions ou des mots composés avec espace (p. ex. cuillère à table)</p> <p>(-) Les classes grammaticales des mots ne tiennent pas compte de la nouvelle grammaire</p>
	Type de vocabulaire			
	Courant, mixte			
Liste orthographique du MELS (Ministère de l'Éducation, du Loisir et du Sport, 2014)	Primaire (1 ^{re} à 6 ^e année)	<ul style="list-style-type: none"> - 250 ouvrages de littérature jeunesse sélectionnés sur "Livres ouverts" - 19 112 mots 	<ul style="list-style-type: none"> - Année scolaire - Phonétique - Classe de mots - Genre et nombre - Nouvelles graphies - Variations en genre et en nombre - Verbes modèles - Thèmes - Formation du féminin et du pluriel - Composition 	<p>(+) Contexte québécois</p> <p>(+) Possibilité de naviguer par année scolaire</p> <p>(+) Possibilité de voir le mot en contexte phrastique</p> <p>(-) Il s'agit d'une liste orthographique et non d'une liste de vocabulaire</p>
	Type de vocabulaire			
	Courant, mixte			

Liste	Clientèle cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
			(juxtaposition, soudure, trait d'union) - Dérivation (lettres muettes)	
Lexique 3 (New et al., 2001)	Adultes	<ul style="list-style-type: none"> - 218 textes littéraires (romans) publiés entre 1950 et 2000 (14,7 millions de mots) - Sous-titres de 9474 films ou saisons de séries (50 millions de mots) - 135 000 mots - 55 000 lemmes 	<ul style="list-style-type: none"> - Fréquence écrite - Genre et nombre - Catégorie grammaticale - Phonologie - Nombre de voisins orthographiques et phonologiques - Fréquence des lettres, phonèmes, bigrammes, trigrammes, syllabes - Etc. 	(+) Plusieurs variables sur les mots (-) Le corpus n'est pas constitué à l'aide de matériel destiné aux enfants
	Type de vocabulaire			
	Courant, mixte			
Vocabulaire général d'orientation scientifique (V.G.O.S.; Phal, 1971)	Secondaire	<ul style="list-style-type: none"> - 24 manuels de 4^e à 1^{re} et terminales en France : physique, mathématiques, chimie, sciences naturelles - 1,8 million de mots 	<ul style="list-style-type: none"> - Fréquence et dispersion - Antonymes 	(-) Le corpus n'est pas constitué à l'aide de matériel destiné aux enfants (-) Non récent
	Type de vocabulaire			
	Transdisciplinaire			

Liste	Cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
Lexique scientifique transdisciplinaire (Drouin, 2007)	Adultes	<ul style="list-style-type: none"> - Corpus de thèses : psychologie, droit, histoire, géographie, archéologie, physique, ingénierie, informatique, chimie - 2,3 millions de mots - Mots dispersés dans la moitié des tranches de textes et avec une spécificité > 3,09 par rapport au corpus de référence : <i>Le Monde</i> 	<ul style="list-style-type: none"> - Fréquence, dispersion et spécificité - Collocations - Classes grammaticales : noms, adjectifs, verbes, adverbes - Lemmes 	<p>(+) Lemmatisation automatique avec l'étiqueteur morphosyntaxique TreeTagger(-)</p> <p>) Le corpus n'est pas constitué à l'aide de matériel destiné aux enfants</p> <p>(-) Tous les mots ne sont pas répertoriés (p. ex. déterminants)</p>
	Type de vocabulaire			
	Transdisciplinaire			
Liste transdisciplinaire (Tutin, 2008)	Adultes	<ul style="list-style-type: none"> - Corpus d'articles scientifiques, de thèses, de rapports de recherche, de cours (inclusion du corpus KIAP du français) : médecine, linguistique, économie - 2 millions de mots - Comprend les mots apparaissant plus de 15 fois dans les trois disciplines. 	<ul style="list-style-type: none"> - Fréquence par discipline et totale - Classe grammaticale - Distance sémantique entre les mots 	<p>(-) Vocabulaire transdisciplinaire uniquement</p> <p>(-) Le corpus n'est pas constitué à l'aide de matériel destiné aux enfants</p> <p>(-) Tous les mots ne sont pas répertoriés (p. ex. déterminants)</p>
	Type de vocabulaire			
	Transdisciplinaire			

Liste	Cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
KIAP (Fløttum <i>et al.</i> , 2013)	Adultes	<ul style="list-style-type: none"> - 450 articles de recherche : économie, linguistique, médecine (3 millions de mots) - Publiés de 1992 à 2003 - 3 langues : anglais, français, norvégien 		(-) La liste ne semble plus disponible en ligne (-) Le corpus n'est pas constitué à l'aide de matériel destiné aux enfants
	Type de vocabulaire			
	Courant/mixte			
Lexique Scientifique Transdisciplinaire Junior (Tremblay <i>et al.</i> , 2023; Tremblay <i>et al.</i> , 2018)	Primaire	<ul style="list-style-type: none"> - Manuels scolaires du 1^{er} cycle du primaire au 3^e cycle du primaire : mathématiques, éthique et culture religieuse, français, histoire, géographie et éducation à la citoyenneté - 90 manuels scolaires et 35 cahiers d'exercices des éditions La Chenelière 	<ul style="list-style-type: none"> - Fréquence, dispersion, spécificité - Classe grammaticale - Lemmes 	(+) Contexte québécois (+) Lemmatisation automatique avec l'étiqueteur morphosyntaxique TreeTagger(-)) Retrait du vocabulaire courant (-) Aucun déterminant, aucune préposition (-) Liste en construction, donc certaines informations absentes
	Type de vocabulaire			
	Transdisciplinaire et disciplinaire			

Liste	Clientèle cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
Mots simples du lexique scientifique transdisciplinaire (Hatier, 2013)	Adultes	<ul style="list-style-type: none"> - 339 articles de revue - 3 511 716 mots provenant de 10 disciplines en sciences humaines et sociales : anthropologie, économie, géographie, histoire, linguistique, sciences de l'éducation, sciences politiques, sciences de l'information, sociologie, psychologie - Corpus de référence : Lexique 3 - Les mots doivent apparaître dans un minimum de 50 tranches, et un minimum de 5 disciplines sur 10. 	<ul style="list-style-type: none"> - Dispersion - Fréquence - Spécificité - 3 classements (LST : lexique scientifique transdisciplinaire et lexique abstrait général; LT : lexique terminologique; LG : lexique de la langue générale) 	(-) Noms et verbes uniquement
	Type de vocabulaire			
	Transdisciplinaire			

Liste	Clientèle cible	Composition	Informations contenues	Forces (+) et faiblesses (-)
ÉQOL (Stanké <i>et al.</i> , 2019)	Primaire (1 ^{re} à 6 ^e année)	<ul style="list-style-type: none"> - 15 000 mots fréquents tirés des manuels scolaires des élèves 	<ul style="list-style-type: none"> - Forme phonologique - Catégorie grammaticale - Genre et nombre - Nombre de lettres - Nombre de lettres géminées - Nombre de phonèmes - Nombre de syllabes orales et écrites - Fréquence - Taux de réussite orthographique 	<p>(+) Contexte québécois</p> <p>(-) Il s'agit d'une liste orthographique (orthographe lexicale) de 1200 mots sélectionnés selon leur fréquence et non une liste de vocabulaire exhaustive</p> <p>(-) L'absence de la variable de dispersion ne permet pas de distinguer le vocabulaire courant, transdisciplinaire et disciplinaire</p> <p>(-) Les variables linguistiques sont tirées de Manulex (Lété <i>et al.</i>, 2004)</p>
	Type de vocabulaire	<ul style="list-style-type: none"> - Sélection des 1200 mots les plus fréquents pour concevoir la liste orthographique 		
	Courant/mixte	<ul style="list-style-type: none"> - 28 manuels scolaires des Éditions Erpi : mathématique, français, univers social - 24 livrets de littérature - 90 extraits de littérature jeunesse - Corpus de référence : Manulex 		

ANNEXE D

PROPRIÉTÉS PSYCHOLINGUISTIQUES DES STIMULUS DE LA TÂCHE DE DÉCISION LEXICALE ET DE LEURS RACINES

Tableau D.1 Propriétés psycholinguistiques des mots formés de deux racines liées latines et de leurs racines

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	A midi	4	2	4	172	1471	2096	3	15	12	1	1	1
	B interroge	9	3	7	183	8955	317	3	29	3	1	1	1
Expérimentation	1 ignifuge	8	3	7	0	26	136	0	2	5	—	3	1
	2 auditorium	10	4	8	0	150	112	0	7	2	—	1	1
	3 ambidextre	10	3	9	0	0	6	0	0	1	—	—	3
	4 apicole	7	3	7	0	0	4845	0	0	8	—	—	1
	5 aqueduc	7	3	6	32	175	9880	0	5	3	2	1	1
	6 bénévole	8	3	8	18	56	46	1	4	4	1	1	1
	7 binocle	7	2	7	0	930	99	0	9	3	—	1	1
	8 homicide	8	3	8	0	7	56	0	1	5	—	3	1
	9 quintuple	9	2	7	0	0	1961	0	0	8	—	—	1
	10 privilège	9	3	9	24	103	186	1	3	4	2	1	2
	11 équinoxe	8	3	7	28	1578	71	1	15	1	1	1	1
	12 introverti	10	4	9	0	291	738	0	2	15	—	1	1
	13 funambule	9	3	8	21	21	53	1	1	3	1	1	1
	14 transpire	9	2	8	19	3072	980	2	15	13	2	1	1
	15 majordome	9	3	9	0	1013	757	0	7	8	—	1	1
	16 manuscrit	9	3	9	29	1837	1539	2	15	12	2	1	1
	17 multicolore	11	4	11	38	1914	1089	1	8	6	1	1	1
	18 naufrage	8	2	7	35	957	150	2	3	4	2	1	1
	19 sacrifice	9	3	9	23	426	2842	1	6	25	2	1	1
	20 omnivore	8	3	8	9	9	216	1	1	7	2	2	1
	21 ovipare	7	3	7	14	150	1293	1	6	10	2	1	1
	22 pétrole	7	2	7	164	182	344	2	3	3	1	1	1
	23 pédicure	8	3	8	0	352	846	0	6	13	—	1	1

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
24	péninsule	9	3	8	26	46	26	1	2	1	2	2	2
25	plurilingue	11	3	9	0	1092	63	0	2	2	—	1	1
26	posthume	8	2	8	0	8	0	0	1	0	—	2	—
27	somnifère	9	3	9	0	16	6723	0	2	4	—	1	1
28	trimestre	9	2	9	0	1051	24	0	10	0	—	1	—
29	univoque	8	3	7	0	6618	514	0	19	3	—	1	1
30	verdict	7	2	7	0	368	3293	0	4	8	—	1	1
Moyenne ¹		8,53	2,83	8,00	16,00	748,27	1296,27	0,57	5,30	6,03			
Écart type ¹		1,14	0,59	1,05	30,92	1335,48	2253,04	0,73	5,09	5,42			

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.2 Propriétés psycholinguistiques des mots formés de deux racines liées grecques et de leurs racines

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	C astronaute	10	3	9	208	509	280	1	6	4	1	1	1
	D bibliothèque	12	4	10	430	515	430	2	5	1	1	1	1
Expérimentation	31 misogynie	8	3	8	0	0	0	0	0	0	—	—	—
	32 aérodrome	9	4	9	0	92	29	0	3	2	—	1	1
	33 antonyme	8	3	7	128	253	523	1	6	4	2	1	2
	34 archéologue	11	4	9	136	414	2537	3	4	43	2	1	1
	35 autonome	8	3	7	0	1132	13665	1	12	22	—	1	1
	36 orthodoxe	9	3	8	84	1620	90	2	5	3	2	1	2
	37 chronophage	11	3	9	0	143	12	0	4	1	—	1	2
	38 décathlon	9	3	7	0	13	191	0	1	2	—	2	1
	39 démocrate	9	3	9	0	304	264	0	2	2	—	2	2
	40 épiderme	8	3	8	6	105	17	1	4	2	2	1	2
	41 hétérogène	10	4	10	0	0	332	0	0	9	—	—	1
	42 hexagone	8	3	8	61	105	858	3	3	11	2	2	1
	43 hippocampe	10	3	8	37	47	37	1	2	1	1	1	1
	44 hydrophobe	10	3	9	0	284	21	0	7	1	—	2	3
	45 hypocrite	9	3	9	0	239	86	0	3	1	—	2	2
	46 isotherme	9	3	8	0	168	264	0	2	4	—	1	1
	47 métropole	9	3	9	0	0	366	0	0	3	—	—	1
	48 microscope	10	3	10	37	132	198	2	7	4	1	1	1
	49 monochrome	10	3	9	0	12	0	0	1	0	—	2	—
	50 octopode	8	3	8	0	64	0	0	4	0	—	1	—
51 ornithorynque	13	4	10	13	31	13	1	3	1	2	2	2	
52 pédiatre	8	3	8	0	254	20	1	4	1	—	1	2	
53 philosophe	10	3	8	33	39	33	2	3	2	2	2	2	
54 photographe	11	3	9	0	1231	3556	2	5	31	—	1	1	
55 polyglotte	10	3	9	0	1013	0	0	6	0	—	1	—	
56 psychopathe	11	3	9	0	44	46	0	2	1	—	2	1	
57 ptérodactyle	12	4	12	0	0	0	0	0	0	—	—	—	
58 pyromane	8	3	8	0	0	12	0	0	1	—	—	2	
59 rhinocéros	10	4	10	11	11	11	1	1	1	1	1	1	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
	60 tétraèdre	9	3	9	0	6	337	0	1	2	—	2	1
Moyenne ¹		9,50	3,20	8,70	18,20	258,53	783,93	0,70	3,17	5,17			
Écart type ¹		1,31	0,41	1,06	37,08	418,54	2552,77	0,95	2,68	9,82			

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.3 Propriétés psycholinguistiques des mots formés de deux racines libres et de leurs racines

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	E chauvesouris	12	3	9	226	236	576	2	3	5	1	1	1
	F beaucoup	8	2	5	2078	3720	4032	1	4	12	1	1	1
Expérimentation	61 abribus	7	3	7	0	309	246	0	2	1	—	1	1
	62 vinaigre	8	2	7	24	38	38	2	2	2	1	1	1
	63 aussitôt	8	3	6	220	5107	1386	1	2	5	1	1	1
	64 autrefois	9	3	7	309	8452	4933	1	4	5	1	1	1
	65 bassecour	9	2	7	17	1035	1937	1	7	3	1	1	1
	66 tirebouchon	12	3	8	0	1243	44	0	14	1	—	1	1
	67 bienêtre	8	2	7	52	8527	1608	1	12	4	1	1	1
	68 bonsoir	7	2	5	20	4867	794	1	14	3	1	1	1
	69 contresens	10	3	8	0	2131	2327	0	16	9	—	1	1
	70 courtrepoin	12	3	9	49	817	3548	1	4	12	1	1	1
	71 doubleclic	10	3	9	0	390	57	0	5	2	—	1	1
	72 entredeux	9	3	7	0	585	91	0	11	4	—	1	1
	73 essuietout	10	3	8	16	0	5009	1	2	9	1	1	1
	74 fairepart	9	2	8	0	1897	12041	0	18	26	—	1	1
	75 gentilhomme	11	3	8	0	220	2035	0	3	5	—	1	1
	76 hautbois	8	2	6	0	1869	1431	0	8	6	—	1	1
	77 mainforte	9	2	7	0	2571	2327	0	4	16	—	1	1
	78 malpropre	9	2	9	0	1905	2062	0	25	9	—	1	1
	79 marchepied	10	3	9	0	2288	824	0	9	1	—	1	1
80 millepatte	10	2	7	7	799	385	1	9	2	1	1	1	
81 passetemps	10	2	8	26	4222	6453	1	13	11	1	1	1	
82 plateforme	10	2	10	97	734	10067	1	8	28	1	1	1	
83 porteclé	8	3	8	10	9486	465	1	35	5	1	1	1	
84 rondpoint	9	2	6	0	917	3548	0	5	12	—	1	1	
85 sagefemme	9	2	7	6	178	1373	1	5	2	2	1	1	
86 superhéros	10	4	10	101	932	309	2	19	5	1	1	1	
87 terreplein	10	2	7	0	6911	710	0	20	1	—	1	1	
88 tournevis	9	3	8	0	1398	68	0	10	2	—	1	1	
89 pourboire	9	2	7	0	4009	370	0	12	2	—	1	1	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
	90 soucoupe	8	2	6	14	8242	151	1	32	3	1	1	1
Moyenne ¹		9,23	2,50	7,53	32,27	2735,97	2221,23	0,57	11,00	6,53			
Écart type ¹		1,22	0,57	1,20	69,47	2905,56	2927,18	0,63	8,59	6,79			

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.4 Propriétés psycholinguistiques des mots monomorphémiques

		Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
						Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	G	caribou	7	3	6	75	—	—	1	—	—	1	—	—
	H	girafe	6	2	6	117	—	—	1	—	—	1	—	—
Expérimentation	91	caoutchouc	10	3	7	38	—	—	1	—	—	1	—	—
	92	salamandre	10	3	9	11	—	—	1	—	—	1	—	—
	93	artichaut	9	3	7	8	—	—	1	—	—	1	—	—
	94	bouteille	9	2	6	305	—	—	1	—	—	1	—	—
	95	chevreuil	9	2	6	91	—	—	1	—	—	1	—	—
	96	chimpanzé	9	3	6	44	—	—	1	—	—	1	—	—
	97	crocodile	9	3	9	49	—	—	1	—	—	1	—	—
	98	sésame	6	2	6	7	—	—	1	—	—	3	—	—
	99	macédoine	9	3	8	0	—	—	0	—	—	—	—	—
	100	orchestre	9	2	8	0	—	—	0	—	—	—	—	—
	101	éléphant	8	3	6	199	—	—	2	—	—	1	—	—
	102	macaroni	8	4	8	41	—	—	1	—	—	1	—	—
	103	perruque	8	2	6	22	—	—	1	—	—	1	—	—
	104	pingouin	8	2	5	25	—	—	1	—	—	1	—	—
	105	rutabaga	8	4	8	0	—	—	0	—	—	—	—	—
	106	toboggan	8	3	6	51	—	—	1	—	—	1	—	—
	107	cactus	6	2	6	18	—	—	1	—	—	1	—	—
	108	safari	6	3	6	0	—	—	0	—	—	—	—	—
	109	luzerne	7	2	7	0	—	—	0	—	—	—	—	—
110	langouste	9	2	7	9	—	—	1	—	—	3	—	—	
111	nénufar	7	3	7	38	—	—	1	—	—	1	—	—	
112	salami	6	3	6	0	—	—	0	—	—	—	—	—	
113	opossum	7	3	5	0	—	—	0	—	—	—	—	—	
114	perdrix	7	2	7	26	—	—	1	—	—	1	—	—	
115	pupitre	7	2	7	0	—	—	0	—	—	—	—	—	
116	tambour	7	2	5	96	—	—	2	—	—	1	—	—	
117	tartare	7	2	7	0	—	—	0	—	—	—	—	—	
118	castor	6	2	6	212	—	—	1	—	—	1	—	—	
119	goéland	7	3	6	18	—	—	1	—	—	1	—	—	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
120	tulipe	6	2	6	50	—	—	1	—	—	1	—	—
Moyenne ¹		7,73	2,57	6,63	45,27	—	—	0,77	—	—		—	—
Écart type ¹		1,26	0,63	1,07	72,00	—	—	0,57	—	—		—	—

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.5 Propriétés psycholinguistiques des pseudomots formés de deux racines liées latines et de leurs racines

		Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
						Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	I	interspect	10	3	9	—	8955	2424	—	29	10	—	1	1
	J	miroge	6	2	6	—	1471	317	—	15	3	—	1	1
Expérimentation	121	pluriple	8	2	8	—	1092	1961	—	2	8	—	1	1
	122	univole	7	3	7	—	6618	46	—	19	4	—	1	1
	123	transole	8	2	7	—	3072	344	—	15	3	—	1	1
	124	introfice	9	3	8	—	291	2842	—	2	25	—	1	1
	125	équimestre	10	3	9	—	1578	24	—	15	0	—	1	—
	126	somnirium	9	3	8	—	16	112	—	2	2	—	1	1
	127	aquascrit	9	3	8	—	175	1539	—	5	12	—	1	1
	128	multivore	9	3	9	—	1914	216	—	8	7	—	1	1
	129	quintilingue	12	3	8	—	0	63	—	0	2	—	—	1
	130	bénédextre	10	3	10	—	56	6	—	4	1	—	1	3
	131	ovicure	7	3	7	—	150	846	—	6	13	—	1	1
	132	triverti	8	3	8	—	1051	738	—	10	15	—	1	1
	133	homispire	9	3	9	—	7	980	—	1	13	—	3	1
	134	apidome	7	3	7	—	0	757	—	0	8	—	—	1
	135	pedambule	9	3	8	—	352	53	—	6	3	—	1	1
	136	ignivoque	9	3	7	—	26	514	—	2	3	—	3	1
	137	vérilège	8	3	8	—	368	186	—	4	4	—	1	2
	138	omnifuge	8	3	8	—	9	136	—	1	5	—	2	1
	139	majinsule	9	3	8	—	1013	26	—	7	1	—	1	2
140	postfrage	9	2	9	—	8	150	—	1	4	—	2	1	
141	sacricide	9	3	9	—	426	56	—	6	5	—	1	1	
142	bicole	6	2	6	—	930	4845	—	9	8	—	1	1	
143	pétrocle	8	2	8	—	182	99	—	3	3	—	1	1	
144	pénhume	7	2	7	—	46	0	—	2	0	—	2	—	
145	manucolore	10	4	10	—	1837	1089	—	15	6	—	1	1	
146	funifère	8	3	8	—	21	6723	—	1	4	—	1	1	
147	ambinoxe	8	3	7	—	0	71	—	0	1	—	—	1	
148	audiduc	7	3	6	—	150	9880	—	7	3	—	1	1	
149	naupare	7	2	6	—	957	1293	—	3	10	—	1	1	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
	150 prividict	9	3	9	—	103	3293	—	3	8	—	1	1
Moyenne ¹		8,43	2,80	7,90	—	748,27	1296,27	—	5,30	6,03	—		
Écart type ¹		1,22	0,48	1,06	—	1335,48	2253,04	—	5,09	5,42	—		

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.6 Propriétés psycholinguistiques des pseudomots formés de deux racines liées grecques et de leurs racines

		Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
						Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	K	astrothèque	11	3	9	—	49	430	—	6	1	—	1	1
	L	biblioteute	11	4	10	—	515	280	—	5	4	—	1	1
Expérimentation	151	ptérosophe	10	3	9	—	0	33	—	0	2	—	—	2
	152	anticrite	9	3	8	—	253	86	—	6	1	—	1	2
	153	octoglotte	10	3	9	—	64	0	—	4	0	—	1	—
	154	monogone	8	3	8	—	12	858	—	1	11	—	2	1
	155	hétérophage	11	4	10	—	0	12	—	0	1	—	—	2
	156	chroniatre	10	2	9	—	143	20	—	4	1	—	1	2
	157	orthophobe	10	3	8	—	1620	21	—	5	1	—	1	3
	158	isocampe	8	3	7	—	168	37	—	2	1	—	1	1
	159	hypoèdre	8	3	8	—	239	337	—	3	2	—	2	1
	160	photogyne	9	3	8	—	1231	0	—	5	0	—	1	—
	161	polyscope	9	3	9	—	1013	198	—	6	4	—	1	1
	162	métrogène	9	3	9	—	0	332	—	0	9	—	—	1
	163	épirynque	9	3	7	—	105	13	—	4	1	—	1	2
	164	hexachrome	10	3	9	—	105	0	—	3	0	—	2	—
	165	microcéros	10	4	10	—	132	11	—	7	1	—	1	1
	166	philomane	9	3	8	—	39	12	—	3	1	—	2	2
	167	rhinopode	9	3	9	—	11	0	—	1	0	—	1	—
	168	aéropathe	9	4	8	—	92	46	—	3	1	—	1	1
	169	psychocrate	11	3	10	—	44	264	—	2	2	—	2	2
170	orninome	8	3	8	—	31	13665	—	3	22	—	2	1	
171	pyrathlon	9	3	7	—	0	191	—	0	2	—	—	1	
172	archéoderme	11	4	10	—	414	17	—	4	2	—	1	2	
173	démologue	9	3	8	—	304	2537	—	2	43	—	2	1	
174	pédopole	8	3	8	—	254	366	—	4	3	—	1	1	
175	décadrome	9	3	9	—	13	29	—	1	2	—	2	1	
176	hydrodactyle	12	4	12	—	284	0	—	7	0	—	2	—	
177	hippodoxe	9	3	8	—	47	90	—	2	3	—	1	2	
178	misonyme	8	3	8	—	0	523	—	0	4	—	—	2	
179	tétragraphe	11	3	10	—	6	3556	—	1	31	—	2	1	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
	180 autotherme	10	3	8	—	1132	264	—	12	4	—	1	1
Moyenne ¹		9,40	3,13	8,63	—	258,53	783,93	—	3,17	5,17	—		
Écart type ¹		1,07	0,43	1,10	—	418,54	2552,77	—	2,68	9,82	—		

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.7 Propriétés psycholinguistiques des pseudomots formés de deux racines libres et de leurs racines

		Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
						Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	M	beausouris	10	3	7	—	3720	576	—	4	5	—	1	1
	N	chauvecoup	10	2	7	—	236	4032	—	3	12	—	1	1
Expérimentation	181	bienpart	8	2	7	—	8527	12041	—	12	26	—	1	1
	182	hautbouchon	11	3	7	—	1869	44	—	8	1	—	1	1
	183	aussifois	9	3	6	—	5107	4933	—	2	5	—	1	1
	184	fairecoupe	10	2	8	—	1897	151	—	18	3	—	1	1
	185	platecour	9	2	8	—	734	1937	—	8	3	—	1	1
	186	contrefemme	11	3	8	—	2131	1373	—	16	2	—	1	1
	187	portesens	9	3	8	—	9486	2327	—	35	9	—	1	1
	188	gentilplein	11	3	8	—	220	710	—	3	1	—	1	1
	189	souhéros	8	3	7	—	8242	309	—	32	5	—	1	1
	190	entrepointe	11	3	8	—	585	3548	—	11	12	—	1	1
	191	passecélé	8	2	7	—	4222	465	—	13	5	—	1	1
	192	bonclit	7	2	6	—	4867	57	—	14	2	—	1	1
	193	terrebois	9	2	7	—	6911	1431	—	20	6	—	1	1
	194	supersoir	8	3	8	—	932	794	—	19	3	—	1	1
	195	mainboire	9	2	6	—	2571	370	—	4	2	—	1	1
	196	tournebus	9	3	8	—	1398	246	—	10	1	—	1	1
	197	sageforte	9	2	9	—	178	2327	—	5	16	—	1	1
	198	marchevis	9	3	8	—	2288	68	—	9	2	—	1	1
	199	rondpropre	10	2	9	—	917	2062	—	5	9	—	1	1
	200	basseforme	10	2	9	—	1035	10067	—	7	28	—	1	1
	201	doubleped	10	3	9	—	390	824	—	5	1	—	1	1
202	tiredeux	8	2	7	—	1243	91	—	14	4	—	1	1	
203	pourhomme	9	2	7	—	4009	2035	—	12	5	—	1	1	
204	abritout	8	3	7	—	309	5009	—	2	9	—	1	1	
205	autretôt	8	3	7	—	8452	1386	—	4	5	—	1	1	
206	vinêtre	7	2	7	—	38	1608	—	2	4	—	1	1	
207	malaigre	8	2	7	—	1905	38	—	25	2	—	1	1	
208	millepoint	10	2	6	—	799	3548	—	9	12	—	1	1	
209	courtepatte	11	3	9	—	817	385	—	4	2	—	1	1	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
	210 essuietemps	11	3	9	—	0	6453	—	2	11	—	1	1
Moyenne ¹		9,17	2,50	7,57	—	2735,9 7	2221,2 3	—	11,00	6,53	—		
Écart type ¹		1,23	0,51	0,97	—	2905,5 6	2927,1 8	—	8,59	6,79	—		

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

Tableau D.8 Propriétés psycholinguistiques des non-mots

		Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
						Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
Entraînement	O	carirafe	8	3	8	—	—	—	—	—	—	—	—	—
	P	gibou	5	2	4	—	—	—	—	—	—	—	—	—
Expérimentation	211	lanchestre	10	2	8	—	—	—	—	—	—	—	—	—
	212	tampitre	8	2	7	—	—	—	—	—	—	—	—	—
	213	orlipe	6	2	6	—	—	—	—	—	—	—	—	—
	214	cacri	5	2	5	—	—	—	—	—	—	—	—	—
	215	opochouc	8	3	6	—	—	—	—	—	—	—	—	—
	216	éléchaut	8	3	6	—	—	—	—	—	—	—	—	—
	217	savreuil	8	2	6	—	—	—	—	—	—	—	—	—
	218	tobofar	7	3	7	—	—	—	—	—	—	—	—	—
	219	chimbaga	8	3	6	—	—	—	—	—	—	—	—	—
	220	pecodile	8	3	8	—	—	—	—	—	—	—	—	—
	221	macapanzé	9	4	8	—	—	—	—	—	—	—	—	—
	222	safazerne	9	3	9	—	—	—	—	—	—	—	—	—
	223	pintare	7	2	6	—	—	—	—	—	—	—	—	—
	224	tartus	6	2	6	—	—	—	—	—	—	—	—	—
	225	séland	6	2	5	—	—	—	—	—	—	—	—	—
	226	bousame	7	2	6	—	—	—	—	—	—	—	—	—
	227	nénudrix	8	3	8	—	—	—	—	—	—	—	—	—
	228	crodoine	8	2	7	—	—	—	—	—	—	—	—	—
	229	salarruque	10	3	8	—	—	—	—	—	—	—	—	—
	230	goéssum	7	3	5	—	—	—	—	—	—	—	—	—
231	tugouste	8	2	7	—	—	—	—	—	—	—	—	—	
232	caoutteille	11	3	7	—	—	—	—	—	—	—	—	—	
233	perlami	7	3	7	—	—	—	—	—	—	—	—	—	
234	rutaggan	8	3	6	—	—	—	—	—	—	—	—	—	
235	lumandre	8	2	7	—	—	—	—	—	—	—	—	—	
236	artibour	8	3	7	—	—	—	—	—	—	—	—	—	
237	chegouin	8	2	5	—	—	—	—	—	—	—	—	—	
238	macétor	7	3	7	—	—	—	—	—	—	—	—	—	
239	casphant	8	2	6	—	—	—	—	—	—	—	—	—	

	Stimulus	N ^{bre} lettres	N ^{bre} syll. orales	N ^{bre} graph.	Indices de fréquence			Taille de la famille morphologique			Cycle d'apparition dans les manuels scolaires		
					Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine	Stimulus entier	1 ^{re} racine	2 ^e racine
240	puroni	6	3	6	—	—	—	—	—	—	—	—	—
Moyenne ¹		7,73	2,57	6,60	—	—	—	—	—	—	—	—	—
Écart type ¹		1,28	0,57	1,04	—	—	—	—	—	—	—	—	—

Note. N^{bre} lettres = Nombre de lettres; N^{bre} syll. orales = Nombre de syllabes orales; N^{bre} graph. = Nombre de graphèmes; — = non disponible; Cycle d'apparition dans les manuels scolaires : 1 = 1^{er} cycle; 2 = 2^e cycle; 3 = 3^e cycle.

¹ La moyenne et l'écart type sont calculés uniquement sur les stimulus de l'expérimentation.

ANNEXE E

PROTOCOLE DE PASSATION DE LA TÂCHE DE DÉCISION LEXICALE

E. 1 Étape 1 : Création d'artéfacts

Dire au participant : « Maintenant que tu es installé, je vais aller partir l'enregistrement de l'autre côté. Puis, je vais te demander de faire différents mouvements. Ça nous servira plus tard à bien analyser les données. »

Dans cet ordre, faire faire les mouvements suivants au participant. Dire :

- « En regardant droit devant, vers l'écran, sans bouger la tête...
 - o ferme les yeux [attendre 3 secondes];
 - o ouvre les yeux;
 - o cligne des yeux normalement quelques fois;
 - o cligne rapidement des yeux (battement rapide des paupières);
 - o cligne lentement des yeux, comme si tu t'endormais.
- Toujours sans bouger la tête,
 - o regarde à gauche [pointer à la gauche du participant];
 - o regarde à droite [pointer à la droite du participant];
 - o encore une fois, regarde à gauche [pointer à la gauche du participant];
 - o regarde à droite [pointer à la droite du participant].
- Toujours sans bouger la tête,
 - o regarde vers le haut [pointer vers le haut];
 - o regarde vers le bas [pointer vers le bas];
 - o encore une fois, regarde vers le haut [pointer vers le haut];
 - o regarde vers le bas [pointer vers le bas].
- Serre très fort la mâchoire [attendre 3 secondes].
- Prend une grande inspiration, puis expire longtemps. »

E. 2 Étape 2 : Tâche de décision lexicale

Dire au participant : « Maintenant, tu vas faire une activité de lecture de vrais mots et de mots inventés. Il est possible que les mots inventés soient formés de deux mots existants en français ou qu'ils contiennent des parties de mots que tu connais. Je serai juste de l'autre côté du miroir et tu resteras seul ici pour pouvoir bien te concentrer. De l'autre côté, je peux te voir et t'entendre. Je peux aussi te parler. Tu vas devoir lire dans ta tête, tu ne dois pas lire à haute voix. Tu vas appuyer sur la touche avec le collant vert si les lettres présentées en noir forment un vrai mot. Si les lettres forment un mot inventé, tu

appuies sur la touche avec le collant rouge. Tu peux placer tes deux index, un doigt sur chaque bouton. Tu dois appuyer le plus rapidement possible. Tu vas t'asseoir confortablement pour garder la tête bien droite et essayer de ne pas bouger. Les lettres apparaîtront au centre de l'écran. Avant chaque suite de lettres, tu vas voir une croix au centre de l'écran. Quand tu la vois, essaie de ne pas cligner des yeux. Les mots s'affichent rapidement tout de suite après et il faudrait s'assurer que tu les as bien vus. Essaie de retenir ton clignement jusqu'à ce que tu aies appuyé sur le bouton. Puis, tu pourras cligner des yeux. On va faire une pratique ensemble. »

Partir le fichier sur DMDX et indiquer le numéro du participant. Lire la consigne affichée à l'écran :

« Bienvenue!

Appuie sur la touche avec le cercle **vert**, si les lettres forment un vrai mot.

Appuie sur la touche avec le cercle **rouge**, si les lettres forment un mot inventé.

Les premiers essais sont une pratique. Prêt?

Appuie sur la barre d'espace pour commencer. »

Rester près du participant pendant qu'il réalise les items de pratique. S'assurer qu'il comprend bien la tâche, qu'il ne bouge pas et qu'il cligne des yeux au bon moment. Fournir une rétroaction au participant. Au besoin, si le participant n'avait pas bien compris la tâche, appuyer sur Échap et repartir le fichier sur DMDX (en indiquant à nouveau le numéro de participant) et refaire les items de pratique avec le participant.

Dire au participant : « Maintenant que la pratique est terminée, on va commencer pour vrai. Il y a un grand nombre de mots à faire, et les mots s'affichent en continu quoi qu'il arrive. Donc pendant toute la tâche tu devras rester concentré et bouger le moins possible. Pour te permettre de relaxer un peu, on a mis plusieurs écrans où c'est écrit "Pause". Quand tu arrives sur cet écran, tu peux en profiter pour respirer un peu, cligner des yeux comme tu veux (mais ne pas trop bouger). Après 30 secondes, un message apparaîtra pour t'indiquer d'appuyer sur la barre d'espace pour continuer. Tu peux prendre plus de temps de pause si tu le désires et recommencer quand tu es prêt. Il y en a environ 200 mots et non-mots qui seront affichés, et nous avons mis 3 pauses en tout. Quand tu auras tout terminé, je viendrai te voir. Je vais aller dans l'autre salle, derrière le miroir. Attends que je referme la porte derrière moi pour commencer. »

Retourner dans la salle d'observation, derrière le miroir sans tain.

E.3 Script DMDX présenté aux participants

Étapes du déroulement	Affichage à l'écran
1. Message d'accueil	<p style="text-align: center;">« Bienvenue!</p> <p>Appuie sur la touche avec le cercle vert, si les lettres forment un vrai mot. Appuie sur la touche avec le cercle rouge, si les lettres forment un mot inventé.</p> <p style="text-align: center;">Les premiers essais sont une pratique. Prêt? Appuie sur la barre d'espace pour commencer.</p>
2. Bloc 0	16 items de pratique
3. Message de transition	<p style="text-align: center;">« Tu as bien compris! Nous pourrons bientôt commencer.</p> <p>Appuie sur la touche avec le cercle vert, si les lettres forment un vrai mot. Appuie sur la touche avec le cercle rouge, si les lettres forment un mot inventé.</p> <p style="text-align: center;">Lorsque tu es prêt, appuie sur la barre d'espace pour commencer. »</p>
4. Bloc 1	60 items aléatoires
5. Pause 1	<p style="text-align: center;">« Pause 1 » [après 30 secondes]</p> <p style="text-align: center;">« Lorsque tu es prêt, appuie sur la barre d'espace pour commencer. »</p>
6. Bloc 2	60 items aléatoires
7. Pause 2	<p style="text-align: center;">« Pause 2 » [après 30 secondes]</p> <p style="text-align: center;">« Lorsque tu es prêt, appuie sur la barre d'espace pour commencer. »</p>
8. Bloc 3	60 items aléatoires
9. Pause 3	<p style="text-align: center;">« Pause 3 » [après 30 secondes]</p> <p style="text-align: center;">« Lorsque tu es prêt, appuie sur la barre d'espace pour commencer. »</p>
10. Bloc 4	60 items aléatoires
11. Fin de la tâche	<p style="text-align: center;">« Tu as terminé. Merci!</p> <p style="text-align: center;">Ne bouge pas, quelqu'un viendra te voir. »</p>

ANNEXE F
ITEMS DU TEST DYNAMIQUE

Tableau E.1 Mots du test dynamique

	#	Phrase	Sens des racines
MOTS			
	A	La pause est à <u>midi</u> .	<i>mi</i> signifie « milieu » <i>di</i> signifie « journée »
Racines latines	1	La réunion aura lieu au <u>trimestre</u> prochain.	<i>tri</i> signifie « trois » <i>mestre</i> signifie « mois »
	2	Cette école contient un <u>auditorium</u> .	<i>aud</i> signifie « entendre » <i>rium</i> signifie « lieu »
	3	Cet homme est <u>funambule</u> .	<i>fun</i> signifie « corde » <i>ambule</i> signifie « se promener »
	4	Dans cet enclos, on retrouve un animal <u>ovipare</u> .	<i>ovi</i> signifie « oeuf » <i>pare</i> signifie « mettre au monde »
	5	Le 20 mars avait lieu l' <u>équinoxe</u> du printemps.	<i>équi</i> signifie « égal » <i>noxe</i> signifie « nuit »
Racines Grecques	6	Cet objet est un <u>microscope</u> .	<i>micro</i> signifie « petit » <i>scope</i> signifie « observer »
	7	Mon frère est un brillant <u>archéologue</u> .	<i>archéo</i> signifie « ancien » <i>logue</i> signifie « étude »
	8	Mon chien est <u>hydrophobe</u> depuis toujours.	<i>hydro</i> signifie « eau » <i>phobe</i> signifie « peur »
	9	Ce fruit a un <u>épiderme</u> lisse.	<i>épi</i> signifie « à la surface de, au-dessus » <i>derme</i> signifie « peau »
	10	L'enseignante dessine un <u>tétraèdre</u> au tableau.	<i>tétra</i> signifie « quatre » <i>èdre</i> signifie « face, base »

Note. Le sens des racines est tiré de l'ouvrage de Cellard (2013).

Tableau E.2 Pseudomots du test dynamique

	#	Phrase	Sens des racines
PSEUDOMOTS			
	B	Lorsque je serai grande, j'aimerais être <u>biblionaute</u> .	<i>biblio</i> signifie « livre » <i>naute</i> signifie « qui navigue »
Racines latines	1	Cet animal est <u>multivore</u> .	<i>multi</i> signifie « nombreux » <i>vore</i> signifie « manger, avaler »
	2	Cette femme impressionnante est <u>quintilingue</u> .	<i>quinti</i> signifie « cinq » <i>lingue</i> signifie « langue »
	3	Ce vêtement est <u>omnifuge</u> .	<i>omni</i> signifie « tout » <i>fuge</i> signifie « fuir »
	4	Je construis un énorme <u>apidome</u> .	<i>api</i> signifie « abeille » <i>dome</i> signifie « maison »
	5	L' <u>aguascrit</u> est une nouvelle technique.	<i>aqua</i> signifie « eau » <i>scrit</i> signifie « écrire »
Racines grecques	6	Cette forme semble <u>monogone</u> .	<i>mono</i> signifie « seul » <i>gone</i> signifie « angle »
	7	Ces deux personnes sont <u>hétérophages</u> .	<i>hétéro</i> signifie « différent » <i>phage</i> signifie « manger, grand mangeur »
	8	J'étudie pour devenir <u>chroniatre</u> .	<i>chron</i> signifie « temps » <i>iatre</i> signifie « médecin »
	9	Ce crayon <u>hexachrome</u> est très utile.	<i>hexa</i> signifie « six » <i>chrome</i> signifie « couleur »
	10	Cette plante <u>autotherme</u> pousse dans mon jardin.	<i>auto</i> signifie « soi-même » <i>therme</i> signifie « chaleur »

Note. Le sens des racines est tiré de l'ouvrage de Cellard (2013).

ANNEXE G

STATISTIQUES DESCRIPTIVES DE L'AMPLITUDE ET DE LA LATENCE DE 200 MS À 350 MS ET DE 350 MS À 500 MS

Les tableaux suivants (G.1 à G.4) présentent les statistiques descriptives de l'amplitude négative maximale et de la latence aux deux fenêtres temporelles ciblées, soit de 200 ms à 350 ms et de 350 ms à 500 ms. À la fenêtre temporelle de 200 ms à 350 ms poststimulus, la négativité moyenne est plus grande en région antérieure que postérieure, et ce pour toutes les catégories de stimulus. Les non-mots (p. ex. **pecodile*) et les mots polymorphémiques formés de deux racines libres (p. ex. *porteclé*) génèrent les plus grandes négativités, surtout dans la région antérieure médiane. De façon générale, les plus petites négativités, autant en région antérieure que postérieure, sont générées par les pseudomots (p. ex. **polyscope*) et les mots polymorphémiques formés de deux racines grecques (p. ex. *microscope*). Toutes catégories d'items confondues, la latence du marqueur de potentiel évoqué identifié à cette fenêtre temporelle survient entre 272,38 ms et 333,48 ms. On constate qu'en région antérieure, les latences sont généralement plus précoces à gauche qu'à droite. En région postérieure, le patron inverse tend à se dessiner : les latences sont plus tardives à gauche qu'à droite. Cela pourrait indiquer une topographie de ce marqueur de potentiel évoqué se déplaçant généralement de la région postérieure droite à antérieure gauche.

En ce qui concerne la fenêtre temporelle de 350 ms à 500 ms poststimulus, le même cas de figure se dessine en présentant des négativités moyennes plus grandes en région antérieure que postérieure. À nouveau, les non-mots (p. ex. **pecodile*) et les mots polymorphémiques formés de deux racines libres (p. ex. *porteclé*) génèrent les plus grandes négativités. De façon générale, les plus petites négativités sont générées par les mots polymorphémiques formés de deux racines grecques (p. ex. *microscope*) ainsi que les pseudomots polymorphémiques formés de deux racines libres (p. ex. **passeclé*). Dans cette fenêtre temporelle, la latence moyenne du marqueur de potentiel évoqué varie de 394,72 ms à 438,10 ms. Elle est habituellement plus tardive en région antérieure que postérieure, même si ce n'est pas tout à fait le cas pour certaines catégories d'items, notamment à gauche (p. ex. non-mots).

Tableau G.1 Amplitude négative maximale (μV) de 200 ms à 350 ms poststimulus à la tâche de décision lexicale pour chaque catégorie d'items.

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines liées latines			
Antérieure gauche	-7,50	4,27	[-9,12; -5,88]
Antérieure médiane	-8,22	4,15	[-9,80; -6,64]
Antérieure droite	-6,91	4,31	[-8,55; -5,27]
Postérieure gauche	-3,17	2,72	[-4,20; -2,14]
Postérieure médiane	-4,58	3,62	[-5,96; -3,21]
Postérieure droite	-3,71	3,78	[-5,15; -2,27]
(2) Mots polymorphémiques formés de deux racines liées grecques			
Antérieure gauche	-7,00	4,35	[-8,66; -5,35]
Antérieure médiane	-7,00	4,63	[-8,76; -5,24]
Antérieure droite	-6,06	3,89	[-7,54; -4,58]
Postérieure gauche	-3,31	2,99	[-4,45; -2,18]
Postérieure médiane	-4,08	4,17	[-5,66; -2,49]
Postérieure droite	-3,51	3,97	[-5,02; -2,00]
(3) Mots polymorphémiques formés de deux racines libres			
Antérieure gauche	-8,31	3,96	[-9,81; -6,80]
Antérieure médiane	-8,61	4,65	[-10,38; -6,84]
Antérieure droite	-7,24	3,41	[-8,53; -5,94]
Postérieure gauche	-4,56	2,60	[-5,55; -3,57]
Postérieure médiane	-7,09	5,95	[-9,36; -4,83]
Postérieure droite	-5,50	4,58	[-7,24; -3,75]
(4) Mots monomorphémiques			
Antérieure gauche	-7,74	3,76	[-9,17; -6,31]
Antérieure médiane	-8,04	4,13	[-9,61; -6,47]
Antérieure droite	-6,17	2,83	[-7,25; -5,10]
Postérieure gauche	-3,87	2,76	[-4,92; -2,82]
Postérieure médiane	-5,84	3,83	[-7,30; -4,39]
Postérieure droite	-3,82	3,49	[-5,15; -2,49]
(5) Pseudomots polymorphémiques formés de deux racines latines			
Antérieure gauche	-7,30	3,71	[-8,71; -5,89]
Antérieure médiane	-7,73	4,10	[-9,29; -6,18]
Antérieure droite	-6,46	2,91	[-7,57; -5,35]
Postérieure gauche	-3,84	2,54	[-4,81; -2,88]
Postérieure médiane	-4,86	4,21	[-6,46; -3,26]
Postérieure droite	-3,99	3,51	[-5,33; -2,66]
(6) Pseudomots polymorphémiques formés de deux racines grecques			
Antérieure gauche	-7,17	3,98	[-8,68; -5,65]
Antérieure médiane	-6,77	4,25	[-8,39; -5,16]
Antérieure droite	-5,73	3,46	[-7,04; -4,41]
Postérieure gauche	-3,50	2,91	[-4,60; -2,39]
Postérieure médiane	-3,77	3,76	[-5,20; -2,34]
Postérieure droite	-3,27	3,46	[-4,58; -1,95]

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(7) Pseudomots polymorphémiques formés de deux racines libres			
Antérieure gauche	-7,41	4,11	[-8,97; -5,84]
Antérieure médiane	-7,66	4,50	[-9,37; -5,94]
Antérieure droite	-6,32	4,03	[-7,86; -4,79]
Postérieure gauche	-3,15	2,21	[-3,99; -2,31]
Postérieure médiane	-4,44	3,97	[-5,96; -2,93]
Postérieure droite	-3,20	3,09	[-4,38; -2,03]
(8) Non-mots			
Antérieure gauche	-9,21	4,22	[-10,81; -7,60]
Antérieure médiane	-9,30	4,72	[-11,09; -7,50]
Antérieure droite	-7,32	3,84	[-8,78; -5,86]
Postérieure gauche	-4,47	3,29	[-5,72; -3,22]
Postérieure médiane	-4,48	4,38	[-6,15; -2,81]
Postérieure droite	-3,50	3,93	[-5,00; -2,01]

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau G.2 Latence du pic d'amplitude maximale (ms) de 200 ms à 350 ms poststimulus à la tâche de décision lexicale pour chaque catégorie d'items

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines liées latines			
Antérieure gauche	317,17	23,25	[308,33; 326,02]
Antérieure médiane	316,86	28,39	[306,06; 327,66]
Antérieure droite	319,28	28,88	[308,29; 330,26]
Postérieure gauche	306,52	47,16	[288,58; 324,46]
Postérieure médiane	286,38	59,12	[263,89; 308,87]
Postérieure droite	292,72	60,08	[269,87; 315,58]
(2) Mots polymorphémiques formés de deux racines liées grecques			
Antérieure gauche	309,00	26,45	[298,94; 319,06]
Antérieure médiane	309,79	28,09	[299,11; 320,48]
Antérieure droite	319,72	23,08	[310,95; 328,50]
Postérieure gauche	307,66	48,95	[289,04; 326,27]
Postérieure médiane	296,55	55,10	[275,59; 317,51]
Postérieure droite	286,10	62,56	[262,31; 309,90]
(3) Mots polymorphémiques formés de deux racines libres			
Antérieure gauche	321,45	26,76	[311,27; 331,63]
Antérieure médiane	312,66	31,83	[300,55; 324,76]
Antérieure droite	321,03	24,68	[311,65; 330,42]
Postérieure gauche	322,83	37,06	[308,73; 336,92]
Postérieure médiane	284,97	60,69	[261,88; 308,05]
Postérieure droite	272,38	58,97	[249,95; 294,81]
(4) Mots monomorphémiques			
Antérieure gauche	313,90	24,14	[304,72; 323,08]
Antérieure médiane	319,62	19,73	[312,11; 327,13]
Antérieure droite	317,28	20,42	[309,51; 325,04]
Postérieure gauche	311,55	47,30	[293,56; 329,54]
Postérieure médiane	304,38	53,81	[283,91; 324,85]
Postérieure droite	284,34	65,73	[259,34; 309,35]
(5) Pseudomots polymorphémiques formés de deux racines latines			
Antérieure gauche	314,90	21,96	[306,55; 323,25]
Antérieure médiane	318,21	23,90	[309,12; 327,30]
Antérieure droite	323,28	19,27	[315,95; 330,61]
Postérieure gauche	311,34	44,54	[294,40; 328,29]
Postérieure médiane	291,90	54,45	[271,19; 312,61]
Postérieure droite	269,24	61,43	[245,87; 292,61]
(6) Pseudomots polymorphémiques formés de deux racines grecques			
Antérieure gauche	311,72	25,42	[302,06; 321,39]
Antérieure médiane	315,28	25,49	[305,58; 324,97]
Antérieure droite	316,31	23,66	[307,31; 325,31]
Postérieure gauche	309,69	46,61	[291,96; 327,42]
Postérieure médiane	292,48	55,57	[271,35; 313,62]
Postérieure droite	284,41	59,56	[261,76; 307,07]

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(7) Pseudomots polymorphémiques formés de deux racines libres			
Antérieure gauche	314,72	21,76	[306,45; 323,00]
Antérieure médiane	310,28	21,82	[301,98; 318,57]
Antérieure droite	312,34	20,44	[304,57; 320,12]
Postérieure gauche	333,48	19,97	[325,89; 341,08]
Postérieure médiane	288,07	55,38	[267,00; 309,13]
Postérieure droite	276,10	63,04	[252,12; 300,08]
(8) Non-mots			
Antérieure gauche	321,66	19,49	[314,24; 329,07]
Antérieure médiane	320,72	23,61	[311,74; 329,71]
Antérieure droite	317,66	28,34	[306,87; 328,44]
Postérieure gauche	319,62	38,35	[305,03; 334,21]
Postérieure médiane	301,14	53,46	[280,80; 321,47]
Postérieure droite	306,48	55,94	[285,20; 327,76]

Note. IC 95 % = intervalle de confiance à 95 %.

Tableau G.3 Amplitude négative maximale (μV) de 350 ms à 500 ms poststimulus à la tâche de décision lexicale pour chaque catégorie d'items

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines liées latines			
Antérieure gauche	-8,36	3,90	[-9,85; -6,88]
Antérieure médiane	-8,71	3,50	[-10,04; -7,38]
Antérieure droite	-7,97	3,39	[-9,26; -6,68]
Postérieure gauche	-4,75	1,98	[-5,05; -4,00]
Postérieure médiane	-5,88	2,83	[-6,95; -4,80]
Postérieure droite	-4,46	2,79	[-5,53; -3,40]
(2) Mots polymorphémiques formés de deux racines liées grecques			
Antérieure gauche	-7,39	3,75	[-8,82; -5,97]
Antérieure médiane	-7,82	3,04	[-8,97; -6,67]
Antérieure droite	-6,52	3,02	[-7,67; -5,37]
Postérieure gauche	-4,57	2,23	[-5,42; -3,72]
Postérieure médiane	-5,78	3,30	[-7,04; -4,53]
Postérieure droite	-4,01	3,64	[-5,39; -2,62]
(3) Mots polymorphémiques formés de deux racines libres			
Antérieure gauche	-9,36	4,28	[-10,98; -7,73]
Antérieure médiane	-9,32	4,27	[-10,94; -7,69]
Antérieure droite	-8,51	4,05	[-10,05; -6,97]
Postérieure gauche	-5,71	3,67	[-7,11; -4,32]
Postérieure médiane	-7,50	5,80	[-9,71; -5,29]
Postérieure droite	-5,64	3,75	[-7,06; -4,21]
(4) Mots monomorphémiques			
Antérieure gauche	-8,67	3,70	[-10,07; -7,26]
Antérieure médiane	-8,62	3,27	[-9,86; -7,37]
Antérieure droite	-7,64	2,75	[-8,69; -6,60]
Postérieure gauche	-5,62	2,89	[-6,72; -4,52]
Postérieure médiane	-6,86	4,00	[-8,38; -5,34]
Postérieure droite	-5,35	3,18	[-6,56; -4,14]
(5) Pseudomots polymorphémiques formés de deux racines latines			
Antérieure gauche	-8,60	3,27	[-9,85; -7,36]
Antérieure médiane	-9,27	3,32	[-10,53; -8,01]
Antérieure droite	-8,39	3,04	[-9,54; -7,23]
Postérieure gauche	-4,70	2,51	[-5,66; -3,75]
Postérieure médiane	-5,50	2,72	[-6,53; -4,47]
Postérieure droite	-4,49	3,02	[-5,64; -3,34]
(6) Pseudomots polymorphémiques formés de deux racines grecques			
Antérieure gauche	-7,83	3,47	[-9,15; -6,51]
Antérieure médiane	-8,48	3,00	[-9,62; -7,34]
Antérieure droite	-7,97	2,93	[-9,08; -6,85]
Postérieure gauche	-4,48	3,25	[-5,71; -3,24]
Postérieure médiane	-5,36	3,40	[-6,65; -4,07]
Postérieure droite	-4,35	2,51	[-5,30; -3,39]

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(7) Pseudomots polymorphémiques formés de deux racines libres			
Antérieure gauche	-7,70	3,76	[-9,12; -6,27]
Antérieure médiane	-8,07	4,26	[-9,69; -6,45]
Antérieure droite	-6,73	3,67	[-8,12; -5,33]
Postérieure gauche	-4,10	2,13	[-4,91; -3,28]
Postérieure médiane	-5,58	2,76	[-6,63; -4,53]
Postérieure droite	-4,31	3,45	[-5,62; -3,00]
(8) Non-mots			
Antérieure gauche	-9,71	3,59	[-11,07; -8,34]
Antérieure médiane	-10,14	4,34	[-11,79; -8,49]
Antérieure droite	-8,48	3,31	[-9,74; -7,22]
Postérieure gauche	-5,49	3,00	[-6,64; -4,35]
Postérieure médiane	-6,19	3,77	[-7,62; -4,75]
Postérieure droite	-4,44	2,83	[-5,52; -3,37]

Note. A = antérieure; P = postérieure; IC 95 % = intervalle de confiance à 95 %.

Tableau G.4 Latence du pic d'amplitude maximale (ms) de 350 ms à 500 ms poststimulus à la tâche de décision lexicale pour chaque catégorie d'items

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(1) Mots polymorphémiques formés de deux racines liées latines			
Antérieure gauche	414,69	51,42	[395,13; 434,25]
Antérieure médiane	411,00	46,29	[393,39; 428,61]
Antérieure droite	411,93	44,73	[394,92; 428,94]
Postérieure gauche	413,52	52,13	[393,69; 433,35]
Postérieure médiane	401,21	38,40	[386,60; 415,81]
Postérieure droite	407,14	40,60	[391,69; 422,58]
(2) Mots polymorphémiques formés de deux racines liées grecques			
Antérieure gauche	419,10	46,15	[401,55; 436,66]
Antérieure médiane	423,03	45,09	[405,88; 440,19]
Antérieure droite	413,93	47,802	[395,75; 432,11]
Postérieure gauche	413,79	48,45	[395,36; 432,22]
Postérieure médiane	405,17	41,29	[389,47; 420,88]
Postérieure droite	415,07	47,27	[397,09; 433,05]
(3) Mots polymorphémiques formés de deux racines libres			
Antérieure gauche	413,55	53,44	[393,23; 433,88]
Antérieure médiane	418,41	48,28	[400,05; 436,78]
Antérieure droite	423,41	51,78	[403,72; 443,11]
Postérieure gauche	406,76	46,81	[388,95; 424,56]
Postérieure médiane	403,72	47,33	[385,72; 421,73]
Postérieure droite	411,97	55,73	[390,77; 433,16]
(4) Mots monomorphémiques			
Antérieure gauche	415,24	45,35	[397,99; 432,49]
Antérieure médiane	406,34	44,94	[389,25; 423,44]
Antérieure droite	427,62	44,84	[410,56; 444,68]
Postérieure gauche	403,48	44,78	[386,45; 420,52]
Postérieure médiane	394,72	32,88	[382,22; 407,23]
Postérieure droite	416,14	40,62	[400,69; 431,59]
(5) Pseudomots polymorphémiques formés de deux racines latines			
Antérieure gauche	419,00	42,96	[402,66; 435,34]
Antérieure médiane	429,03	40,88	[413,48; 444,58]
Antérieure droite	421,24	37,28	[407,06; 435,42]
Postérieure gauche	414,21	50,58	[394,97; 433,45]
Postérieure médiane	415,69	39,19	[400,78; 430,60]
Postérieure droite	425,45	45,16	[408,27; 442,63]
(6) Pseudomots polymorphémiques formés de deux racines grecques			
Antérieure gauche	418,41	42,12	[402,39; 434,43]
Antérieure médiane	420,28	40,38	[404,92; 435,64]
Antérieure droite	422,21	39,96	[407,01; 437,41]
Postérieure gauche	421,83	48,92	[403,22; 440,44]
Postérieure médiane	407,86	38,67	[393,15; 422,57]
Postérieure droite	405,21	38,95	[390,39; 420,02]

Catégories d'items	<i>M</i>	<i>ÉT</i>	IC 95 %
(7) Pseudomots polymorphémiques formés de deux racines libres			
Antérieure gauche	418,76	47,31	[400,76; 436,76]
Antérieure médiane	438,10	45,16	[420,93; 455,28]
Antérieure droite	433,86	44,42	[416,97; 450,76]
Postérieure gauche	413,38	45,98	[395,89; 430,87]
Postérieure médiane	405,79	39,95	[390,60; 420,99]
Postérieure droite	413,03	43,58	[396,46; 429,61]
(8) Non-mots			
Antérieure gauche	418,79	41,69	[402,90; 434,61]
Antérieure médiane	425,62	39,09	[410,75; 440,49]
Antérieure droite	424,59	42,56	[408,40; 440,78]
Postérieure gauche	425,24	52,35	[405,33; 445,15]
Postérieure médiane	405,34	41,48	[389,57; 421,12]
Postérieure droite	406,21	44,81	[389,16; 423,25]

Note. IC 95 % = intervalle de confiance à 95 %.

APPENDICE A

MODÈLES D'IDENTIFICATION DES MOTS ÉCRITS QUI N'ABORDENT PAS EXPLICITEMENT LA MORPHOLOGIE

Différentes visions de l'identification des mots écrits permettent un classement des modèles. D'abord, les modèles à double voie qui s'appuient sur le traitement des mots écrits selon deux procédures distinctes sont présentés. De ces modèles, découlent les modèles à étapes qui tentent d'expliquer l'identification des mots écrits de façon plus séquentielle que dichotomique en considérant des processus intermédiaires mis en œuvre par les lecteurs débutants. Ensuite, les modèles par analogie qui avancent que les lecteurs s'appuient sur les mots déjà connus pour identifier d'autres mots sont présentés. Enfin, des modèles connexionnistes sont exposés afin de mettre en relief leur hypothèse universelle de l'identification des mots écrits.

Le modèle à double voie de Coltheart (1978) prévoit deux types de procédures s'activant pour identifier les mots écrits, selon différents critères, dont la fréquence du mot. La procédure la plus rapide, utilisée notamment pour les mots fréquents, est la procédure directe, dite d'adressage, qui permet de récupérer la forme phonologique du mot et son sens directement dans le lexique interne mental à partir de sa représentation orthographique. La seconde procédure est plus lente et permet la lecture de mots plus rares et de non-mots. Il s'agit de la procédure d'assemblage. Comme son nom laisse sous-entendre, la suite orthographique présentée visuellement au lecteur est scindée en graphèmes. Chacun de ces graphèmes est ensuite associé au phonème correspondant. Puis, les phonèmes sont assemblés pour former et prononcer le mot. Ce modèle a été élaboré pour expliquer le fonctionnement de l'identification des mots chez le lecteur expert. Ensuite, des chercheurs s'en sont inspirés pour proposer des modèles développementaux permettant de rendre compte du développement de l'identification des mots chez le lecteur débutant.

Les modèles développementaux, ou à étapes, prévoient que les lecteurs passent par trois principales étapes avant d'atteindre la compétence de lecteur expert : l'étape logographique, l'étape alphabétique et l'étape orthographique (Frith, 1985). Pour certains modèles (Harris et Coltheart, 1986; Marsh *et al.*, 1981), ces étapes sont au nombre de quatre simplement dû à une scission différente des principales procédures. La première étape des modèles développementaux est l'étape logographique. Cette stratégie s'installe avant l'apprentissage formel de la lecture. Le lecteur utilise divers indices et traits

visuels pour reconnaître un mot, par exemple sa forme ou la couleur du logo associé (Morton, 1989; Sprenger-Charolles et Casalis, 1996). L'enfant peut aussi tenter de reconnaître un mot par la présence d'une ou plusieurs lettres qu'il connaît, sans considérer la présence d'autres lettres ou la configuration de celles-ci (Sprenger-Charolles et Casalis, 1996). Une jeune enfant peut donc penser que le mot *Ketchup* est son prénom puisqu'il commence par la même lettre que son prénom, *Kathleen*. Enfin, l'étape logographique comprend aussi les mots appris globalement et mémorisés ainsi par les lecteurs (Frith, 1985; Sprenger-Charolles et Casalis, 1996). La deuxième étape est celle alphabétique, bien qu'elle porte différentes appellations selon les modèles (p. ex. *phase alphabétique*, Morton, 1989; *décodage phonologique*, Harris et Coltheart, 1986). Il s'agit de la procédure qui s'assure de faire correspondre chaque graphème à un phonème afin de lire des mots connus ou inconnus ainsi que des pseudomots (Sprenger-Charolles et Casalis, 1996). Pour le lecteur débutant qui ne maîtrise pas encore parfaitement les correspondances graphème-phonème, cette procédure est plus laborieuse et demande beaucoup de ressources cognitives. La troisième et dernière étape est celle associée aux habiletés du lecteur expert, l'étape orthographique. Cette étape correspond à la voie d'adressage de Coltheart (1978) où les mots sont reconnus globalement, sans passer par un assemblage phonologique. Cette étape permet d'ailleurs d'identifier des mots qui ne sont pas phonologiquement transparents, par exemple les mots qui contiennent des lettres muettes (p. ex. *bavard*) ou des graphèmes particuliers (p. ex. *monsieur*). Selon la revue de littérature de Sprenger-Charolles et Casalis (1996), les unités traitées par l'étape orthographique correspondent aux morphèmes puisque l'identification des mots se fait ici par unité de sens. Or, aucune précision n'est apportée quant à l'identification des mots écrits polymorphémiques dans ce modèle. Les modèles à étapes ont reçu diverses critiques notamment en ce qui concerne leur rigidité. Effectivement, selon ces modèles, les étapes sont successives et doivent absolument être franchies pour passer à la suivante, et ce, de la même façon pour tous les lecteurs.

Les modèles par analogie sont un autre type de modèles qui expliquent l'identification des mots écrits. Ils critiquent surtout l'aspect non lexical des modèles à double voie (Sprenger-Charolles et Casalis, 1996), c'est-à-dire le fait que dans la voie indirecte, les mots déjà connus et présents dans le lexique mental du lecteur n'influencent pas la lecture de nouveaux mots. Selon les modèles analogiques, le rôle de la syllabe, particulièrement de la rime, est central dans l'apprentissage de la lecture (Sprenger-Charolles et Casalis, 1996).

Le modèle par analogie de Goswami et Bryant (1990) prévoit trois précurseurs à l'identification des mots écrits pour le lecteur débutant. D'abord, la capacité à segmenter les mots en syllabes et les syllabes en unités sous-syllabiques (attaque et rime) à l'oral permet au prélecteur d'identifier des patrons orthographiques qui correspondent aux rimes. Ce modèle postule donc que le lexique orthographique se bâtit sur la base des rimes. La lecture de nouveaux mots se fait par analogie aux mots connus partageant une même rime. Ensuite, avec l'enseignement formel de la lecture, les connaissances phonologiques au niveau phonémique du lecteur se développent, d'abord en écriture puis en lecture. C'est ce qui explique pourquoi les auteurs soutiennent l'influence mutuelle de la lecture et de l'écriture dans le développement de la littératie.

Des limites ont été soulevées quant au modèle de Goswami et Bryant (1990). Effectivement, des résultats expérimentaux contradictoires ont d'abord mené les chercheurs à émettre l'hypothèse que le développement des processus d'identification des mots dépend du système d'écriture de la langue de l'apprenant (Sprenger-Charolles et Casalis, 1996). Enfin, sans remettre totalement en question l'apport des unités sous-syllabiques dans l'apprentissage de la lecture, des recherches ont permis de démontrer que, à long terme, l'entraînement à la lecture serait plus efficace par l'enseignement des correspondances graphème-phonème que par l'analogie (Bruck et Treiman, 1992; Wise *et al.*, 1990). Il ne serait donc pas optimal d'utiliser l'analogie comme unique procédure d'identification des mots et des pseudomots.

Enfin, les modèles connexionnistes forment le dernier type de modèles qui expliquent l'identification des mots écrits. Il s'agit de modèles computationnels qui simulent sur ordinateur les processus cognitifs mis en jeu lors de l'identification des mots écrits et les réponses comportementales associées. Ils se basent sur des régularités statistiques au sein des stimulus pour prévoir l'apprentissage. En opposition aux modèles à double voie qui prévoient un traitement différencié selon la nature du mot, les modèles connexionnistes postulent plutôt que tous les mots sont traités et identifiés selon un même système de nœuds interconnectés (Sprenger-Charolles et Casalis, 1996). Pour les connexionnistes, identifier un mot n'est pas synonyme de le récupérer en mémoire, mais bien d'activer ses différentes représentations. L'apprentissage de l'écrit consiste à optimiser le poids des interconnexions entre les nœuds.

Le modèle à activation interactive [ma traduction de *Interactive Activation Model*] de McClelland et Rumelhart (1981) a été pensé pour expliquer le traitement de mots de quatre lettres. Il prévoit trois

niveaux hiérarchiques d'activation : les traits visuels, les lettres et les mots. Des connexions excitatrices et inhibitrices sont présentes entre les niveaux et au sein de chaque niveau. Le modèle prévoit un traitement parallèle selon la modalité orale ou écrite lors de l'identification des mots. Cependant, seul le traitement des mots écrits est décrit ici. Ainsi, le trait visuel « | » active toutes les lettres qui partagent ce trait (p. ex. T, I, L) en inhibant toutes les autres (p. ex. Z, X, O). L'activation en parallèle de différents traits visuels pour une même lettre, puis pour une combinaison de lettres associées à leur position dans le mot, permet ensuite d'activer le mot souhaité et d'inhiber les mots qui ne correspondent pas aux propriétés des niveaux précédents. Le mot recueillant le plus grand nombre d'activations est correctement identifié. En plus de ces activations ascendantes, le modèle reconnaît un niveau de traitement dit *supérieur* qui permet de fournir des activations descendantes directement aux mots. Selon ce modèle, les mots fréquents sont reconnus plus rapidement que les mots de basse fréquence puisqu'ils disposent d'un niveau d'activation plus grand, avant même l'activation de traits visuels (Ferrand et Ayora, 2009). Cet apport accélère le processus en attribuant plus rapidement une forte activation pour un item fréquent du dernier niveau du modèle, le mot.

En reprenant les principes du modèle à activation interactive de McClelland et Rumelhart (1981), Grainger et Holcomb (2009) présentent le modèle bimodal à activation interactive [ma traduction de *bimodal interactive-activation model (BIAM)*]. Il s'agit d'une architecture hiérarchique où sont traités successivement les traits visuels ou phonologiques, les unités sublexicales et les mots afin d'accéder aux unités sémantiques, et ce, autant en modalité visuelle qu'auditive. Les auteurs tentent de faire correspondre les étapes du modèle aux principales composantes de potentiels évoqués d'identification des mots écrits. Donc, si on se concentre exclusivement sur la modalité visuelle, le N/P150 reflète le traitement des traits visuels du mot. La composante N250 reflète le traitement d'unités sublexicales, comme des lettres ou des groupes de lettres, permettant d'identifier le mot entier. Selon les auteurs, la composante P325 sert à renseigner sur la reconnaissance du mot entier. Enfin, le N400 témoigne de l'effort cognitif pour arrimer les représentations orthographiques aux représentations sémantiques.

Dans le même ordre d'idées, le modèle d'activation-vérification (Paap *et al.*, 1982) ressemble en plusieurs points au modèle à activation interactive (McClelland et Rumelhart, 1981). Effectivement, le modèle prévoit trois opérations : l'encodage, la vérification et la prise de décision. L'encodage renvoie aux opérations de bas niveau qui permettent de reconnaître les caractéristiques uniques aux lettres ou aux mots afin de les activer inconsciemment en mémoire. Une fois ces caractéristiques activées, la

procédure de vérification s'enclenche lors de laquelle le système détermine si le stimulus correspond à un mot stocké dans le lexique mental. Selon les auteurs (Paap *et al.*, 1982), il convient de concevoir cette étape comme une analyse indépendante de type descendant puisque l'analyse du stimulus se fait en fonction des items déjà présents dans le lexique mental. Cette étape est aussi influencée par des facteurs sémantiques et par la fréquence des mots. Selon Paap et ses collaborateurs (1982), il s'agit d'ailleurs de la principale différence avec le modèle à activation interactive (McClelland et Rumelhart, 1981). Enfin, la vérification mène à la prise de décision, à savoir si le stimulus correspond à un mot. Dans ce modèle, tous les mots sont stockés dans le lexique mental, mais aussi chaque lettre de façon individuelle à la manière d'un abécédaire.

Seidenberg et McClelland (1989a) s'inspirent du modèle à activation interactive (McClelland et Rumelhart, 1981) pour proposer le modèle à traitement parallèle distribué (PDP) composé d'une architecture visant à reproduire l'apprentissage de la lecture de mots. Ce modèle prévoit trois types d'unités : orthographiques, phonologiques et sémantiques. Lors de la lecture des mots, le modèle traite les suites orthographiques de trois lettres comme unités d'entrées et leur fait correspondre, à la sortie, des suites phonétiques de trois phonèmes en passant par des unités cachées. Ces dernières permettent d'établir des connexions entre les différents types d'unités afin de récupérer la connexion la plus forte. Le système ne fonctionne donc pas par l'élaboration de règles de correspondances entre les graphèmes et les phonèmes, mais s'appuie plutôt sur le poids des connexions au sein des unités cachées (Sprenger-Charolles et Casalis, 1996). Les mots fréquents sont accédés de façon précoce aux unités sémantiques par leurs unités orthographiques, sans passer par les unités cachées ni par les unités phonologiques (Ferrand et Ayora, 2009), ce qui est insuffisant pour les mots de basse fréquence.

Enfin, Coltheart *et al.* (1993) ont, quant à eux, implanté un modèle connexionniste permettant de construire graduellement un ensemble de règles des correspondances entre les graphèmes et les phonèmes par les multiples rencontres avec les mots. Le modèle prévoit trois niveaux de détection (traits visuels, lettres, mots) avant de passer par la composante sémantique puis la composition phonologique de sortie permettant la prononciation du mot. Ce modèle permet non seulement d'associer un phonème à chaque lettre d'un mot simple (p. ex. *papa* = [papa]), mais aussi d'apprendre de nouvelles règles comme celles des graphèmes complexes (p. ex. *oi* = [wa]) et des graphies contextuelles (p. ex. *c* = [s] ou [k]) (Sprenger-Charolles et Casalis, 1996).

Les modèles connexionnistes reçoivent eux aussi des critiques. Ils sont conçus pour traiter les mots en les scindant en graphèmes/phonèmes. Or, il est difficile d'expliquer les théories mises de l'avant par les cognitivistes, notamment l'apport des unités syllabiques et sous-syllabiques (Goswami et Bryant, 1990) ou encore l'identification globale des mots fréquents (p. ex. Coltheart, 1978; Seymour, 1999) à l'aide des modèles connexionnistes.

APPENDICE B

RÉSULTATS DES ANALYSES DE COMPARAISON DES MOYENNES DES DIFFÉRENTES PROPRIÉTÉS PSYCHOLINGUISTIQUES DES CATÉGORIES DES ITEMS DE LA TÂCHE DE DÉCISION LEXICALE

À la section 3.3.1.1., nous présentons les résultats des analyses de comparaison des moyennes des indices de fréquence (fréquence du stimulus entier, fréquence de la première racine et fréquence de la deuxième racine) entre les catégories des items de la tâche de décision lexicale. D'autres propriétés psycholinguistiques ont été mesurées post hoc (Annexe D), mais ne font pas l'objet des analyses principales de la présente étude. Ainsi, les résultats des analyses de comparaisons des moyennes de ces propriétés psycholinguistiques entre les catégories des items de la tâche de décision lexicale sont présentés dans cette section.

Des ANOVAS à un facteur ont été réalisées pour mesurer la différence des moyennes des variables continues, soit le nombre de lettres, le nombre de syllabes orales, le nombre de graphèmes, la taille de la famille morphologique du stimulus entier, la taille de la famille morphologique de la première racine ainsi que la taille de la famille morphologique de la deuxième racine. La correction Bonferroni est appliquée aux comparaisons multiples. Lorsque les variances sont hétérogènes, les résultats de l'ANOVA de Welch sont rapportés et la correction Games-Howell est utilisée pour les comparaisons multiples. En raison de la multitude de comparaisons à effectuer et par souci de pertinence et d'espace, seuls les résultats significatifs sont rapportés.

Les résultats de l'ANOVA à un facteur signalent la présence d'au moins une différence significative entre les moyennes des huit catégories d'items à la tâche de décision lexicale quant au nombre de lettres, $F(7, 232) = 10,38, p < 0,001, \eta^2 = 0,24$. Les analyses des comparaisons multiples confirment que les mots formés de racines grecques comptent plus de lettres que les pseudomots formés de racines latines ($p = 0,023, d = 0,44$). Aussi, les mots monomorphémiques et la catégorie homologue, les non-mots, sont significativement plus courts que les mots ($p < 0,001, d = 0,72$) et pseudomots formés de racines grecques ($p < 0,001, d = 0,68$) ainsi que les mots ($p < 0,001, d = 0,62$) et pseudomots formés de racines libres ($p < 0,001, d = 0,59$).

Le nombre total de syllabes orales dans les items diffère entre les catégories selon les résultats de l'ANOVA à un facteur, $F(7, 99,24) = 10,15, p < 0,001, \eta^2 = 0,20$. Les mots et les pseudomots formés de

racines grecques comptent plus de syllabes que les autres items de la tâche de décision lexicale. Plus précisément, les analyses des comparaisons multiples montrent que les mots formés de racines grecques comptent plus de syllabes que les mots formés de racines libres ($p < 0,001$, $d = 0,70$), que les mots monomorphémiques ($p < 0,001$, $d = 0,60$), que les pseudomots formés de racines latines ($p = 0,021$, $d = 0,45$), que les pseudomots formés de racines libres ($p < 0,001$, $d = 0,76$) et que les non-mots ($p < 0,001$, $d = 0,64$). Quant à eux, les pseudomots formés de racines grecques comptent plus de syllabes que les mots formés de racines libres ($p < 0,001$, $d = 0,62$), que les mots monomorphémiques ($p = 0,004$, $d = 0,53$), que les pseudomots formés de racines libres ($p < 0,001$, $d = 0,67$) et que les non-mots ($p = 0,002$, $d = 0,56$).

En ce qui concerne le nombre de graphèmes, les résultats de l'ANOVA à un facteur signale des différences significatives entre les différentes catégories d'items à la tâche de décision lexicale, $F(7, 232) = 16,49$, $p < 0,001$, $\eta^2 = 0,33$. Les comparaisons multiples confirment que les mots monomorphémiques et les non-mots sont les items qui comptent le moins de graphèmes. Plus précisément, les mots monomorphémiques comptent significativement moins de graphèmes que les mots formés de racines latines ($p < 0,001$, $d = 0,64$), que les mots formés de racines grecques ($p < 0,001$, $d = 0,97$), que les mots formés de racines libres ($p = 0,036$, $d = 0,42$), que les pseudomots formés de racines latines ($p < 0,001$, $d = 0,59$), que les pseudomots formés de racines grecques ($p < 0,001$, $d = 0,94$) et que les pseudomots formés de racines libres ($p = 0,024$, $d = 0,44$). Quant à eux, les non-mots comptent aussi moins de graphèmes que les mots formés de racines latines ($p < 0,001$, $d = 0,66$), que les mots formés de racines grecques ($p < 0,001$, $d = 0,98$), que les mots formés de racines libres ($p < 0,024$, $d = 0,44$), que les pseudomots formés de racines latines ($p < 0,001$, $d = 0,61$), que les pseudomots formés de racines grecques ($p < 0,001$, $d = 0,95$) et que les pseudomots formés de racines libres ($p = 0,015$, $d = 0,45$). Aussi, les mots formés de racines grecques comptent plus de graphèmes que les mots ($p < 0,001$; $d = 0,55$) et pseudomots formés de racines libres ($p = 0,002$; $d = 0,53$). Le même patron se dessine en ce qui concerne les pseudomots formés de racines grecques. Ces derniers comptent plus de graphèmes que les mots ($p = 0,003$; $d = 0,51$) et pseudomots formés de racines libres ($p = 0,004$; $d = 0,50$).

La taille de la famille morphologique a aussi été répertoriée pour tous les stimulus entiers, en plus de la taille de la famille morphologique de la première et de la deuxième racine des items. Les résultats de l'ANOVA à un facteur ne montrent aucune différence significative entre les moyennes des quatre

catégories de mots à la tâche de décision lexicale (mots formés de racines latines, mots formés de racines grecques, mots formés de racines libres et mots monomorphémiques) pour la taille de la famille morphologique du stimulus entier, $F(3, 63,60) = 0,74, p = 0,533, \eta^2 = 0,01$. De plus, concernant la taille de la famille morphologique de la deuxième racine, aucune différence significative n'est présente entre les catégories d'items polymorphémiques (mots et pseudomots formés de racines latines, mots et pseudomots formés de racines grecques, mots et pseudomots formés de racines libres), $F(5, 174) = 0,200, p = 0,962, \eta^2 = 0,01$.

Le seul indice de la taille de la famille morphologique qui diffère significativement entre les catégories d'items polymorphémiques est celui de la première racine, $F(5, 78,91) = 9,53, p < 0,001, \eta^2 = 0,24$. Les racines libres présentes en tête de mots dans nos items, que ce soit les mots ou les pseudomots puisqu'ils partagent les mêmes racines, produisent des familles morphologiques beaucoup plus étendues que les racines latines ($p = 0,034; d = 0,47$) et grecques ($p < 0,001; d = 0,71$).

Enfin, trois autres propriétés psycholinguistiques ont été recueillies pour chacun des items de la tâche de décision lexicale. Il s'agit du cycle scolaire (1^{er} cycle, 2^e cycle ou 3^e cycle) d'apparition de la première occurrence du stimulus entier et de chacune des racines dans les manuels scolaires du primaire. Comme il s'agit de variables catégorielles, des tests de khi carré ont été effectués. Cependant, ceux-ci se sont avérés invalides étant donné que l'effectif théorique minimum n'est pas respecté. En effet, de façon générale, les mots et les racines qui n'apparaissent pas avant le 2^e cycle du primaire dans les manuels scolaires ne sont pas répertoriés. Très peu de mots et d'items apparaissent pour la première fois au 3^e cycle du primaire. Par conséquent, seuls les tableaux croisés de ces variables sont présentés afin de donner un aperçu de la répartition des données. Le Tableau B.1 présente la répartition des mots des catégories de la tâche de décision lexicale selon leur première occurrence dans les manuels scolaires au primaire. Les Tableaux B.2 et B.3 présentent respectivement celle de la première et de la deuxième racine des items des différentes catégories de la tâche de décision lexicale, toujours selon leur première occurrence dans les manuels scolaires au primaire.

Tableau B.1 Répartition des mots des catégories de la tâche de décision lexicale selon leur première occurrence dans les manuels scolaires au primaire

Catégories d'items	1 ^{er} cycle		2 ^e cycle		3 ^e cycle		Non répertorié		Total	
	n	%	n	%	n	%	n	%	n	%
(1) Mots formés de racines latines	5	16,7	9	30,0	0	0,0	16	53,3	30	100,0
(2) Mots formés de racines grecques	3	10,0	7	23,3	0	0,0	20	66,7	30	100,0
(3) Mots formés de racines libres	14	46,7	1	3,3	0	0,00	15	50,0	30	100,0
(4) Mots monomorphémiques	19	63,3	0	0,0	2	6,7	9	30,3	30	100,0
Total	41	34,2	17	14,2	2	1,7	60	50,0	120	100,0

Tableau B.2 Répartition des premières racines des items des catégories de la tâche de décision lexicale selon leur première occurrence dans les manuels scolaires au primaire

Premières racines des items	1 ^{er} cycle		2 ^e cycle		3 ^e cycle		Non répertoriée		Total	
	n	%	n	%	n	%	n	%	n	%
1 ^{es} racines latines	22	73,3	3	10,0	2	6,7	3	10,0	30	100,0
1 ^{es} racines grecques	15	50,0	10	33,3	0	0,0	5	16,7	30	100,0
1 ^{es} racines libres	30	100,0	0	0,0	0	0,0	0	0,0	30	100,0
Total	67	74,4	13	14,4	2	2,2	8	8,9	90	100,0

Tableau B.3 Répartition des deuxièmes racines des items des catégories de la tâche de décision lexicale selon leur première occurrence dans les manuels scolaires au primaire

Deuxièmes racines des items	1 ^{er} cycle		2 ^e cycle		3 ^e cycle		Non répertoriée		Total	
	n	%	n	%	n	%	n	%	n	%
2 ^{es} racines latines	25	83,3	2	6,7	1	3,3	2	6,7	30	100,0
2 ^{es} Racines grecques	14	46,7	10	33,3	1	3,3	5	16,7	30	100,0
2 ^{es} Racines libres	30	100,0	0	0,0	0	0,0	0	0,0	30	100,0
Total	69	76,7	12	13,3	2	2,2	7	7,8	90	100,0

RÉFÉRENCES

- Amenta, S. et Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, 3(232), 1-12. <https://doi.org/10.3389/fpsyg.2012.00232>
- American Psychological Association. (s.d.). *APA dictionary of psychology*. Récupéré le 31 août 2022, de <https://dictionary.apa.org/>
- Amiot, D. (2010). Paradigmes, radicaux supplétifs et constituants néoclassiques en morphologie constructionnelle. *Paradigmes en morphologie constructionnelle*, 21-36. <https://doi.org/halshs-00725972>
- Analyse et traitement informatique de la langue française [ATILF]. (s.d.). *TLFi : Trésor de la langue Française informatisé*. ATILF - CNRS & Université de Lorraine. Récupéré le 23 septembre 2021 de <http://www.atilf.fr/tlfi>
- Anctil, D., Singcastar, M. et Tardif, M. (2018). Pratiques d'enseignement du lexique en classe de primaire au Québec. *La lettre*, 64, 19-25. <https://doi.org/10.3406/airdf.2018.2247>
- Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive science*, 26, 85-112. [https://doi.org/10.1016/S0364-0213\(01\)00062-3](https://doi.org/10.1016/S0364-0213(01)00062-3)
- Andrews, S. et Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, 68, 279-296. <http://dx.doi.org/10.1016/j.jml.2012.12.001>
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10), 1-166. <https://doi.org/10.2307/1166112>
- Arcara, G., Marelli, M., Buodo, G. et Mondini, S. (2013). Compound headedness in the mental lexicon: an event-related potential study. *Cognitive Neuropsychology*, 31(1-2), 164-183. <https://doi.org/10.1080/02643294.2013.847076>
- Aronoff, M. et Fudeman, K. (2023). *What is morphology?* (3^e éd). Wiley.
- Bandettini, P. A. et Ungerleider, L. G. (2001). From neuron to BOLD: new connections. *Nature Neurosciences*, 4(9), 864-866. <https://doi.org/10.1038/nn0901-864>
- Bar-Ilan, L. et Berman, R. A. (2007). Developing register differentiation: The Latinate-Germanic divide in English. *Linguistics*, 45(1), 1-35. <https://doi.org/10.1515/LING.2007.001>
- Bara, F., Gentaz, É. et Colé, P. (2008). Littérature précoce et apprentissage de la lecture : comparaison entre des enfants à risque, scolarisés en France dans des réseaux d'éducation prioritaire, et des enfants de classes régulières. *Revue des sciences de l'éducation*, 34(1), 27-45. <https://doi.org/https://doi.org/10.7202/018988ar>
- Baumann, J. F., Edwards, E. C., Boland, E. et Font, G. (2012). Teaching word-learning strategies. Dans E. J. Kame'enui et J. F. Baumann (dir.), *Vocabulary instruction. Research to practice*. (2^e éd., p. 139-166). The Guilford Press.
- Baumann, J. F., Edwards, E. C., Boland, E. M., Olejnik, S. et Kame'enui, E. J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Journal*, 40(2), 447-494. <https://doi.org/10.3102/00028312040002447>
- Baumann, J. F., Edwards, E. C., Font, G., Tereshinski, C. A., Kame'enui, E. J. et Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly*, 37(2), 150-176. <https://doi.org/10.1598/RRQ.37.2.3>
- Baumann, J. F. et Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy*, 54(1), 4-12. <https://doi.org/10.1598/JAAL.54.1.1>

- Beck, I. L., McKeown, M. G. et Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. The Guilford Press.
- Beck, I. L., McKeown, M. G. et Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2^e éd.). The Guilford Press.
- Beck, I. L., McKeown, M. G. et Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. Dans M. G. McKeown et M. E. Curtis (dir.), *The nature of vocabulary acquisition* (p. 147-163). Lawrence Erlbaum Associates.
- Bertrand, O. (2011). *Histoire du vocabulaire français*. Éditions de l'école Polytechnique.
- Beyersmann, E. et Grainger, J. (2018). Support from the morphological family when unembedding the stem. *Journal of experimental psychology: Learning, Memory, and Cognition*, 44(1), 135-142. <https://doi.org/10.1037/xlm0000435>
- Beyersmann, E., Iakimova, G., Ziegler, J. C. et Colé, P. (2014). Semantic processing during morphological priming: An ERP study. *Brain Research*, 1579(C), 45-55. <https://doi.org/10.1016/j.brainres.2014.07.010>
- Beyersmann, E., Kezilas, Y., Coltheart, M., Castles, A., Ziegler, J. C., Taft, M. et Grainger, J. (2018). Taking the book from the bookshelf: Masked constituent priming effects from compound words and nonwords. *Journal of Cognition*, 1(1), 1-13. <https://doi.org/10.5334/joc.11>
- Beyersmann, E., Ziegler, J., Castles, A., Coltheart, M., Kezilas, Y. et Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, 23(2), 533-539. <https://doi.org/10.3758/s13423-015-0927-z>
- Bianco, M. (2015). *Du langage oral à la compréhension de l'écrit*. Presses universitaires de Grenoble.
- Blachowicz, C. L. Z. et Fisher, P. J. (2010). *Teaching vocabulary in all classrooms* (4^e éd.). Allyn & Bacon.
- Blachowicz, C. L. Z. et Fisher, P. J. L. (2000). Vocabulary Instruction. Dans R. Barr, M. L. Kamil, P. B. Mosenthal et P. D. Pearson (dir.), *Handbook of reading research* (vol. 3, p. 503-523). Lawrence Erlbaum Associates.
- Bouffartigue, J. (1983). *Trésors des racines latines*. Belin.
- Bouffartigue, J. et Delrieu, A.-M. (1981). *Trésors des racines grecques*. Belin.
- Boudewyn, M. A., Luck, S. J., Farrens, J. L. et Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, 55(6), e13049. <https://doi.org/10.1111/psyp.13049>
- Bowers, P. N. et Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing*, 23, 515-537. <https://doi.org/10.1007/s11145-009-9172-z>
- Bowers, P. N., Kirby, J. R. et Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, 80(2), 144-179. <https://doi.org/10.3102/0034654309359353>
- Boyle, R. (2009). The legacy of diglossia in English vocabulary: What learners need to know. *Language Awareness*, 18(1), 19-30. <https://doi.org/10.1080/09658410802147360>
- Brain Products GmbH. (2018, 25 mai). *Analysis of Event-Related Potentials* [Webinaire]. GoTo Stage. <https://www.gotostage.com/channel/a26a09b972b84c52a391a16dfb4298e9/recording/081b4a7c72294eee9b8786603c6f6a6d/watch?source=WEBSITE>
- Brain Vision Solutions. (2020). *BrainVision Recorder Professional* (version 1.21.0402).
- Brewer, M. B. et Crano, W. D. (2014). Research design and issues of validity. Dans H. T. Reis et C. M. Judd (dir.), *Handbook of research methods in social and personality psychology* (2^e éd.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511996481>
- Brooks, T. L. et Cid de Garcia, D. (2015). Evidence for morphological composition in compound words using MEG. *Frontiers in Human Neuroscience*, 9(215). <https://doi.org/10.3389/fnhum.2015.00215>

- Bruck, M. et Treiman, R. (1992). Learning to pronounce words : The limitations of analogies. *Reading Research Quarterly*, 27(4), 375-388. <https://doi.org/10.2307/747676>
- Burani, C. et Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, 2(3/4), 217-227. <https://doi.org/10.1080/01690968708406932>
- Burton, V. J. et Watkins, R. V. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders*, 40(5), 335-356. <https://doi.org/10.1016/j.jcomdis.2006.06.015>
- Butterworth, B. (1983). Lexical representation. Dans B. Butterworth (dir.), *Language production. Development, Writing and Other Language Processes* (vol. 2, p. 257-294). Academic Press.
- Cabral, L. (2015). *The acquisition of science vocabulary: A morphological approach to content area literacy* [Thèse de doctorat, Northeastren University]. Northeastren University Library. <https://doi.org/10.17760/D20154938>
- Caramazza, A., Laudanna, A. et Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28(3), 297-332. [https://doi.org/10.1016/0010-0277\(88\)90017-0](https://doi.org/10.1016/0010-0277(88)90017-0)
- Caramazza, A., Miceli, G., Silveri, M. C. et Laudanna, A. (1985). Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology*, 2(1), 81-114. <https://doi.org/10.1080/02643298508252862>
- Carlisle, J. F. (1995). Morphological awareness and early reading achievement. Dans L. B. Feldman (dir.), *Morphological aspects of language processing* (p. 189-209). Erlbaum.
- Carlisle, J. F. et Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, 40(4), 428-449. <https://doi.org/10.1598/RRQ.40.4.3>
- Cartier, S. C. (2007). *Apprendre en lisant au primaire et au secondaire. Mieux comprendre et mieux intervenir*. Les éditions CEC.
- Casalis, S., Dusautoir, M., Colé, P. et Ducrot, S. (2009). Morphological effects in children word reading: a priming study in fourth graders. *British Journal of Developmental Psychology*, 27(3), 761-766. <https://doi.org/10.1348/026151008x389575>
- Casalis, S., Quémart, P. et Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: evidence from a cross-language study. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00452>
- Catach, N. (2016). *L'orthographe française. Traité théorique et pratique*. Armand Colin.
- Cellard, J. (2013). *Les racines grecques et latines du vocabulaire français*. De Boeck Supérieur.
- Cervetti, G. N., Hiebert, E. H., Pearson, P. D. et McClung, N. A. (2015). Factors that influence the difficulty of science words. *Journal of Literacy Research*, 47(2). <https://doi.org/10.1177/1086296X15615363>
- Chatrian, G. E., Lettich, E. et Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*, 25(2), 83-92. <https://doi.org/10.1080/00029238.1985.11080163>
- Chialant, D. et Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. Dans L. Feldman (dir.), *Morphological aspects of language processing: cross-linguistics perspectives* (p. 55-76). Lawrence Erlbaum Associates.
- Clark, E. (1993). *The lexicon in acquisition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511554377>
- Coch, D., Bares, J. et Landers, A. (2013). ERPs and morphological processing: The N400 and semantic composition. *Cognitive, Affective, & Behavioral Neuroscience*, 13, 355-370. <https://doi.org/10.3758/s13415-012-0145-3>
- Coch, D., Hua, J. et Landers-Nelson, A. (2020). All morphemes are not the same: Accuracy and response times in a lexical decision task differentiate types of morphemes. *Journal of Research in Reading*, 43(3), 329-346. <https://doi.org/10.1111/1467-9817.12306>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2e éd.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Colé, P., Beauvillain, C., Pavard, B. et Segui, J. (1986). Organisation morphologique et accès au lexique. *L'année psychologique*, 86(3), 349-365. <https://doi.org/10.3406/psy.1986.29154>
- Colé, P., Beauvillain, C. et Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, 28(1), 1-13. [https://doi.org/10.1016/0749-596X\(89\)90025-9](https://doi.org/10.1016/0749-596X(89)90025-9)
- Coltheart, M. (1978). Lexical access in simple reading tasks. Dans G. Underwood (dir.), *Strategies of Information Processing* (p. 151-216). Academic Press.
- Coltheart, M., Curtis, B., Atkins, P. et Haller, M. (1993). Models of reading aloud : Dual route and parallel processing approaches. *Psychological Review*, 100(4), 589-608. <https://doi.org/10.1037/0033-295X.100.4.589>
- Conseil supérieur de la langue française. (1990). *Les rectifications de l'orthographe* (N°100). https://www.academie-francaise.fr/sites/academie-francaise.fr/files/rectifications_1990.pdf
- Crosson, A. C. et McKeown, M. G. (2016). Middle school learners' use of Latin roots to infer the meaning of unfamiliar words. *Cognition and Instruction*, 34(2), 148-171. <https://doi.org/10.1080/07370008.2016.1145121>
- Crosson, A. C., McKeown, M. G., Moore, D. W. et Ye, F. (2019). Extending the bounds of morphology instruction: Teaching Latin roots facilitates academic word learning for English Learner adolescents. *Reading and Writing*, 32(3), 689-727. <https://doi.org/10.1007/s11145-018-9885-y>
- Crosson, A. C. et Moore, D. (2017). When to take up roots: The effects of morphology instruction for middle school and high school english learners. *Reading Psychology*, 38(3), 262-288. <https://doi.org/10.1080/02702711.2016.1263699>
- Crosson, A. C., Shen, Y., McKeown, M. G., Lei, P., Zhao, H., Li, X., Patrick, K. et Brown, K. (2021). Morphological analysis skill and academic vocabulary knowledge are malleable through intervention and may contribute to reading comprehension for multilingual adolescents. *Journal of Research in Reading*, 44(1), 154-174. <https://doi.org/10.1111/1467-9817.12323>
- Daigle, D. et Gagnon, I. (2014). La morphologie dérivationnelle comme levier de réussite pour un jeune adulte dyslexique et dysorthographique. *Correspondance*, 19(2). <https://correspo.ccdmd.qc.ca/index.php/document/haro-sur-lecole/la-morphologie-derivationnelle-comme-levier-de-reussite-pour-un-jeune-adulte-dyslexique-et-dysorthographique/>
- Dal, G. et Amiot, D. (2008). Composition néoclassique en français et ordre des constituants. Dans D. Amiot (dir.), *La composition dans une perspective typologique* (p. 89-113). Artois Presses Université.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, 42(8), 895-901. <https://www.jstor.org/stable/41385916>
- Daniels, H. (2002). Expositoy text in literature circles. *Voices from the Middle*, 9(4), 7-14.
- Davis, C. P., Libben, G. et Segalowitz, S. J. (2019). Compounding matters: Event- related potential evidence for early semantic access to compound words. *Cognition*, 184, 44-52. <https://doi.org/10.1016/j.cognition.2018.12.006>
- Deacon, S. H., Parrila, R. et Kirby, J. R. (2008). A review of the evidence on morphological processing in dyslexics and poor readers: A strength or weakness? Dans G. Reid, A. Fawcett, F. Manis et L. Siegel (dir.), *The SAGE Handbook of Dyslexia* (p. 212-237). SAGE Publishing. <https://doi.org/10.4135/9780857020987.n10>
- Desrosiers, H. et Tétrault, K. (2012). *Les facteurs liés à la réussite aux épreuves obligatoires de français en sixième année du primaire: un tour d'horizon* (vol. 7, Fascicule 1). Québec : Gouvernement du Québec, Institut de la statistique du Québec.

- Diependaele, K., Grainger, J. et Sandra, D. (2012). Derivational Morphology and Skilled Reading. An Empirical Overview. Dans M. J. Spivey, K. McRae et M. F. Joanisse (dir.), *The Cambridge handbook of psycholinguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139029377>
- Diependaele, K., Sandra, D. et Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, 37(6), 895-908. doi:10.3758/MC.37.6.895
- Dressler, W. U. (2006). Compound Types. Dans G. Libben et G. Jarema (dir.), *The representation and processing of compound words* (p. 23-44). Oxford university press. <https://doi.org/10.1093/acprof:oso/9780199228911.003.0002>.
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, 12(2), 45-64. <https://doi.org/10.3917/rfla.122.0045>
- Druide informatique inc. (s.d.). *Antidote 10*. Récupéré le 23 septembre 2021 de <https://www.antidote.info/fr>
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B. et Mével, J.-P. (2002) *Dictionnaire de linguistique*. Larousse.
- Duñabeitia, J. A., Laka, I., Perea, M. et Carreiras, M. (2009). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *European Journal of Cognitive Psychology*, 21(4), 615-640. <https://doi.org/10.1080/09541440802079835>
- Duñabeitia, J. A., Perea, M. et Carreiras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review*, 14(6), 1171-1176. <https://doi.org/10.3758/bf03193108>
- Dymock, S. (2005). Teaching expository text structure awareness. *The Reading Teacher*, 59(2), 177-181. <https://doi.org/10.1598/RT.59.2.7>
- El Yagoubi, R., Chiarelli, V., Mondini, S., Perrone, G., Danieli, M. et Semenza, C. (2008). Neural correlates of Italian nominal compounds and potential impact of headedness effect: An ERP study. *Cognitive Neuropsychology*, 25(4), 559-581. <https://doi.org/10.1080/02643290801900941>
- Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, 28(5), 491-520. <https://doi.org/10.1080/09500690500339092>
- Fazio, X. et Gallagher, T. L. (2014). Morphological development levels of science content vocabulary: Implications for science-based texts in elementary classrooms. *International Journal of Science and Mathematics Education*, 12(6), 1407-1423. <https://doi.org/10.1007/s10763-013-9470-4>
- Fejzo, A. (2011). *Les effets d'un programme de développement de la compétence morphologique sur la compétence morphologique, l'identification et la production des mots écrits chez des élèves arabophones de 2^e cycle du primaire scolarisés en français* [Thèse de doctorat, Université du Québec à Montréal]. Archipel. <http://archipel.uqam.ca/id/eprint/4486>
- Fejzo, A. (2015). Portrait des lecteurs et des scripteurs de 4^e année et effets d'une intervention en conscience morphologique sur l'orthographe lexicale [Manuscrit non publié]. *Département de didactique des langues, Université du Québec à Montréal*.
- Fejzo, A. (2021). The contribution of morphological awareness to vocabulary among L1 and L2 French-speaking 4th-graders. *Reading and Writing*, 34(3), 659–679. <https://doi.org/10.1007/s11145-020-10084-8>
- Fejzo, A., Desrochers, A. et Deacon, S. H. (2018). The acquisition of derivational morphology in children. Dans R. Berthiaume, D. Daigle et A. Desrochers (dir.), *Morphological Processing and Literacy Development: Current Issues and Research* (p. 112-132). Routledge. <https://doi.org/10.4324/9781315229140>.

- Fejzo, A., Godard, L. et Laplante, L. (2014). L'analyse morphologique: pierre angulaire dans l'enrichissement du vocabulaire. *Québec français*, 171, 80-82. <https://id.erudit.org/iderudit/71231ac>
- Ferrand, L. (2001). *Cognition et lecture. Processus de base de la reconnaissance des mots écrits chez l'adulte*. De Boeck Supérieur.
- Ferrand, L. et Ayora, P. (2009). *Psychologie cognitive de la lecture. Reconnaissance des mots écrits chez l'adulte*. De Boeck Supérieur.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4^e éd.). SAGE Publications.
- Fiorentino, R., Naito-Billen, Y., Bost, J. et Fund-Reznicek, E. (2014). Electrophysiological evidence for the morpheme-based combinatoric processing of English compounds. *Cognitive Neuropsychology*, 31(1-2), 123-146. <https://doi.org/10.1080/02643294.2013.855633>
- Fiorentino, R. et Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22(7), 953-1000. <https://doi.org/10.1080/01690960701190215>
- Fløttum, K., Dahl, T., Didriksen, A. A. et Gjesdal, A. M. (2013). KIAP – reflections on a complex corpus. *Bergen Language and Linguistics Studies*, 3(1), 137-150. <https://doi.org/10.15845/bells.v3i1.367>
- Fluss, J., Bertrand, D., Ziegler, J. et Billard, C. (2009). Troubles d'apprentissage de la lecture : Rôle des facteurs cognitifs, comportementaux et socio-économiques. *Développements*, 1(1), 21-33. <https://doi.org/10.3917/devel.001.0021>
- Forster, K. I. et Azuma, T. (2000). Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, 15(4/5), 539-561. <https://doi.org/10.1080/01690960050119698>
- Forster, K. I. et Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116-124. <https://doi.org/10.3758/BF03195503>
- Frauenfelder, U. H. et Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. Dans G. Booij et J. van Marle (dir.), *Yearbook of morphology 1991* (p. 165-183). Springer Netherlands. <https://doi.org/10.1007/978-94-011-2516-1>.
- Freeman, N. D., Townsend, D. et Templeton, S. (2019). Thinking about words: First graders' response to morphological instruction. *The Reading Teacher*, 72(4), 463-473. <https://doi.org/10.1002/trtr.1749>
- Frith, U. (1985). Beneath the surface of developmental dyslexia. Dans K. E. Patterson, J. C. Marshall et M. Coltheart (dir.), *Surface dyslexia : Cognitive and neuropsychological studies of phonological reading* (p. 301-330). Erlbaum. <https://doi.org/10.4324/9781315108346>
- Gala, N., François, T., Bernhard, D. et Fairon, C. (2014, 1-4 juillet). *Un modèle pour prédire la complexité lexicale et graduer les mots* [Acte de conférence]. Traitement Automatique des Langues Naturelles, Marseille, France. <https://hal.archives-ouvertes.fr/hal-01001916>
- Gala, N. et Rey, V. (2008, 9-13 juin). *POLYMOTS : une base de données de constructions dérivationnelles en français à partir de radicaux phonologiques* [Acte de conférence]. Traitement Automatique des Langues Naturelles, Avignon, France. <https://aclanthology.org/2008.jeptalnrecital-court.10>
- Gentaz, E., Sprenger-Charolles, L. et Theurel, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLOS ONE*, 10(3), e0119581. <https://doi.org/10.1371/journal.pone.0119581>
- Gentaz, E., Sprenger-Charolles, L., Theurel, A. et Colé, P. (2013). Reading comprehension in a large cohort of French first graders from low socio-economic status families: A 7-month longitudinal study. *PLOS ONE*, 8(11), e78608. <https://doi.org/10.1371/journal.pone.0078608>
- Giraud, H. (2005). Un modèle supralexicale de représentation de la morphologie dérivationnelle en français. *L'année psychologique*, 105(1), 171-195. <https://doi.org/10.3406/psy.2005.3825>

- Giraud, H. et Grainger, J. (2000). Effects of prime word frequency and morphemic frequency in masked morphological priming. *Language and Cognitive Processes*, 15(4/5), 421-444. <https://doi.org/10.1080/01690960050119652>
- Giraud, H. et Grainger, J. (2001). Priming complex words: Evidence for supralexicalexical representation of morphology. *Psychonomic Bulletin & Review*, 8(1), 127-131. <https://doi.org/10.3758/BF03196148>
- Giraud, H. et Grainger, J. (2003). Supralexicalexical model for French derivational morphology. Dans E. M. H. Assink et D. Sandra (dir.), *Reading complex words, Cross-language studies* (vol. 22, p. 139-157). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4757-3720-2_7.
- Gonnerman, L. M. et Plaut, D. C. (2000). *Semantic and morphological effects in masked priming* [Acte de conférence]. Cognitive Neuroscience Society Annual Meeting.
- Gonnerman, L. M., Seidenberg, M. S. et Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, 136(2), 323-345. <https://doi.org/10.1037/0096-3445.136.2.323>
- González, G. F., Žarić, G., Tijms, J., Bonte, M., Blomert, L., Leppänen, P. et van der Molen, M. W. (2016). Responsivity to dyslexia training indexed by the N170 amplitude of the brain potential elicited by word reading. *Brain and Cognition*, 106, 42-54. <https://doi.org/10.1016/j.bandc.2016.05.001>
- Goodwin, A. P. (2016). Effectiveness of word solving: Integrating morphological problem-solving within comprehension instruction for middle school students. *Reading and Writing*, 29, 91-116. <https://doi.org/10.1007/s11145-015-9581-0>
- Goodwin, A. P., Gilbert, J. K. et Cho, S.-J. (2013). Morphological contributions to adolescent word reading: An item response approach. *Reading Research Quarterly*, 48(1), 39-60. <https://doi.org/10.1002/rrq.037>
- Goswami, U. et Bryant, P. (1990). *Phonological skills and learning to read*. Erlbaum.
- Grainger, J. et Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. *Psychology of Learning and Motivation*, 67, 285-317. <https://doi.org/10.1016/bs.plm.2017.03.009>
- Grainger, J., Colé, P. et Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, 30(3), 370-384. [https://doi.org/10.1016/0749-596X\(91\)90042-I](https://doi.org/10.1016/0749-596X(91)90042-I)
- Grainger, J. et Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, 3(1), 128-156. <https://doi.org/10.1111/j.1749-818X.2008.00121.x>
- Green, T. M. (2008). *The Greek and Latin roots of English* (4^e éd.). Rowman & Littlefield.
- Green, T. M. (2015). *The Greek and Latin roots of English* (5^e éd.). Rowman & Littlefield.
- Gross, M. (1982). Une classification des phrases « figées » du français. *Revue québécoise de linguistique*, 11(2), 151-185. <https://doi.org/10.7202/602492ar>
- Harris, M. et Coltheart, M. (1986). *Language processing in children and adults : An introduction*. Routledge & Kegan.
- Hatier, S. (2013, 17-21 juin). *Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines: une première expérimentation* [Acte de conférence]. Traitement Automatique des Langues Naturelles, Les Sables d'Olonne, France.
- Henderson, L. et Chard, J. (1980). The reader's implicit knowledge of orthographic structure. Dans U. Frith (dir.), *Cognitive processes in spelling* (p. 85-116). Academic Press.
- Henry, M. (2010). *Unlocking literacy: Effective decoding and spelling instruction* (2^e éd.). Brookes.
- Hiebert, E. H. (2011, 7 juin). The 90-10 rule of vocabulary in increasing students' capacity for complex text. *TextProject*. <https://textproject.org/the-90-10-rule-of-vocabulary-in-increasing-students-capacity-for-complex-text/>

- Hiebert, E. H. (2020). The core vocabulary: The foundation of proficient comprehension. *Reading Teacher*, 73(6), 757-768. <https://doi.org/10.1002/trtr.1894>
- Hiebert, E. H. et Cervetti, G. N. (2012). What differences in narrative and informational texts mean for the learning and instruction of vocabulary. Dans E. J. Kame'enui et J. F. Baumann (dir.), *Vocabulary instruction. Research to practice*. (2^e éd., p. 322-344). The Guilford Press.
- Hiebert, E. H., Scott, J., Castaneda, R. et Spichtig, A. (2019). An analysis of the features of words that influence vocabulary difficulty. *Education Sciences*, 9(1). <https://doi.org/10.3390/educsci9010008>
- Holcomb, P. J. et Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, 18(10), 1631-1643. <https://doi.org/10.1162/jocn.2006.18.10.1631>
- Hoover, W. A. et Tunmer, W. E. (2020). *The cognitive foundations of reading and its acquisition. A framework with applications connecting teaching and learning* (vol. 20). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-030-44195-1>
- Hruby, G. G. (2017). Literacy, Comprehension, and the Neurosciences. Dans S. E. Israel (dir.), *Handbook of research on reading comprehension* (2^e éd., p. 191-216). The Guilford Press.
- Hruby, G. G. et Goswami, U. (2011). Neuroscience and reading: A review for reading education researchers. *Reading Research Quarterly*, 46(2), 156-172. <https://doi.org/10.1598/RRQ.46.2.4>
- Huot, H. (2005). *La morphologie: Forme et sens des mots du français* (2e éd.). Armand Colin.
- IBM Corp. (2021). *IBM SPSS Statistics for Windows, Version 28.0*.
- Janosz, M., Pascal, S., Belleau, L., Archambault, I., Parent, S. et Pagani, L. (2013). *Les élèves du primaire à risque de décrocher au secondaire: caractéristiques à 12 ans et prédicteurs à 7 ans* (vol. 7, Fascicule 2). Québec : Gouvernement du Québec, Institut de la statistique du Québec.
- Jarema, G. (2006). Compound representation and processing: A cross-language perspective. Dans G. Libben et G. Jarema (dir.), *The representation and processing of compound words* (p. 45-70). Oxford university press. <https://doi.org/10.1093/acprof:oso/9780199228911.003.0003>.
- Just, M. A. et Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87(4), 329-354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kaczer, L., Timmer, K., Bavassi, L. et Schiller, N. O. (2015). Distinct morphological processing of recently learned compound words: An ERP study. *Brain Research*, 1629, 309-317. <https://doi.org/10.1016/j.brainres.2015.10.029>
- Kieffer, M. J. et Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, 25, 347-373. <https://doi.org/10.1007/s11145-010-9272-9>
- Kim, A. E., McKnight, S. M. et Miyake, A. (2024). How variable are the classic ERP effects during sentence processing? A systematic resampling analysis of the N400 and P600 effects. *Cortex*, 177, 130-149. <https://doi.org/10.1016/j.cortex.2024.05.007>
- Kirby, J. R. et Bowers, J. S. (2018). The effects of morphological instruction on vocabulary learning, reading, and spelling. Dans R. Berthiaume, D. Daigle et A. Desrochers (dir.), *Morphological processing and literacy development* (p. 217-243). Routledge. <https://doi.org/10.4324/9781315229140>
- Kirby, J. R. et Bowers, P. N. (2017). Morphological instruction and literacy. Binding phonological, orthographic and semantic features of words. Dans K. Cain, D. L. Compton et R. K. Parrilla (dir.), *Theories of Reading Development* (p. 437-462). John Benjamins Publishing Company. <https://doi.org/10.1075/swll.15>
- Koester, D., Gunter, T. C. et Wagner, S. (2007). The morphosyntactic decomposition and semantic composition of German compound words investigated by ERPs. *Brain and Language*, 102, 64-79. <https://doi.org/10.1016/j.bandl.2006.09.003>

- Krott, A., Baayen, R. H. et Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1), 51-93. <https://doi.org/10.1515/ling.2001.008>
- Kuo, L.-j. et Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, 41(3), 161-180. https://doi.org/10.1207/s15326985ep4103_3
- Kuperman, V., Schreuder, R., Bertram, R. et Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology Human Perception & Performance*, 35(3), 876-895. <https://doi.org/10.1037/a0013484>
- 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>
- Larousse. (s.d.). Larousse. Récupéré le 23 septembre 2021 de <https://www.larousse.fr/dictionnaires/francais>
- Laudanna, A. et Burani, C. (1985). Address mechanisms to decomposed lexical entries. *Linguistics*, 23, 775-792. <https://doi.org/10.1515/ling.1985.23.5.775>
- Laudanna, A., Burani, C. et Cermele, A. (1994). Prefixes as processing units. *Language and Cognitive Processes*, 9(3), 295-316. <https://doi.org/10.1080/01690969408402121>
- Le Robert. (2021). *Le Robert dico en ligne*. Récupéré le 23 septembre 2021 de <https://dictionnaire.lerobert.com/>
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3^e éd.). Guérin.
- Leminen, A., Smolka, E., Duñabeitia, J. A. et Pliatsikas, C. (2018). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex*, 116, 4-44. <https://doi.org/10.1016/j.cortex.2018.08.016>
- Lesaux, N. K. et Kieffer, M. J. (2010, 2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, 47(3), 596-632. <https://doi.org/10.3102/0002831209355469>
- Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156-166. <https://doi.org/10.3758/BF03195560>
- Libben, G. (2006). Why study compound processing? An overview of the issues. Dans G. Libben et G. Jarema (dir.), *The representation and processing of compound words* (p. 1-22). Oxford university press. <https://doi.org/10.1093/acprof:oso/9780199228911.003.0001>
- LIF-TALEP. (2012). *Polymots*. Récupéré le 6 juin 2021 de <http://polymots.lif.univ-mrs.fr/v2/>
- Longtin, C.-M. et Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, 53(1), 26-41. <https://doi.org/10.1016/j.jml.2005.02.008>
- Longtin, C.-M., Segui, J. et Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18(3), 313-334. <https://doi.org/10.1080/01690960244000036>
- Lopes da Silva, F. H. (2013). EEG and MEG: Relevance to neuroscience. *Neuron*, 80(5), 1112-1128. <https://doi.org/10.1016/j.neuron.2013.10.017>
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (2^e éd.). Massachusetts Institute of Technology.
- MacGregor, L. J. et Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, 126, 217-229. <https://doi.org/10.1016/j.bandl.2013.04.002>
- Manelis, L. et Tharp, D. A. (1977). The processing of affixed words. *Memory & Cognition*, 5(6), 690-695. <https://doi.org/10.3758/BF03197417>

- Manyak, P. C., Baumann, J. F. et Manyak, A. M. (2018). Morphological analysis instruction in the elementary grades: Which morphemes to teach and how to teach them. *Reading Teacher*, 72(3), 289-300. <https://doi.org/10.1002/trtr.1713>
- Marsh, G., Friedman, M., Welsch, V. et Desberg, P. (1981). A Cognitive developmental theory of reading acquisition. Dans G. E. MacKinnon et T. G. Waller (dir.), *Reading research : Advances in theory and practice* (vol. 3, p. 199-221). Erlbaum.
- Marzano, R. J. (2020). *Teaching basic, advanced, and academic vocabulary: A comprehensive framework for elementary instruction*. Marzano Resources.
- Mattiello, E. (2016). Analogical neologisms in English. *Italian Journal of Linguistics*, 28(2), 103-142. https://www.italian-journal-linguistics.com/app/uploads/2021/05/4_Mattiello.pdf
- McClelland, J. L. et Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375-407. <https://doi.org/10.1037/0033-295X.88.5.375>
- McKeown, M. G., Beck, I. L. et Sandora, C. (2012). Direct and rich vocabulary instruction needs to start early. Dans E. J. Kame'enui et J. F. Baumann (dir.), *Vocabulary instruction. Research to practice*. (2^e éd., p. 17-33). The Guildford Press.
- McKeown, M. G., Crosson, A. C., Moore, D. W. et Beck, I. L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal*, 55(3), 572-616. <https://doi.org/10.3102/0002831217744181>
- McKinnon, R., Allen, M. et Osterhout, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *Cognitive neuroscience and neuropsychology*, 14(6), 883-886. <https://doi.org/10.1097/01.wnr.0000070192.28954.7e>
- Mercier, P., Gagnon, M. et Clément, M. (2005). Les protocoles de recherche pré, quasi et expérimentaux. Dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale: Pour harmoniser recherche et pratique* (2^e éd., p. 87-148). Presses de l'Université du Québec.
- Milin, P., Smolka, E. et Feldman, L. B. (2017). Models of lexical access and morphological processing. Dans E. M. Fernández et H. S. Cairns (dir.), *The Handbook of Psycholinguistics* (p. 240-268). Wiley. <https://doi.org/10.1002/9781118829516.ch11>.
- Ministère de l'Éducation. (2006a). *Programme de formation de l'école québécoise. Éducation préscolaire. Enseignement primaire*. <http://www.education.gouv.qc.ca/enseignants/pfeq/primaire/>
- Ministère de l'Éducation. (2006b). *Programme de formation de l'école québécoise. Enseignement secondaire, premier cycle*. <http://www.education.gouv.qc.ca/enseignants/pfeq/secondaire/>
- Ministère de l'Éducation, du Loisir et du Sport. (2007). *Programme de formation de l'école québécoise. Enseignement secondaire, deuxième cycle*. <http://www.education.gouv.qc.ca/enseignants/pfeq/secondaire/>
- Ministère de l'Éducation, du Loisir et du Sport. (2009). *Progression des apprentissages. Français, langue d'enseignement*. http://www.education.gouv.qc.ca/fileadmin/site_web/documents/education/jeunes/pfeq/PDA_PFEQ_francais-langue-enseignement-primaire_2011.pdf
- Ministère de l'Éducation, du Loisir et du Sport. (2014). *Liste orthographique à l'usage des enseignantes et des enseignants*. http://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/formation_jeunes/Liste_Orthographique_Primaire.pdf
- Morais, J. (1999). *L'art de lire*. Odile Jacob.
- Morris, J., Frank, T., Grainger, J. et Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, 44, 506-521. <https://doi.org/10.1111/j.1469-8986.2007.00538.x>

- Morris, J., Grainger, J. et Holcomb, P. J. (2008). An electrophysiological investigation of early effects of masked morphological priming. *Language and Cognitive Processes*, 23(7-8), 1021-1056. <https://doi.org/10.1080/01690960802299386>
- Morton, J. (1989). An information-processing account of reading acquisition. Dans A. M. Galaburda (dir.), *From neurons to reading* (p. 43-66). MIT Press.
- Mousty, P. et Leybaert, J. (1999). Evaluation des habiletés de lecture et d'orthographe au moyen de BELEC: Données longitudinales auprès d'enfants francophones testés en 2ème et 4ème années primaire. *Revue européenne de psychologie appliquée*, 49, 325-342.
- Mousty, P., Leybaert, J., Alegria, J., Content, A. et Morais, J. (1994). BELEC. Batterie d'évaluation du langage écrit et de ses troubles. Dans J. Grégoire et B. Piérart (dir.), *Évaluer les troubles de la lecture. Les nouveaux modèles théoriques et leurs implications diagnostiques* (p. 127-145). De Boeck.
- Nagy, W. E. et Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19(3), 304-330. <https://doi.org/10.2307/747823>
- Nagy, W. E., Anderson, R. C. et Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270. <https://doi.org/10.3102/00028312024002237>
- Nagy, W. E., Anderson, R. C., Schommer, M., Scott, J. A. et Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24(3), 262-282. <https://doi.org/10.2307/747770>
- Nagy, W. E., Carlisle, J. F. et Goodwin, A. P. (2013, 2013). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, 47(1), 3-12. <https://doi.org/10.1177/0022219413509967>
- Nagy, W. E., Herman, P. A. et Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253. <https://doi.org/10.2307/747758>
- Nagy, W. E. et Scott, J. A. (2000). Vocabulary processes. Dans M. Kamil, P. Mosenthal, P. D. Pearson et R. Barr (dir.), *Handbook of reading research* (vol. III, p. 269-284). Lawrence Erlbaum Associates.
- Nagy, W. E. et Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91-108. <https://doi.org/10.1002/RRQ.011>
- Nation, I. S. P. (2001). Knowing a word. Dans I. S. P. Nation (dir.), *Learning Vocabulary in Another Language* (p. 23-59). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.1353/cml.2006.0049>
- New, B., Pallier, C., Ferrang, L. et Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'année psychologique*, 101, 447-462.
- Nippold, M. A. et Sun, L. (2008). Knowledge of morphologically complex words: A developmental study of older children and young adolescents. *Language, Speech, and Hearing Services in Schools*, 39(3), 365-373. [https://doi.org/10.1044/0161-1461\(2008/034\)](https://doi.org/10.1044/0161-1461(2008/034))
- Observatoire National de la Lecture. (2000). *Maîtriser la lecture*. Odile Jacob.
- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Results (Volume 1): What students know and can do*. PISA, OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Organisation for Economic Co-operation and Development. (s.d.). *PISA Frequently Asked Questions (FAQs)*. OECD. Programme for International Student Assessment (PISA). <https://www.oecd.org/en/about/programmes/pisa/pisa-frequently-asked-questions-faqs.html>
- Orme, J. E. (1966). Hypothetically true norms for the progressive matrices tests. *Human Development*, 9(4), 222-230. <https://doi.org/10.1159/000270425>

- Paap, K. R., Newsome, S. L., McDonald, J. E. et Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89(5), 573-594. <https://doi.org/10.1037/0033-295X.89.5.573>
- Padak, N., Newton, E., Rasinski, T. V. et Newton, R. M. (2008). Getting to the root of word study: Teaching Latin and Greek word roots in elementary and middle grades. Dans A. E. Farstrup et S. J. Samuels (dir.), *What research has to say about vocabulary instruction* (p. 6-31). International Reading Association.
- Pagani, L. S., Fitzpatrick, C., Belleau, L. et Janosz, M. (2011). *Prédire la réussite scolaire des enfants en quatrième année à partir de leurs habiletés cognitives, comportementales et motrices à la maternelle* (vol. 6, Fascicule 1). Québec : Gouvernement du Québec, Institut de la statistique du Québec.
- Palumbo, A., Kramer-Vida, L. et Hunt, C. V. (2015). Teaching vocabulary and morphology in intermediate grades. *Preventing School Failure: Alternative Education for Children and Youth*, 59(2), 109-115. <https://doi.org/10.1080/1045988X.2013.850649>
- Pavard, B. (1983). Traitement perceptif des mots affixés: mise en évidence d'un contrôle cognitif. *L'année psychologique*, 83(2), 443-464. <https://doi.org/10.3406/psy.1983.28476>
- Pearson. (2007). *Raven's standard progressive matrices : Evidence of reliability and validity*. NCS Pearson.
- Pearson, P. D., Hiebert, E. H. et Kamil, M. L. (2012). Vocabulary assessment. Making do with what we have while we create the tools we need. Dans E. J. Kame'enui et J. F. Baumann (dir.), *Vocabulary instruction. Research to practice*. (2^e éd., p. 231-255). The Guildford Press.
- Peereman, R., Lété, B. et Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme—phoneme mappings, and infralexicale and lexical units in child-directed written material. *Behavior Research Methods*, 39(3), 579-589. <https://doi.org/10.3758/BF03193029>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. <https://doi.org/10.1080/10888430701530730>
- 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>
- Phal, A. (1971). *Vocabulaire général d'orientation scientifique (V.G.O.S.) - Part du lexique commun dans l'expression scientifique*. Didier.
- Picoche, J. (2015). *Dictionnaire d'étymologie du français*. Dictionnaires Le Robert-Sejer.
- Pikulski, J. J. et Templeton, S. (2004). Teaching and developing vocabulary : Key to long-term reading success. *Current Research in Reading/Language Arts*, 1-12.
- Plaut, D. C. et Gonnerman, L. M. (2000). Are non- semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15(4-5), 445-485. <https://doi.org/10.1080/01690960050119661>
- Poehner, M. E. et Lantolf, J. P. (2010). Vygotsky's teaching-assessment dialectic and L2 education: The case for dynamic assessment. *Mind, Culture, and Activity*, 17(4), 312-330. <https://doi.org/10.1080/10749030903338509>
- Polguère, A. (2016). *Lexicologie et sémantique lexicale : notions fondamentales* (3^e éd.). Les Presses de l'Université de Montréal. <https://doi.org/https://doi.org/10.2307/j.ctv69t90p>
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513-536. <https://doi.org/10.1111/1467-9922.00193>
- Quémart, P. et Casalis, S. (2015). Visual processing of derivational morphology in children with developmental dyslexia: Insights from masked priming. *Applied Psycholinguistics*, 36, 345-376. <https://doi.org/10.1017/S014271641300026X>

- Quémart, P., Casalis, S. et Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology*, 109, 478-496. <https://doi.org/10.1016/j.jecp.2011.02.008>
- Quémart, P., Janiot, M. et Casalis, S. (2012). Les représentations orthographiques et morphologiques activées lors de la reconnaissance des mots écrits chez les dyslexiques [Manuscrit non publié]. *Cahiers Romains de Sciences cognitives*.
- Rasinski, T. V., Padak, N. et Newton, J. (2017). The roots of comprehension: Studying Latin and Greek word origins makes vocabulary instruction resonate. *Educational Leadership*, 74(5), 41-45.
- Rasinski, T. V., Padak, N., Newton, J. et Newton, E. (2011). The Latin-Greek connection building vocabulary through morphological study. *The Reading Teacher*, 65(2), 133-141. <https://doi.org/10.1002/TRTR.01015>
- Rasinski, T. V., Padak, N., Newton, R. M. et Newton, E. (2007). *Research-based curriculum building vocabulary from word roots levels 3-8*. Beach City Press.
- Rasinski, T. V., Padak, N., Newton, R. M. et Newton, E. (2008). *Greek and Latin roots: Keys to building vocabulary*. Shell Education.
- Raven, J. (1983). *Standard progressive matrices*. H.K. Lewis.
- Reichle, E. D. et Perfetti, C. A. (2003). Morphology in word identification: A word- experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, 7(3), 219-237. https://doi.org/10.1207/S1532799XSSR0703_2
- Rey-Debove, J. (1984). Le domaine de la morphologie lexicale. *Cahiers de lexicologie*, 2(45), 3-19. <https://doi.org/10.15122/isbn.978-2-8124-4298-8.p.0005>
- Rey, A. et Verdier, F. (2017). *Le Petit Robert*. Le Robert.
- Royle, P., Steinhauer, K., Dessureault, É., Herbay, A. C. et Brambati, S. M. (2019). Aging and language: Maintenance of morphological representations in older adults. *Frontiers in Communication*, 4(16), <https://doi.org/10.3389/fcomm.2019.00016>
- Rubin, G. S., Becker, C. A. et Freeman, R. H. (1979). Morphological structure and its effect on visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 757-797. [https://doi.org/10.1016/S0022-5371\(79\)90467-5](https://doi.org/10.1016/S0022-5371(79)90467-5)
- Sandra, D. (1991). La reconnaissance visuelle des mots composés. Dans R. Kolinsky, J. Morais et J. Segui (dir.), *La Reconnaissance des mots dans les différentes modalités sensorielles : Etudes de psycholinguistique cognitive* (p. 207-236). Presses Universitaires de France.
- Statistique Canada. (2012). Enfants et jeunes. Dans *Annuaire du Canada 2012* (p. 180-193). Statistique Canada.
- Schreuder, R. et Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, 33, 357-375. <https://doi.org/10.1006/jmla.1994.1017>
- Schreuder, R. et Baayen, R. H. (1995). Modeling morphological processing. Dans L. B. Feldman (dir.), *Morphological aspects of language processing* (p. 131-154). Erlbaum.
- Seidenberg, M. S. et Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4(9), 353-361. [https://doi.org/10.1016/S1364-6613\(00\)01515-1](https://doi.org/10.1016/S1364-6613(00)01515-1)
- Seidenberg, M. S. et McClelland, J. L. (1989a). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568. <https://doi.org/10.1037//0033-295X.96.4.523>
- Seidenberg, M. S. et McClelland, J. L. (1989b). Visual word recognition and pronunciation: A computational model of acquisition, skilled performance, and dyslexia. Dans A. M. Galaburda (dir.), *From neurons to reading* (p. 255-305). The MIT Press.
- Seymour, P. H. K. (1999). Cognitive architecture of early reading. Dans I. Lundberg, F. E. Tønnessen et I. Austad (dir.), *Dyslexia: Advances in Theory and Practice* (p. 59-73). Springer Netherlands.

- Seymour, P. H. K. (2007). 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 Perspectives* (p. 1-24). Lawrence Erlbaum Associates.
- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453. <https://doi.org/10.3102/00346543075003417>
- Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education*, 6(1), 5-10. <https://doi.org/10.1080/02568549109594817>
- Snow, C. E., Griffin, P. et Burns, S. M. (2005). *Knowledge to support the teaching of reading: preparing teachers for a changing world*. Jossey-Bass.
- Solomyak, O. et Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, 22(9), 2042-2057. <https://doi.org/10.1162/jocn.2009.21296>
- Sommet, N. et Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, 30(1), 203-218. <https://doi.org/10.5334/irsp.90>
- Sprenger-Charolles, L. et Casalis, S. (1996). *Lire. Lecture et écriture: acquisition et troubles du développement*. Presses universitaires de France.
- Stanké, B., Le Mené, M., Rezzonico, S., Moreau, A. C., Dumais, C., Robidoux, J., Dault, C. et Royle, P. (2019). ÉQOL : Une nouvelle base de données québécoise du lexique scolaire du primaire comportant une échelle d'acquisition de l'orthographe lexicale. *CORPUS*, 19. <https://doi.org/10.4000/corpus.3818>
- Statistique Canada. (2017). *Québec [Province] et Canada [Pays] (tableau). Profil du recensement*. (Recensement de 2016, produit n° 98-316-X2016001). <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=F>
- Steinhauer, K. et Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, 83, 62-65.
- Stites, M. C., Federmeier, K. D. et Christianson, K. (2016). Do morphemes matter when reading compound words with transposed letters? Evidence from eye-tracking and event-related potentials. *Language, Cognition and Neuroscience*, 31(10), 1299-1319. <https://doi.org/10.1080/23273798.2016.1212082>
- Stockwell, R. et Minkova, D. (2001). *English words: History and structure*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791161>
- Swanborn, M. S. L. et de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261-285. <https://doi.org/10.3102/00346543069003261>
- Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 262-272. <https://doi.org/10.3758/BF03197599>
- Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9(3), 271-294. <https://doi.org/10.1080/01690969408402120>
- Taft, M. et Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638-647. [https://doi.org/10.1016/S0022-5371\(75\)80051-X](https://doi.org/10.1016/S0022-5371(75)80051-X)
- Templeton, S. (1983). Using the spelling/meaning connection to develop word knowledge in older students. *Journal of Reading*, 27(1), 8-14.
- Templeton, S. (2004). The vocabulary-spelling connection: Orthographic development and morphological knowledge at the intermediate grades and beyond. Dans J. F. Baumann et E. J. Kame'enui (dir.), *Vocabulary instruction : Research to practice* (p. 118-138). Guilford Press.
- Thomas, M. S. C. et Ansari, D. (2020). Educational neuroscience: Why is neuroscience relevant to education?. Dans M. S. C. Thomas, D. Mareschal, et I. Dumontheil (dir.), *Educational*

- neuroscience : Development across the life span.* (p. 3-22). Routledge. <https://doi.org/10.4324/9781003016830>
- Tremblay, O. (2015). L'analyse de corpus pour l'étude du lexique en classe de français. *Correspondance*, 21(1), 1-5.
- Tremblay, O., Drouin, P. et Saidane, R. (2023). Constitution d'un corpus de manuels scolaires en usage à l'école primaire pour l'élaboration d'une liste de vocabulaire transdisciplinaire. *Corpus*, 24. <https://doi.org/10.4000/corpus.7585>
- Tremblay, O., Saidane, R. et Drouin, P. (2018). Une liste de vocabulaire transdisciplinaire pour les élèves francophones du deuxième cycle du primaire. Dans M.-P. Jacques et A. Tutin (dir.), *Lexique transversal et formules discursives des sciences humaines* (p. 163-189). ISTE Editions.
- Tutin, A. (2008). Sémantique lexicale et corpus: l'étude du lexique transdisciplinaire des écrits scientifiques. *Lublin studies in modern languages and literature*, 32, 242-260.
- Ullman, M. T. (2001a). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105-122. <https://doi.org/10.1017/S1366728901000220>
- Ullman, M. T. (2001b). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2, 717-726. <https://doi.org/10.1038/35094573>
- Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovotto, C., Pancheva, R., Ozawa, K., Mordecai, K. et Maki, P. (2002). Sex differences in the neurocognition of language. *Brain and Language*, 83, 141-143.
- Ullman, M. T., Miranda, R. A. et Travers, M. L. (2007). Sex Differences in the Neurocognition of Language. Dans J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. P. Herman et E. Young (dir.), *Sex Differences in the Brain: From Genes to Behavior* (p. 291-310). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195311587.003.0015>.
- Vanier, M. (1991). *Matrices progressives standard. J.C. Raven, J.H. Court et J. Raven (1re édition 1938; 6e édition 1986)*. <https://docplayer.fr/40366913-Matrices-progressives-standard.html>
- Vergara-Martínez, M., Duñabeitia, J. A., Laka, I. et Carreiras, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Research*, 1257, 53-64. <https://doi.org/10.1016/j.brainres.2008.12.040>
- Wall, A. (2016). From the roots up: Building students' vocabulary knowledge through Latin and Greek. *Delta Kappa Gamma Bulletin*, 82(4), 1-12.
- Whissell-Turner, K. et Fejzo, A. (2021). Knowledge of Greek and Latin roots is related to reading comprehension. *The Canadian Journal of Applied Linguistics*, 24(3), 61-78. <https://doi.org/10.37213/cjal.2021.30473>
- Whissell-Turner, K., Fejzo, A. et Saidane, R. (2022). Facteurs linguistiques liés à la connaissance des racines latines et grecques: origine et fréquence. *Bellaterra Journal of Teaching & Learning Language & Literature*, 15(3), e1069. <https://doi.org/10.5565/rev/jtl3.1069>
- White, T. G., Power, M. A. et White, S. (1989, 1989). Morphological analysis: Implications for teaching and understanding vocabulary growth. *Reading Research Quarterly*, 24(3), 283-304. <https://doi.org/10.2307/747771>
- Wise, B. W., Olson, R. K. et Treiman, R. (1990). Subsyllabic units in computerised reading instruction : Onset-rime vs. postvowel segmentation. *Journal of Experimental Child Psychology*, 49(1), 1-19. [https://doi.org/10.1016/0022-0965\(90\)90046-b](https://doi.org/10.1016/0022-0965(90)90046-b)
- Worthington, D. et Nation, I. S. P. (1996). Using texts to sequence the introduction of new vocabulary in an EAP course. *RELC Journal*, 27(2), 1-11. <https://doi.org/10.1177/003368829602700201>
- Yoncheva, Y. N., Blau, V. C., Maurer, U. et McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental neuropsychology*, 35(4), 423-445. <https://doi.org/10.1080/87565641.2010.480918>

Zoski, J. L., Nellenbach, K. M. et Erickson, K. A. (2018). Using morphological strategies to help adolescents decode, spell, and comprehend big words in science. *Communication Disorders Quarterly*, 40(1), 57-64. <https://doi.org/10.1177/1525740117752636>