



HAL
open science

Suivi longitudinal des performances en lecture-écriture du CP au CM2 : analyse des liens entre lecture et orthographe et des relations entre orthographe lexicale et sublexicale

Carole Hanner

► **To cite this version:**

Carole Hanner. Suivi longitudinal des performances en lecture-écriture du CP au CM2 : analyse des liens entre lecture et orthographe et des relations entre orthographe lexicale et sublexicale. Education. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALH032 . tel-04255644

HAL Id: tel-04255644

<https://theses.hal.science/tel-04255644>

Submitted on 24 Oct 2023

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : SHPT - Sciences de l'homme, du Politique et du Territoire

Spécialité : Sciences de l'Éducation

Unité de recherche : Laboratoire de Recherche sur les Apprentissages en Contexte

**Suivi longitudinal des performances en lecture-écriture du CP au CM2
: analyse des liens entre lecture et orthographe et des relations entre
orthographe lexicale et sublexicale**

**Longitudinal study on reading and writing performances throughout
elementary school: an analysis of the relationships between reading
and spelling and between lexical and sublexical orthographic
knowledge**

Présentée par :

Carole HANNER

Direction de thèse :

Pascal BRESSOUX

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Marie-Line BOSSE

PROFESSEURE DES UNIVERSITES, Université Grenoble Alpes

Co-directrice de thèse

Rapporteurs :

Marie-France MORIN

PROFESSEUR, Université de Sherbrooke

Catherine MARTINET

PROFESSEUR, HEP Vaud

Thèse soutenue publiquement le **14 décembre 2022**, devant le jury composé de :

PASCAL BRESSOUX

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Marie-Line BOSSE

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Co-directrice de thèse

Marie-France MORIN

PROFESSEUR, Université de Sherbrooke

Rapporteuse

Catherine MARTINET

PROFESSEUR, HEP Vaud

Rapporteuse

Catherine BRISSAUD

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Présidente

Michel FAYOL

PROFESSEUR EMERITE, Université Clermont Auvergne

Examinateur



Remerciements

Je remercie en premier lieu les membres du jury, qui ont accepté de lire et de questionner ce travail : Catherine Brissaud, Michel Fayol, Catherine Martinet et Marie-France Morin. Merci à vous d'avoir pris le temps de participer à la discussion finale de ce travail.

Je remercie Marie-Line Bosse et Pascal Bressoux pour leur encadrement pendant ces six années : merci pour vos éclairages, votre compréhension, votre grande capacité d'adaptation et votre aide durant tout ce temps : votre bienveillance est sans nul doute un des facteurs qui m'aura permis d'arriver au bout de cette aventure !

Je remercie toutes les personnes, et elles furent nombreuses, impliquées dans le grand projet qu'était ACQUISLONGIT : ceux qui ont participé à l'élaboration des passations, particulièrement Maryse Bianco et Laurent Lima ; bien sûr, tous les acteurs impliqués dans la circonscription du Haut-Grésivaudan : les IEN Messieurs Luc Sindirian et Rémy Ducouset, les conseillers pédagogiques, les écoles, directeurs et directrices, et les enseignants qui nous ont ouvert les portes de leur établissement et de leurs classes, année après année, pendant plusieurs semaines : merci à vous pour votre accueil, votre adaptabilité et votre disponibilité ! Merci aux familles et surtout aux élèves, sans qui rien de tout cela n'aurait été possible. Quel plaisir d'arriver avec nos cartons remplis de livrets et de les voir venir nous saluer, pleins d'enthousiasme (pour la plupart), en s'écriant « c'est Longit ! ». Merci enfin aux nombreux vacataires qui ont passé de longues semaines à parcourir chaque année la vallée du Grésivaudan, à transporter des cartons, merci pour nos repas avec vue sur la piste d'atterrissage des parapentes, pour les cafés-tricot entre deux passations ; ces semaines étaient intenses et épuisantes, mais c'était un plaisir de les passer avec vous. Merci à Sandrine Lescure, notre super gestionnaire, pour son aide et sa réactivité dans la préparation de ces périodes de passation. Pour Longit comme pour tout le reste, votre aide aura été précieuse !

Maintenant que la fin de cette étape arrive, je me sens comme Frodo après son retour du Mordor. Ce long chemin parcouru n'aurait sans doute pas été possible sans ma Communauté de la Thèse ! Je remercie tous ceux qui ont eu la chance (ou pas...) de partager un bureau avec moi et mon incapacité à rester en place et dans le calme : Gremlin, Laurent, Olga, Hoa, Fernando, Jérémy, Dany, Fanny, Marieke ; à ceux avec qui j'ai pu partager des discussions, des pauses, des visios ou des congrès : Manon, Mathilde, Valentina, Ludivine, Irène, Ophélie, Séverine, Sandrine, Marina, Erika, Emilie, Marlène, Céline, Romain... Aux collègues docteurs

(ou plus !), pour leur écoute et leurs conseils : Anne-Laure, Benjamin, Camille, Elisa, Anissa, Alexis, Salomé, Anna, Vicky, Laurent, Nadia au pluriel... Merci à tous, merci pour tous ces moments passés ensemble, que ça soit autour d'un café, entre deux portes de salles de cours, pour les apéros au Martins, les brunchs ...

Durant ce voyage épique, j'avais emmené avec moi les souvenirs de tous ceux qui me sont chers, et qui, de près ou de loin, m'ont soutenue, sans faillir, sans comprendre parfois, mais surtout sans jugement. A mes amis, ma famille, merci pour tout, pour votre soutien inconditionnel et votre amour, dans ce voyage si particulier. Dans les moments de doute, et ils furent nombreux, savoir que vous croyiez en moi sans doute plus que moi-même, a été un réel moteur pour me donner la force et le courage de continuer mon chemin. Merci à ma mère, mon soutien de toujours, à ma fratrie adorée, à mon Ananas que j'aime *jusqu'à la Lune*, et à Matthieu, pour qui les derniers mois ont dû être bien difficiles... C'est la fin d'une étape pour vous tous aussi !

Enfin, *last but definitely not least*, à mon père, celui qui me guide de là-haut en étant l'étoile la plus brillante du ciel, même dans la nuit la plus sombre : je sais que tu me poussais pour que je réalise cet objectif qui me tenait à cœur tant sur le plan personnel que professionnel : ce travail est pour toi, *Daddy* !

« *Not all those who wander are lost* »

J.R.R. Tolkien

Résumé

Ce travail de thèse s'intéresse à l'évolution des habiletés de lecture et d'écriture sur la période de l'école élémentaire en France, et fait partie du projet ACQUISLONGIT soutenu par l'Institut Carnot de l'Éducation de la région Auvergne Rhône-Alpes. Il a pour objectif dans un premier temps d'étudier sur un temps long et en français le lien entretenu par les habiletés de lecture et d'écriture ; en effet, si certaines études font état d'un lien unilatéral de la lecture vers l'écriture en début d'enseignement formel (e.g. Chun et al., 2022 ; Deacon et al., 2012 ; Schaars et al., 2017), il n'y a pas, à notre connaissance, d'étude longitudinale en langue française qui ait mesuré l'évolution de ce lien sur l'ensemble de l'école élémentaire. Nous avons donc testé 561 élèves, de leur entrée au CP à la fin du CM2, afin d'étudier l'évolution de leurs performances en orthographe lexicale, lecture de mots isolés et lecture de mots en contexte. Les résultats du modèle *cross-lagged* ont montré un lien bilatéral entre les trois variables étudiées sur tous les temps de mesure, même si le poids des habiletés de lecture restait plus fort que celui de l'écriture. Les résultats montrent également une distinction entre lecture de mots isolés et lecture de mots en contexte.

Dans un second temps, nous nous sommes intéressés plus spécifiquement au développement de la compétence orthographique, en étudiant les prédicteurs en début de scolarité de l'orthographe lexicale (via des tâches d'écriture sous la dictée ou de reconnaissance orthographique) et de l'orthographe sublexicale, c'est-à-dire de la sensibilité aux régularités graphotactiques de la langue (via des tâches de choix forcé de pseudomots). Des analyses multiniveaux ont montré une différence majeure entre orthographe lexicale et sublexicale : l'orthographe lexicale est en partie influencée par le niveau de la classe (comme l'enseignement explicite des structures orthographiques), tandis que ce niveau n'a pas d'effet sur l'orthographe sublexicale. De plus, si certaines habiletés communes ont pu être mises en avant par les résultats (dictée de pseudomots, dictée de mots), il reste difficile d'évaluer la nature du lien entre orthographe lexicale et sublexicale, notamment car le pouvoir explicatif des régressions concernant l'orthographe sublexicale est très faible, laissant donc penser que cette habileté n'est que peu influencée par les prédicteurs connus de la lecture et de l'écriture.

Abstract

Our main objective is to investigate the evolution of reading and writing abilities throughout elementary school in France. Our first aim is to study the relationships between reading and spelling skills in French, in a longitudinal way. Actually, although some studies show a unilateral relationship from reading to spelling in the early years of formal education (e.g. Chun et al., 2022; Deacon et al., 2012; Schaars et al., 2017), there has not been, to our knowledge, a french longitudinal study on the evolution of this relationship during the whole elementary school years. We tested 561 children, from the beginning of first Grade (french equivalent) to the end of fifth Grade, in order to study the evolution of their performances on word-specific orthographic knowledge, isolated word reading and text reading fluency. A cross-lagged model showed a bilateral relationship between the three variables on all measured times, even though the weight of reading skills remained stronger than the weight of spelling skills. Results also showed a distinction between isolated word reading and text reading fluency.

Our second aim is to investigate the orthographic component, by studying the predictors of word-specific orthographic knowledge and sublexical orthographic knowledge at the beginning of formal education, using dictation and recognition tasks for the former and forced-choice tasks for the latter. Multilevel analysis showed a major difference between lexical and sublexical orthographic knowledge: lexical orthographic knowledge is partly influenced by class-level (e.g. explicit instruction on orthographic structures), whereas it has no effect on sublexical orthographic knowledge. Moreover, if some shared abilities have been outlined by our results (nonword dictation, word dictation), investigating the nature of the relationship between lexical and sublexical orthographic knowledge remains complex, especially because regression analysis on sublexical orthographic knowledge only explain very little of the global variance, which leads us to think that this ability is not much influenced by known predictors of reading and spelling.

Sommaire

Remerciements	2
Résumé	5
Abstract	7
Liste des tableaux	11
Liste des figures	13
Glossaire.....	15
Introduction	17
PREMIÈRE PARTIE : THÉORIE.....	23
Chapitre 1 : Généralités sur l’acquisition de la lecture-écriture et spécificités de la langue française	25
1.1 Définitions	25
1.2 Les différents systèmes d’écriture	26
1.3 Comparaison des langues alphabétiques européennes	28
1.4 Description de l’orthographe française.....	29
1.4.1 Le code alphabétique, ou les correspondances graphème-phonème	30
1.4.2 Les connaissances orthographiques	31
1.4.2.1 Les connaissances orthographiques lexicales	31
1.4.2.2 Les connaissances orthographiques sublexicales.....	31
1.4.3 Principe morphographique : les connaissances morphologiques.....	33
1.4.3.1 La morphologie dérivationnelle	33
1.4.3.2 La morphologie flexionnelle.....	34
Chapitre 2 : les modèles théoriques	37
2.1 Modèles experts de la lecture et de l’orthographe	37
2.2 Développement et acquisition de la lecture et de l’orthographe	41
2.2.1 Universalité de la lecture et de l’orthographe	41

2.2.2	Les modèles développementaux.....	43
2.2.3	Les modèles interactifs.....	44
2.2.4	Théories spécifiques au développement orthographique.....	47
Chapitre 3 : les facteurs de l'environnement scolaire impliqués dans l'apprentissage de la lecture et de l'écriture.....		51
3.1	Des facteurs liés au type d'enseignement.....	51
3.1.1	Effet-maitre et effet-classe.....	51
3.1.2	Effet des modalités d'enseignement.....	52
3.2	Les facteurs liés aux méthodes d'enseignement du code en début d'apprentissage..	54
3.3	Les facteurs liés à l'orthographe.....	56
3.3.1	Les orthographe approchées en période préscolaire.....	56
3.3.2	Le cas de l'écriture sous la dictée.....	57
3.3.3	L'effet du temps passé sur les activités liées à l'écrit.....	59
Chapitre 4 : les liens entre lecture et écriture.....		63
4.1	Les prédictors de la lecture et de l'écriture à l'école élémentaire.....	63
4.2	Liens entre lecture et écriture pendant l'apprentissage formel.....	70
Synthèse théorique.....		77
SECONDE PARTIE : EMPIRIE.....		81
Chapitre 5 : Méthode générale.....		83
5.1	Méthode.....	83
5.1.1	Participants.....	83
5.1.2	Matériel.....	85
5.1.3	Compétences générales.....	86
5.1.4	Langage oral.....	87
5.1.5	Code écrit.....	88
5.1.6	Orthographe.....	90
5.1.7	Compréhension écrite.....	92

5.2	Procédure	93
Chapitre 6 : Etude 1 : évolution longitudinale du lien lecture-écriture		95
6.1	Hypothèses.....	95
6.2	Matériel et méthode	95
6.3	Résultats.....	96
6.3.1	Résultats descriptifs de l'évolution des performances en lecture et écriture du CP au CM2	97
6.3.2	Corrélations entre les habiletés de lecture et d'orthographe	99
6.3.3	Résultats du modèle <i>cross-lagged</i> sur l'évolution de la lecture et de l'écriture du CE1 au CM2.....	103
6.4	Conclusion.....	107
Chapitre 7 : Etudes 2 et 3 : études du lien entre orthographe lexicale et sublexicale en début d'apprentissage formel.....		111
7.1	Hypothèse	111
7.2	Matériel et méthode	111
7.3	Résultats.....	112
7.3.1	Prédicteurs de l'orthographe sublexicale	113
7.3.2	Prédicteurs de l'orthographe lexicale	114
7.4	Conclusion	117
Chapitre 8 : Discussion générale.....		122
8.1	Synthèse.....	122
8.2	Limites	125
8.3	Perspectives	127
Conclusion.....		132
Bibliographie.....		134
Liste des annexes		149
Annexes		150

Liste des tableaux

Tableau 1. Nombre moyen d'erreurs à la dictée de Fénelon, en fonction de l'année du test et du niveau scolaire des élèves (Manesse & Cogis, 2007)	19
Tableau 2. Nombre moyen d'élèves par classe et écart-type pour chaque niveau de classe étudié	84
Tableau 3. Nombre d'élèves ayant participé à l'étude à chaque temps de mesure, du CP au CM2	84
Tableau 4. Synthèse des épreuves utilisées en fonction des années scolaires pour l'ensemble du suivi longitudinal.....	85
Tableau 5. Présentation des épreuves utilisées pour les études 1, 2 et 3 et temps de mesure auxquelles elles ont été testées	86
Tableau 6. Durée moyenne des épreuves collectives et individuelles à chaque temps de mesure.	94
Tableau 7. Nombre de mots correctement lus en une minute à l'épreuve de lecture de mots et de fluence de texte : moyennes et écart-types par année scolaire.	97
Tableau 8. Scores moyens obtenus à l'épreuve de dictée Le corbeau : score phonologique, score d'usage et score grammatical, du CE1 au CM2.....	98
Tableau 9. Corrélations de l'épreuve de lecture de mots isolés du CP au CM2.....	100
Tableau 10. Corrélations de la fluence de textes du CP au CM2.....	100
Tableau 11. Corrélations de l'orthographe d'usage du CE1 au CM2	101
Tableau 12. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CE1	101
Tableau 13. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CE2.....	101
Tableau 14. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CM1.....	102
Tableau 15. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CM2.....	102
Tableau 16. Analyse de régression montrant les prédicteurs en CP de l'orthographe sublexicale en CE2.	113
Tableau 17. Analyse de régression montrant les prédicteurs en CE1 de l'orthographe sublexicale en CE2.	114

Tableau 18. Analyse multiniveaux montrant les prédicteurs en CP de l'orthographe lexicale en CE2 (tâche de reconnaissance orthographique)	115
Tableau 19. Analyse multiniveaux montrant les prédicteurs en CE1 de l'orthographe lexicale en CE2 (tâche de reconnaissance orthographique)	116
Tableau 20. Analyse multiniveaux montrant les prédicteurs en CE1 de l'orthographe lexicale en CE2 (tâche de production sous forme d'écriture sous la dictée).....	117

Liste des figures

Figure 1. Modèle cross-lagged sur l'évolution des liens entre orthographe d'usage (USAGE), fluence en contexte (FLUENCE) et lecture de mots isolés (LEC_MOTS) du CE1 au CM2..	104
Figure 2. Modèle cross-lagged complet sur l'évolution des liens entre orthographe d'usage (USAGE), fluence en contexte (FLUENCE) et lecture de mots isolés (LEC_MOTS) du CE1 au CM2.....	174

Glossaire

Certains termes utilisés tout au long de ce manuscrit peuvent être compris de différentes manières. Nous présentons donc ici un glossaire qui répertorie l'ensemble des termes pouvant prêter à confusion, et le sens que nous allons leur attribuer tout au long de ce document.

Compréhension. Selon le modèle simple de la lecture (Gough & Tunmer, 1986), la compréhension est un des deux éléments fondamentaux pour la réussite de l'activité de lecture. La compréhension consiste en l'accès aux représentations mentales associées à ce qui est lu.

Décodage. Le décodage en lecture fait appel à la capacité de transformer un graphème en son phonème correspondant, permettant ainsi la lecture. La maîtrise de ces correspondances grapho-phonémiques est donc un prérequis pour effectuer un décodage adéquat.

Écriture. L'écriture peut être à la fois perçue, dans la littérature, comme faisant référence au geste moteur lié à la production écrite mais également comme l'activité de production écrite correcte, c'est-à-dire l'orthographe. Dans ce document, le terme « écriture » est à considérer comme la production d'un écrit respectant les normes orthographiques de la langue, en référence au terme anglais *spelling*.

Épellation. Principalement utilisé dans certains modèles théoriques de l'écriture, l'épellation réfère à la production d'un mot écrit. Ce terme est utilisé dans le modèle simple de l'écriture (Juel et al., 1986) ; en français, il peut être confondu avec l'épellation orale d'un mot ; nous préférons, pour nous référer à cette production écrite de mots, les termes d'orthographe ou d'écriture.

Fluence. La fluence de lecture consiste en une lecture rapide et de qualité. Dans ce travail, la fluence de lecture est mesurée selon un nombre de mots correctement lus par minute, à haute voix. Nous distinguerons deux types de fluence : la fluence de lecture de mots isolés, et la fluence de texte, ou lecture en contexte.

Grades. Les niveaux scolaires présentés dans ce document sont basés sur les niveaux scolaires français. Toutefois, lors de la présentation de certaines études anglo-saxonnes, nous mentionnerons les Grades auxquels elles se réfèrent pour la majorité d'entre elles : le Grade 1 correspondant au niveau du Cours Préparatoire français, le Grade 2 au CE1, le Grade 3 au CE2,

le Grade 4 au CM1, et le Grade 5 au CM2. La période de la maternelle est quant à elle dénommée, pour sa version anglo-saxonne, Kindergarten.

Identification de mots. L'identification de mots consiste à accéder à la forme orale d'un mot à partir de sa forme écrite, soit en utilisant le décodage, soit en reconnaissant le mot dans sa globalité.

Lecture. La lecture est une activité dont le but est la compréhension d'un message écrit via l'identification des unités qui le composent.

Littéracie. La littéracie correspond aux habiletés fondamentales de la lecture et de l'écriture utilisées par les enfants dès leur plus jeune âge pour comprendre et interagir avec le langage dans lequel ils évoluent.

Orthographe. L'orthographe correspond à la norme d'écriture d'un mot dans le langage auquel elle se réfère. Dans ce document, nous utiliserons le terme d'orthographe pour faire référence à la capacité à écrire correctement les mots, conformément à l'usage qui leur est conféré par le langage de référence.

Introduction

L'apprentissage de la lecture et de l'écriture est une étape fondamentale dans le parcours scolaire de chacun ; c'est donc tout naturellement que depuis plusieurs décennies les chercheurs s'intéressent à ces apprentissages : mise en place, évolution, perfectionnement... tous ces aspects sont étudiés de manière théorique et empirique afin de mieux comprendre leur fonctionnement.

L'étude de ces apprentissages majeurs permet évidemment une meilleure compréhension de cette période développementale, mais également une adaptation potentielle des méthodes d'enseignement au regard des apports de la recherche (MEN, 2019). Il semble donc nécessaire de s'intéresser plus spécifiquement à l'acquisition et au développement des compétences orthographiques du français chez l'élève, afin de mieux comprendre, et par extension de mieux accompagner cette acquisition complexe, qui s'avère être critique plus tard pour la vie professionnelle. En effet, la capacité à s'exprimer correctement à l'écrit pourrait faire partie des critères fondamentaux lors d'une embauche, d'autant plus depuis la crise sanitaire liée à la Covid-19 et le recours massif au télétravail, et donc à l'utilisation plus fréquente de canaux de communications passant par l'écrit. La maîtrise de l'orthographe, et de manière plus générale, de l'expression à l'écrit, revêt donc un intérêt qui dépasse largement le cadre scolaire ; c'est un outil nécessaire à l'intégration dans une société telle que la nôtre.

Plusieurs études, nationales ou internationales, permettent de suivre l'évolution des performances des élèves dans différents domaines scolaires au fil des ans. Ainsi, l'étude PIRLS (*Progress in International Reading Literacy Study*) évalue tous les cinq ans les capacités des élèves de nombreux pays (cinquante, pour la dernière étude réalisée en 2016) ; les élèves évalués sont dans leur quatrième année d'enseignement formel, ce qui correspond pour la France aux élèves de CM1 (MEN-DEPP, 2017) . L'évaluation porte sur la lecture (avec deux objectifs, d'une part lire pour accéder aux textes littéraires, de l'autre lire pour obtenir des informations) ainsi que sur le processus de compréhension. Les résultats de cette dernière étude montrent que, dans l'ensemble, les élèves français ont de moins bonnes performances que leurs voisins, le score global pour la France étant de 511 points, alors que celui de l'ensemble des pays de l'Union Européenne est de 540 et celui des pays de l'OCDE 541. De la même manière, les élèves français réussissent moins bien à chaque évaluation : par rapport à l'étude faite en 2001, il y avait en 2016 3% d'élèves en moins dans le groupe d'élèves ayant un score avancé,

7% de moins dans le groupe ayant un score élevé, 5% de moins pour le groupe intermédiaire et 1% de moins dans le groupe ayant un score bas, toutes ces différences par rapport à l'étude de 2001 étant statistiquement significatives.

Si les performances des élèves français en littéracie semblent donc diminuer avec le temps, qu'en est-il plus spécifiquement de leurs performances en lecture et en écriture ? Du côté des études nationales, l'évaluation CEDRE (Cycle d'Évaluations Disciplinaires Réalisées sur Échantillon) est proposée tous les cinq ans à des élèves de CM2 et se base sur les programmes officiels de l'Éducation Nationale. Pour la discipline qui nous intéresse ici, à savoir « maîtrise de la langue et compétences générales », on évalue le langage oral, la lecture et l'écriture, dans le but d'évaluer différents niveaux de compréhension. Andreu et al (2016) ont étudié les performances moyennes des élèves en 2003, 2009 et 2015 et leur répartition dans les différents groupes de niveaux d'acquisition (du niveau le plus faible soit le groupe <1 au niveau le plus fort, soit le groupe 5). D'après ces résultats, la performance globale des élèves reste sensiblement la même (score moyen de 250 en 2003, 251 en 2009 et 251 en 2015) mais l'écart-type diminue (il était de 50 en 2003 et de 43 en 2015), ce qui montre que les scores des élèves ont tendance à se regrouper autour de la moyenne. Ainsi, le pourcentage d'élèves présents dans le groupe <1 diminue (2.7% en 2003 contre 1% en 2015) ; mais d'un autre côté, le pourcentage d'élèves les plus performants diminue également (10% en 2003 contre 7.2% en 2015). Les auteurs relèvent également une forte disparité dans les performances en fonction du type d'établissement (privé sous contrat, public, REP) et du genre de l'élève, les garçons étant plus représentés dans les groupes de moindre performance que les filles.

Depuis la rentrée scolaire 2018 les élèves entrant en CP et CE1 sont eux aussi évalués, selon le dispositif Repères, dont l'objectif est de mettre en évidence les acquis des élèves afin de permettre aux enseignants d'avoir un état des lieux des compétences des élèves en français et en mathématiques. En ce qui concerne le français, l'évaluation s'intéresse aux champs suivants : les correspondances graphophonologiques, la reconnaissance de lettres, la phonologie, la compréhension orale, la lecture de l'écrit et la compréhension de l'écrit (DEPP-MENJ, 2019). Les résultats obtenus à la rentrée 2019 montrent globalement une stabilité des performances entre la session 2018 et la session 2019 en français ; pour les épreuves identiques sur les deux années, par exemple, en début de CP la reconnaissance de lettres (-2.1 points), la compréhension de mots lus par l'enseignant semblent avoir diminué (-3.1 points), tandis que la compréhension de phrases lues semble avoir légèrement augmenté (+1.9 point). En début de CE1, la lecture à voix haute d'un texte, la lecture à voix haute de mots et l'écriture de mots

semblent avoir augmenté tandis que la compréhension de mots lus semble avoir diminué (76.6% en 2019 contre 78% en 2018). L'effet du genre semble rester stable entre les deux sessions, puisque les filles sont globalement plus représentées dans le groupe le plus performant, pour le domaine du français. Enfin, en ce qui concerne le secteur où se situe l'établissement scolaire, on remarque des écarts de performance en faveur du secteur non prioritaire, comparativement aux établissements situés en REP ou REP+. Si les performances en elles-mêmes sont intéressantes à étudier, le contexte dans lequel les apprentissages ont lieu n'en est donc pas moins important.

D'un point de vue temporel, l'évolution des programmes et de la répartition du temps scolaire est également à prendre en compte lorsque l'on s'intéresse à l'évolution des performances des élèves. Cette comparaison des performances des élèves a justement été étudiée sur un temps très long, via la dictée proposée par Manesse et Cogis (2007). En 2005, les auteurs ont comparé les performances orthographiques des élèves de la classe de CM2 jusqu'à la classe de troisième en leur proposant une dictée qui avait déjà été proposée à des élèves de ces mêmes niveaux scolaires en 1987. La dictée était un extrait d'un livre de Fénelon et comptait 83 mots pour quatre phrases. Les erreurs étaient comptabilisées, et caractérisées selon deux critères : leur aspect descriptif (le type de faute), ainsi que leur aspect quantitatif, puisque la comparaison des deux cohortes se faisait principalement sur la base du nombre d'erreurs moyen de chaque niveau scolaire. Les cohortes se composaient de 3048 élèves en 1987, et de 2767 élèves en 2005.

Le Tableau 1 présente les scores comparés par classe, au demi-point près, des élèves évalués en 1987 et de ceux évalués en 2005. Étant donné que chaque erreur est comptabilisée, le score représente le nombre moyen de fautes pour chaque niveau scolaire. Ainsi, plus le score est élevé, plus il y a eu d'erreurs, et donc moins bonne a été la performance.

Tableau 1. Score d'erreurs à la dictée de Fénelon, en fonction de l'année du test et du niveau scolaire des élèves (Manesse & Cogis, 2007)

	CM2	6e	5e	4e	3e
1987	24,5	21	16,5	10,5	8
2005	36	31,5	27	23	17,5

De manière générale, les élèves évalués en 2005 ont donc fait plus d'erreurs que leurs confrères de 1987, et ce à tous les niveaux scolaires. On remarque toutefois que pour les deux

cohortes, le score d'erreurs diminue avec les années ; cependant, s'il diminue de deux tiers pour les élèves de 1987 entre le CM2 (24.5) et la troisième (8), il diminue seulement de moitié pour les élèves de 2005 (avec un score respectivement de 36 en CM2 et 17.5 en troisième). Ces différences sont à nuancer, selon les auteurs, principalement à cause de l'évolution du système scolaire. En effet, les redoublements, par exemple, étaient plus courants en 1987 qu'en 2005. De la même manière, les auteurs infèrent que le nombre d'heures passées à l'enseignement du français en primaire et au collège a diminué entre 1987 (cinq heures hebdomadaires au collège à l'époque) et 2005, et que la nature même du travail qui est fait sur l'orthographe a été modifiée, passant d'une pratique de type apprentissage par cœur et copie de mots, à un apprentissage plus intégré et contextualisé. En résumé, il semble donc qu'au fil des ans, les performances des élèves en orthographe aient diminué. Cependant, la comparaison par niveau scolaire est à interpréter avec modération compte tenu des changements dans les pratiques éducatives. De plus, la progression des élèves au fur et à mesure de leur scolarité reste visible, même si elle n'est pas aussi prononcée qu'auparavant.

Ainsi, les comparaisons nationales et internationales semblent montrer une diminution globale des performances des élèves concernant la lecture et l'écriture, que ça soit comparativement aux autres pays de l'OCDE ou comparativement aux élèves français d'années antérieures. Cependant, cette diminution diffère en fonction de l'habileté étudiée, et peut également s'expliquer par le contexte d'apprentissage (établissement, programmes, etc.). Il semble donc difficile de mettre en évidence avec certitude des critères spécifiques liés à cette baisse. Dans ce contexte, mieux comprendre les mécanismes de ces acquisitions et les facteurs agissant sur leur évolution durant la scolarité des élèves est un objectif fondamental. C'est donc sur ces aspects-là que va se concentrer notre travail.

Ce travail de thèse comporte deux parties. Dans la première partie seront apportés des éclairages théoriques sur les compétences en lecture et en écriture, à travers quatre chapitres. Dans le chapitre 1, nous présenterons tout d'abord l'état des connaissances générales sur l'acquisition de la lecture et de l'écriture et sur la spécificité de la langue française ; après avoir défini les termes de lecture et d'écriture, nous présenterons les différents systèmes d'écriture, puis la comparaison des langues alphabétiques européennes. Enfin, nous nous focaliserons sur

la description de l'orthographe française. Dans le chapitre 2, centré sur l'apprentissage de la lecture et de l'écriture, nous présenterons plusieurs modèles théoriques liés à ces apprentissages : des modèles experts de la lecture et de l'écriture, puis des modèles développementaux et interactifs. Nous nous intéresserons ensuite plus spécifiquement à l'orthographe, en présentant des modèles théoriques qui lui sont consacrés. Le chapitre 3 sera quant à lui dédié aux facteurs de l'environnement scolaire de l'élève, qui peuvent impacter ses performances scolaires ; nous présenterons ainsi des facteurs liés à l'environnement scolaire, puis des pratiques pédagogiques efficaces, qu'elles soient générales ou plus spécifiquement centrées sur l'enseignement de l'orthographe. Enfin, le chapitre 4 nous permettra de mettre en évidence les liens entre lecture et orthographe, via la présentation des principaux prédicteurs de ces deux habiletés, ainsi que leurs interactions et leur évolution, depuis le stade novice du lecteur-scripteur en début d'école élémentaire, jusqu'à son statut d'expert en fin d'école élémentaire.

La seconde partie de ce travail portera sur les études empiriques réalisées à travers quatre autres chapitres. Le chapitre 5 présentera la méthode ainsi que le matériel utilisés dans les différentes études ; le chapitre 6 présentera une analyse *cross-lagged* visant à mettre en évidence les liens entre lecture et écriture du CE1 au CM2 de manière plus globale. Le chapitre 7 proposera deux études portant plus spécifiquement sur l'orthographe lexicale et sublexicale : prédicteurs communs ou spécifiques, influence réciproque de ces deux compétences orthographiques du CP au CE2. Le chapitre 8 proposera une discussion générale de l'ensemble de ce manuscrit, ainsi que les limites et perspectives mises en lumière à la suite de ce travail.

PREMIÈRE PARTIE : THÉORIE

Chapitre 1 : Généralités sur l'acquisition de la lecture-écriture et spécificités de la langue française

1.1 Définitions

La lecture ainsi que l'écriture sont deux activités basées sur le traitement du langage écrit. Le système d'écriture utilisé pour ce faire est un système conventionnel et arbitraire de signes communs à une même langue. La juxtaposition de ces signes permet de former des unités ayant du sens (e.g., les mots, pour notre système alphabétique français), permettant la communication entre les personnes partageant cette langue commune.

Lire est une activité cognitive dont le but est de pouvoir comprendre le sens d'un message écrit à travers l'identification des unités qui le composent (mots, phrases...). Cette activité est complexe, et mobilise d'une part la capacité à transposer l'information écrite à l'oral (ou capacité d'identification), et d'autre part la capacité de compréhension du langage. Ce sont Gough et Tunmer (1986) qui ont proposé cette conception de l'activité de lecture, qu'ils appellent le modèle simple de la lecture (*Simple View of Reading*, ci-après SVR), et qu'ils représentent par l'équation $R = D \times C$ (Gough & Tunmer, 1986, p.7), où R correspond à la lecture (*Reading*), D au décodage (*Decoding*) et C à la compréhension (*Comprehension*). Ces deux capacités fondamentales pour la lecture que sont le décodage et la compréhension du langage seraient, pour les auteurs, à la fois distinctes et indissociables : pour être efficace et correcte, la lecture doit combiner ces deux capacités. Toutefois, elles peuvent être impactées de manière individuelle ; c'est notamment cette dernière spécificité qui rend ce cadre théorique intéressant, puisqu'il permet de rendre compte à la fois des performances en lecture, mais également des difficultés et troubles qui peuvent y être liés. Il serait ainsi possible de distinguer quatre profils de lecteurs : les lecteurs ayant à la fois des capacités de décodage et de compréhension correctes ; ceux ayant des capacités de décodage faible, mais une compréhension correcte (e.g. Snowling, 2014) ; ceux ayant des capacités de décodage et de compréhension faibles ; et enfin les lecteurs ayant des capacités de décodage correctes, mais une faible compréhension (e.g. Bianco, 2015 ; Grigorenko et al., 2003). Cette vision de la lecture permet l'utilisation de ce cadre théorique dans différentes langues ; ce dernier n'est toutefois pas suffisant, selon certains auteurs (e.g. Castles et al, 2018) pour rendre compte de

l'ensemble des mécanismes impliqués dans l'apprentissage de la lecture (voir Bianco, 2019). Dans le cadre de ce travail de thèse, nous nous attacherons plus spécifiquement à étudier la question de l'identification, et non celle des capacités de compréhension.

L'écriture, quant à elle, correspond à l'action de représenter un message oral, grâce à l'utilisation de signes arbitraires (en français, les lettres qui composent les mots) communs à une même langue. L'acte d'écrire, tout comme l'acte de lire, est une activité cognitive complexe car mobilisant plusieurs capacités simultanément : motricité fine, mémoire, capacités attentionnelles et visuo-spatiales, compréhension générale, monitoring, etc. En se basant sur le modèle SVR, Juel et al. (1986) proposent le modèle simple de l'écriture (*Simple View of Writing*) : dans ce modèle, l'écriture est le produit de deux habiletés clés pour cet apprentissage, à savoir l'épellation (*spelling*) et l'idéation (*ideation*, qui correspond à la génération et à l'organisation des idées) ; on obtient donc l'équation $W = S \times I$, pour reprendre la formule du SVR. De la même manière que pour le modèle simple de la lecture, les deux habiletés qui composent le modèle simple de l'écriture nécessitent d'interagir afin de produire une écriture performante. Elles sont à la fois indissociables pour que cette production écrite soit réalisable efficacement, tout en étant distinguables. Selon ce modèle, l'épellation comprend le code orthographique, commun au décodage du SVR. Lecture et écriture ne sont donc pas, contrairement à ce qui a été envisagé pendant longtemps, des compétences miroirs : l'activité d'écriture, en effet, demande plus de connaissances que l'activité de lecture, et, comme nous le verrons ultérieurement, ces deux habiletés sont dissociables (Perfetti, Rieben & Fayol, 1997).

Les activités de lecture et d'écriture sont pratiquées dans toutes les langues disposant d'un système d'écriture. Cependant, la diversité des systèmes d'écriture est très importante, et peut donc engendrer des différences majeures dans la pratique de la lecture et de l'écriture.

1.2 Les différents systèmes d'écriture

Il s'agit tout d'abord de distinguer clairement les différents systèmes d'écriture (voir Joyce & Borgwaldt, 2011, pour une revue) : les systèmes logographiques utilisent comme unité de langage le morphème, se basant donc sur une unité de sens ; les *kanjis* japonais sont un exemple de système logographique. L'autre système d'écriture japonais se base sur les *kana*,

qui correspondent quant à eux à des syllabes, c'est donc un exemple de système syllabique. Le système d'écriture abjad correspond à un alphabet consonantique, où la plus petite unité se compose d'une consonne à laquelle est implicitement liée une voyelle non écrite, pour y associer la prononciation correcte (comme l'hébreu ou l'arabe par exemple). Les systèmes alphasyllabaires, comme le thaï ou le khmer, se basent sur des unités écrites composées à la fois d'une consonne et d'une voyelle. Enfin, les systèmes alphabétiques ont pour unité le phonème, codé via l'utilisation de lettres (voyelles ou consonnes). La langue latine, d'où le français est issu, est fondée sur un système alphabétique. Perfetti et Harris (2013), après avoir comparé différents systèmes d'écriture, mettent en avant deux points communs à l'ensemble des systèmes : d'une part, le fait qu'il existe une correspondance entre une langue et son écriture, et d'autre part les contraintes imposées par le système écrit de la langue. Pour les auteurs, ces deux points forment les bases communes aux différentes langues et permettent le développement des habiletés de lecture et d'écriture.

Le système alphabétique repose donc sur l'association entre un graphème pour le versant écrit, et un phonème pour le versant oral. Par conséquent, apprendre à lire signifie en premier lieu savoir que, pour les systèmes alphabétiques comme le français, le graphème est un symbole auquel correspond un phonème particulier : c'est le principe alphabétique. La correspondance graphème-phonème (pour une activité de lecture) ainsi que la correspondance phonème-graphème (pour une activité d'écriture de type dictée par exemple) permettent d'établir l'ensemble des possibilités d'écriture ou de prononciation qui composent une langue.

En ce qui concerne les langues reposant sur un système alphabétique, ces correspondances entre phonèmes et graphèmes, ou graphèmes et phonèmes permettent d'établir la notion de transparence, ou au contraire, d'opacité, d'une langue alphabétique donnée. Une langue dite transparente sera constituée d'un ensemble de correspondances graphème-phonème et phonème-graphème consistantes : à chaque graphème correspondra un seul phonème donné, et inversement. *A contrario*, plus il y a de phonèmes différents possibles pour un même graphème, ou à l'inverse, de graphèmes différents possibles pour un même phonème, plus on considérera la langue en question comme étant opaque. Si le continuum allant de la transparence à l'opacité concerne le système orthographique sur un plan plutôt général, la notion de consistance, ou à l'inverse, d'inconsistance est calculée à partir du nombre de graphèmes possibles pour écrire un même phonème, ou inversement pour la correspondance grapho-phonémique (Peereman & Content, 1999).

Ainsi, des langues basées sur un même système d'écriture (comme le système alphabétique, par exemple) et issues d'une même famille de langue pourront différer de manière importante dans leur orthographe, ce qui rendra l'apprentissage de la lecture et de l'écriture pour chacune de ces langues bien différent. Cette distinction sur l'opacité de la langue et le caractère consistant ou non de son orthographe au sein d'un même système d'écriture voire d'une même famille de langues permet de mettre en évidence les spécificités de chaque système d'écriture, justifiant ainsi la nécessité de les étudier de manière individuelle, en plus de leur étude collective et/ou comparative.

1.3 Comparaison des langues alphabétiques européennes

Parmi les langues d'origine européenne basées sur un système alphabétique, l'italien et le finnois, par exemple, sont considérées comme faisant partie des langues les plus transparentes (Ziegler, 2018). De l'autre côté de ce continuum se trouve l'anglais, considérée comme étant la langue la plus opaque, compte tenu des nombreuses possibilités de prononciation d'une même lettre d'une part, et des nombreuses possibilités d'écriture d'un même phonème d'autre part (*e.g.* Ziegler, 2018; Ziegler et al., 2010). Par exemple, un graphème anglais telle que la lettre *a* se prononcera de plusieurs manières, comme dans *hat*, *bake* ou *was*. De la même manière, un phonème anglais tel que /o/ pourra s'écrire *o*, *ough*, *aw*, *au*...

La diversité des systèmes d'écriture vient donc éclairer les différences que l'on peut constater dans l'apprentissage de la lecture et de l'écriture en fonction de la langue concernée. Malgré cela, des théories peuvent être proposées pour un même système d'écriture, comme par exemple les langues basées sur un système alphabétique. Ces théories se fondent en effet sur les mécanismes cognitifs impliqués dans cet apprentissage de la lecture, ainsi que les stratégies mises en œuvre pour arriver à une lecture experte et automatisée.

Ziegler (2018) présente un tableau comparatif de l'opacité de l'anglais et du français, d'après les données de Peereman et Content (1999), qui ont répertorié le degré de consistance de chacune de ces deux langues, dans le sens de la lecture (du graphème vers le phonème) et dans le sens de l'écriture (du phonème vers le graphème) en fonction des différentes unités linguistiques (consonne initiale, finale, voyelle, attaque et rime) pour des mots monosyllabiques. Ziegler (2018) propose, via ce tableau, une mise en lumière des différences d'inconsistance entre le français et l'anglais : ainsi, nous remarquons que l'anglais est plus inconsistant que le

français dans le sens de la lecture, tout particulièrement pour l'attaque et la voyelle, alors que le français reste relativement consistant en lecture. Cependant, dans le sens de l'écriture, les deux langues sont toutes deux inconsistantes, de manière similaire sur l'attaque du mot, ainsi que la voyelle ; le français est quant à lui plus inconsistant au niveau de la rime. Cela nous permet donc de constater que, si le français et l'anglais sont toutes deux considérées comme faisant partie des langues opaques, l'inconsistance de leur orthographe n'est toutefois pas identique, ce qui laisse à penser que les études individuelles, en plus des études comparatives, restent nécessaires pour la compréhension du fonctionnement des apprentissages liés à chaque langue.

Dans cette même optique d'étude des différences entre les langues, Seymour et collaborateurs (2003) ont comparé le niveau de lecture et les capacités de décodage phonologique de plusieurs enfants de pays européens, dont l'orthographe de la langue se situe sur un continuum allant de consistant (e.g. finnois, italien) à inconsistant (e.g. français, anglais). Ils ont ainsi montré que les élèves des langues dont l'orthographe est consistante montrent de bonnes performances en fluence de lecture ainsi qu'un faible pourcentage d'erreurs, et ce, avant même la fin de leur première année d'école élémentaire. A contrario, les élèves dont la langue maternelle appartient aux orthographe inconsistantes ont des performances significativement moindres. Les auteurs attribuent ce décalage à la difficulté d'acquisition de la lecture en fonction du degré de consistance de l'orthographe concernée, la difficulté s'accroissant pour les orthographe inconsistantes tel que le français.

Ainsi, le degré de transparence de la langue peut faire partie des facteurs impactant l'apprentissage de la lecture et de l'écriture, généralement associées au début de l'école élémentaire. Si certains aspects sont spécifiques à chaque langue, les théories explicatives de l'acquisition de la lecture et de l'écriture, généralisables à l'ensemble des systèmes d'écriture, devraient aussi pouvoir tenir compte de cette spécificité linguistique.

1.4 Description de l'orthographe française

Le français écrit est une langue alphabétique qui intègre à la fois des principes phonographiques et morphographiques (Fayol, 2009 ; Fayol & Jaffré, 1999). Dans cette partie descriptive de l'orthographe française, nous aborderons d'abord ses caractéristiques en lien avec le principe phonographique, en distinguant trois niveaux de compétence orthographique :

le niveau des correspondances phono-graphémiques, le niveau des connaissances lexicales et celui des connaissances générales. Le niveau des connaissances morphologiques, correspondant au niveau morphographique, sera quant à lui abordé en fin de partie.

1.4.1 Le code alphabétique, ou les correspondances graphème-phonème

En tout début d'apprentissage, l'enfant acquiert le code alphabétique, qui lui donne la capacité à écrire de manière phonologiquement correcte, c'est-à-dire en faisant correspondre un graphème pour chaque phonème entendu, sans tenir compte du respect de l'orthographe conventionnelle. Cette capacité nécessite de connaître un graphème associé à chaque phonème de la langue. Toutefois, les correspondances entre graphèmes et phonèmes, à savoir les différentes possibilités de lecture d'un même graphème, sont très inconsistantes en français. Celles-ci sont relativement complexes, notamment comparativement à d'autres langues européennes. Ainsi, une même lettre pourra se prononcer de différentes manières : à la lettre *t*, par exemple, pourra correspondre le son /t/ (comme dans *tuteur*), mais aussi le son /s/ (comme dans *population*), ou bien encore ne pas se prononcer, si le *t* est par exemple positionné en fin de verbe (comme dans *il lit*). D'après Fayol et Jaffré (2008), la consistance de l'orthographe française dans la correspondance du graphème au phonème est de 96% ; si cela reste bien supérieur à la consistance de la correspondance du phonème au graphème (71%), les mots dont la prononciation ne correspond pas à la graphie restent toutefois fréquents (e.g., *monsieur*, *femme...*).

Mais c'est surtout dans la complexité des relations phono-graphémiques que cette inconsistance orthographique est la plus visible : selon Fayol et Jaffré (2014), il existerait dans la langue française plus de 130 graphèmes pour seulement une trentaine de phonèmes (Fayol & Jaffré, 2014) ; il y a donc la plupart du temps plusieurs possibilités d'écrire un même phonème (e.g. le phonème /o/ peut s'écrire *o*, *au*, *eau*). Cette dissymétrie dans le nombre de graphèmes et phonèmes disponibles dans la langue augmente considérablement la complexité pour l'acte d'écrire à partir de l'oral, puisqu'un choix sera donc requis de manière régulière lors de la production d'un écrit. Or, cette inconsistance phono-graphémique se retrouve dans quasiment tous les phonèmes en français. La conversion de phonèmes en graphèmes est donc une des caractéristiques de la complexité de l'orthographe française (Ziegler et al., 1996). Compte tenu de cette forte inconsistance de la langue, les connaissances orthographiques lexicales seront d'une grande importance dans le développement de la production orthographique.

1.4.2 Les connaissances orthographiques

1.4.2.1 *Les connaissances orthographiques lexicales*

L'orthographe lexicale correspond à l'orthographe conventionnelle, et est donc à distinguer des capacités à écrire de façon phonologiquement correcte. Dans une langue où l'inconsistance phono-graphémique est aussi importante, on ne peut écrire que peu de mots correctement en faisant appel uniquement aux correspondances phonèmes-graphèmes les plus fréquentes. L'acquisition des connaissances lexicales orthographiques est alors fondamentale pour devenir un bon scripteur. Ce niveau fait appel au lexique orthographique, c'est-à-dire que le scripteur va récupérer en mémoire des informations sur l'orthographe des mots préalablement rencontrés et stockés sous leur forme conventionnelle. Cette capacité à rendre compte correctement de la forme écrite d'un mot se fait progressivement, dès le début de l'apprentissage explicite du langage écrit à l'école élémentaire. L'acquisition et la croissance de ce lexique orthographique peut se faire également de manière implicite, via une exposition répétée à un même mot, afin qu'il puisse être stocké en mémoire : c'est le principe de l'auto-apprentissage, sur lequel nous reviendrons plus précisément ultérieurement.

1.4.2.2 *Les connaissances orthographiques sublexicales*

Le troisième niveau concerne les connaissances orthographiques générales ; ces connaissances sont implicites, c'est-à-dire qu'elles ne correspondent pas à des connaissances conscientes et verbalisables par le scripteur et correspondent à des connaissances statistiques, aussi nommées les régularités graphotactiques. La régularité graphotactique correspond à la probabilité de transcrire un phonème en un certain graphème en fonction de la fréquence à laquelle cette association phonème-graphème existe dans le système orthographique concerné, mais aussi de la régularité de la co-occurrence des lettres qui composent le graphème. Par exemple, s'il est fréquent de rencontrer le graphème *-o*, il est toutefois très rare de le rencontrer

en fin de mot sous cette forme écrite. Cette connaissance des régularités graphotactiques permettrait donc au scripteur de faire des choix basés sur les associations non conscientes qu'il a rencontrées au cours de son expérience de lecteur-scripteur, en lien avec la place du graphème dans le mot, et la spécificité des lettres qui le composent. Elle est également appelée orthographe sublexicale.

Les régularités sont basées sur les probabilités de rencontrer un graphème donné en fonction de différents paramètres : sa position à l'intérieur du mot (position initiale, médiane ou finale, par exemple), ses graphèmes voisins, la fréquence de co-occurrence des lettres qui le composent, ou bien encore la fréquence de doublage d'une lettre qui le compose. C'est ainsi, par exemple, que lorsque l'on demande à un élève de choisir, parmi deux propositions de pseudomots, celui qui ressemble le plus à un mot du français, il aura tendance à choisir majoritairement *tunnar* plutôt que *ttunar*, car il a intégré, de manière implicite, qu'un mot français ne commence jamais par une double consonne (voir Pacton et al., 2013, pour une étude sur les règles implicites du français). S'il se base sur les mots français qu'il a rencontré, il aura donc déjà rencontré des mots comprenant une consonne doublée en position médiane, le *n* étant une consonne fréquemment doublée, mais jamais une consonne doublée en position initiale.

La sensibilité aux régularités graphotactiques d'une langue semble être une habileté présente précocement chez l'enfant : en effet, les recherches montrent que dès l'âge préscolaire, les enfants font preuve de certaines connaissances graphotactiques propres à leur langue (Treiman, 2017 ; Pollo, Kessler & Treiman, 2009). Sénéchal et collaborateurs (2016) ont montré qu'en français, les élèves de début de scolarité élémentaire (Grade 1, 2 et 3, correspondant respectivement au CP, CE1 et CE2) avaient déjà cette capacité, notamment concernant la particularité française des lettres muettes en position finale. Dans leur étude, les élèves devaient en effet écrire sous dictée des mots dont la lettre finale muette était plus ou moins fréquente (*t*, pour la plus fréquente, et *d* pour la moins fréquente). En Grade 1 et Grade 2, les résultats ont montré une meilleure performance pour la lettre fréquente *t* que pour la lettre moins fréquente *d*, confortant l'hypothèse de départ selon laquelle les caractéristiques des lettres muettes finales en français sont donc bien soumises aux propriétés statistiques du système orthographique du français écrit, et ce dès le début de la scolarité élémentaire.

Cet apprentissage graphotactique implicite correspond en fait à la version orthographique de ce qu'on appelle plus généralement l'apprentissage statistique, mode d'apprentissage très puissant dès la naissance et qui permet de faire des prédictions sur le monde qui nous entoure (voir Arciuli, 2017 pour une synthèse sur les connaissances actuelles sur

l'apprentissage statistique). Si cet apprentissage graphotactique se fait généralement hors de la situation d'enseignement formel en classe, il semble toutefois qu'il puisse être enseigné de manière explicite en classe : Treiman et Kessler (2022) mettent en avant l'intérêt de l'enseignement de cet apprentissage, notamment pour permettre à l'enfant de prendre conscience des liens qui peuvent exister entre les lettres ; ce travail réflexif sur la construction de la langue écrite, qui rend l'enfant actif dans ce travail sur la langue semble favoriser le développement de l'orthographe.

Tous les auteurs ne font pas la distinction que nous venons de présenter entre l'orthographe lexicale et les connaissances graphotactiques. Cunningham et collaborateurs (2001) ont voulu par exemple vérifier la contribution unique des habiletés de traitement orthographique (*orthographic processing skills*) aux capacités de reconnaissance de mots. L'objectif pour eux était de montrer que les habiletés de traitement orthographique jouent un rôle spécifique, indépendant des compétences phonologiques, dans l'apprentissage de la lecture en début de scolarité élémentaire. On observe donc ici plutôt une distinction en deux niveaux, avec d'une part le volet phonologique (via des tâches de manipulation de phonèmes et de répétition de pseudomots), et d'autre part un volet orthographique général, regroupant à la fois l'orthographe lexicale et sublexicale, dont le score composite comporte des épreuves de sensibilité graphotactique (choix de combinaison de lettres), une épreuve de choix orthographique (reconnaissance orthographique), une épreuve de choix homophonique, ainsi qu'une épreuve de dictée.

Si la distinction entre l'apport des compétences phonologiques et celui des compétences orthographiques dans l'apprentissage de la lecture-écriture semble généralisée, les habiletés constitutives de la composante orthographique ne font cependant pas encore l'objet d'un consensus établi. Le présent travail s'attachera donc à investiguer cette composante orthographique afin de tenter d'apporter des informations supplémentaires concernant ce point.

1.4.3 Principe morphographique : les connaissances morphologiques

1.4.3.1 La morphologie dérivationnelle

Une autre caractéristique majeure de la difficulté d'écriture de l'orthographe française réside dans la présence de nombreuses lettres muettes en fin de mots, indispensables à la bonne compréhension de l'écrit, mais qui ne se prononcent pas à l'oral. Ces lettres muettes finales se traduisent donc par des graphèmes sans phonème « effectif » correspondant. D'après Mussar et al. (2020), environ 30% des mots de la langue française se terminent par une lettre muette. La morphologie dérivationnelle permet d'expliquer une partie de ces lettres finales muettes, via le champ lexical de ces mots. Par exemple, le mot *chant* finit par un *-t* muet, donc visible à l'écrit mais non prononcé, et le verbe correspondant, le mot *chanter*, permet de comprendre pourquoi la lettre *-t* vient se placer à la fin du nom commun *chant*. Ce *-t* en fin de *chant* est un morphogramme lexical, c'est-à-dire une marque qui permet un lien visuel avec des dérivés (comme *chanter* par exemple). Une étude a d'ailleurs démontré que la prise en compte de la morphologie permet de diminuer le degré d'inconsistance des fins de mots écrits en français (Peereman & Sprenger-Charolles, 2018).

La morphologie dérivationnelle se base sur l'utilisation d'un morphème, c'est-à-dire la plus petite unité de sens qui peut composer un mot (Fayol & Jaffré, 2008). A ce morphème peuvent donc s'ajouter différents composants (affixes, préfixes ou suffixes, par exemple) afin de créer d'autres mots à partir de la base du morphème choisi. Selon notre exemple précédent, au morphème *chant* a été ajoutée la marque du verbe à l'infinitif du premier groupe (la terminaison en *-er*).

1.4.3.2 La morphologie flexionnelle

La morphologie flexionnelle est une autre explication de la complexité de l'orthographe française, en lien cette fois-ci avec les accords grammaticaux (en genre et en nombre, par exemple). Fayol et Jaffré (2008) présentent trois catégories composant la morphologie flexionnelle : la catégorie sémantique (le genre), morphosyntaxique (la fonction du mot dans la phrase), et la morphophonologique sous-lexicale (la terminaison du mot). Ainsi, pour reprendre l'exemple cité précédemment, le pluriel de « *un chanteur* » sera donc « *des chanteurs* », l'ajout du *-s* en fin de mot représentant l'accord en nombre pour le pluriel.

La complexité de l'orthographe du français lors du passage de l'oral à l'écrit n'est donc pas seulement liée à l'orthographe du mot en lui-même, mais également au contexte de la phrase dans laquelle il s'inscrit : *il chante* et *ils chantent* étant des homophones, par exemple, il est

nécessaire pour le scripteur de récupérer différents indices pour pouvoir choisir l'orthographe correcte (voir Bosse et al., 2021, pour plus d'illustrations sur les spécificités du système d'écriture français). L'orthographe du français tire donc ses particularités et sa complexité principalement de ses relations entre phonèmes et graphèmes, ainsi que des spécificités morphologiques qui la composent.

Dans ce premier chapitre, nous avons donc commencé par définir les concepts de lecture et d'écriture ; puis nous avons présenté les différents systèmes d'écriture. Nous avons ensuite proposé une comparaison de quelques langues européennes appartenant au système alphabétique, pour montrer leurs différences ; enfin, nous nous sommes attardés sur les particularités de l'orthographe française, avec notamment l'importance des correspondances graphème-phonème, la présentation de l'orthographe lexicale, de l'orthographe sublexicale et les spécificités liées à la morphologie.

Dans le chapitre suivant, nous allons présenter quelques grandes théories explicatives de la lecture et de l'orthographe, en commençant par les modèles experts les plus répandus dans ce champ de recherche. Nous présenterons ensuite différents modèles, qu'ils soient développementaux ou interactifs, par exemple, ou bien encore centrés sur l'acquisition et le développement de la compétence orthographique.

Résumé du Chapitre 1

La lecture correspond au produit du décodage et de la compréhension (modèle simple de la lecture, Gough & Tunmer, 1986). L'écriture, quant à elle, est le produit de l'épellation et de l'idéation (modèle simple de l'écriture, Juel *et al*, 1986).

Le système alphabétique est basé sur l'association entre des phonèmes et des graphèmes. Le degré de transparence d'une langue a un impact sur ces apprentissages.

En français, l'inégalité de la consistance entre les phonèmes et les graphèmes rend le passage de l'oral à l'écrit d'autant plus complexe. L'orthographe lexicale se base sur le lexique orthographique pour stocker en mémoire la forme orthographique conventionnelle d'un mot. L'orthographe sublexicale se base sur les régularités graphotactiques, c'est-à-dire la probabilité d'écriture d'un graphème en fonction de la fréquence à laquelle cette forme graphémique est rencontrée dans une langue. La morphologie correspond aux caractéristiques liées au sens : elle peut être dérivationnelle (morphèmes) ou flexionnelle (accords grammaticaux).

Chapitre 2 : les modèles théoriques

Dans ce deuxième chapitre, nous allons présenter dans un premier temps les théories concernant le fonctionnement expert de la lecture et de l'orthographe, afin d'explicitier la manière dont lecture et orthographe ont été modélisées. Nous commencerons donc par présenter les modèles experts de la lecture et de l'orthographe, puis nous présenterons des modèles développementaux et interactifs.

2.1 Modèles experts de la lecture et de l'orthographe

Depuis plusieurs décennies, différents modèles théoriques ont été proposés pour essayer d'expliquer les mécanismes de la lecture et de la production orthographique sur le plan cognitif. Nous allons ici présenter les principaux modèles, ainsi que les éventuelles critiques qui leur ont été adressées afin de mettre en avant l'évolution des théories jusqu'à l'état actuel des connaissances.

Coltheart (1978, d'après Coltheart et al., 2001) a proposé un type de modèle appelé *dual route*, ou modèle double voie, qui envisage l'accès à la lecture selon deux chemins possibles : une voie d'assemblage (ou voie indirecte, non basée sur une reconnaissance du mot lu), par laquelle le lecteur doit d'abord « reconstruire » le mot via les règles de correspondances grapho-phonémiques avant de pouvoir accéder au lexique mental pour prononcer le mot lu ; et une voie d'adressage (ou voie directe), par laquelle l'identification du mot se fait directement via l'accès au lexique mental qui reconnaît le mot écrit et permet sa prononciation. Coltheart et collaborateurs (2001) ont implémenté une version computationnelle de ce modèle : le *dual route cascaded (DRC) model*, basé sur les tâches de décision lexicale et de lecture à voix haute. Ce modèle comporte théoriquement trois routes, mais deux seulement ont été implémentées : la route lexicale non sémantique, correspondant à la voie d'adressage, où le lexique orthographique intervient ; et la route basée sur les correspondances graphème-phonème, qui fait écho à la voie d'assemblage, où chaque unité présentée est assimilée à son équivalent phonétique. La route lexicale sémantique, avec la prise en compte notamment des représentations sémantiques, n'a pas été implémentée. La combinaison du traitement fait par

l'ensemble des routes (lexicale non sémantique et lexicale sémantique d'un côté, ainsi que la route basée sur les règles de correspondance graphème-phonème de l'autre) permettrait au modèle de produire les sons correspondants aux mots présentés à l'écrit. D'après ce modèle, l'orthographe lexicale correspondrait donc à la voie d'adressage (ou route lexicale non sémantique), et l'orthographe sublexicale à la voie d'assemblage (où les règles basées sur les correspondances entre graphèmes et phonèmes prédominent).

En ce qui concerne la production de mots écrits, Kandel (à paraître) propose un modèle qui se conçoit, comme certains modèles de la lecture experte, dans une perspective double voie : le modèle APOMI (*Anticipatory Processing of Orthographic and Motor Information*). Ce modèle a la particularité d'intégrer conjointement le traitement orthographique ainsi que l'exécution motrice de l'écriture. Le modèle APOMI propose deux voies, une lexicale et une sublexicale, qui fonctionnent de manière parallèle. L'originalité de ce modèle réside dans l'activation simultanée des traitements orthographiques et moteurs, ce qui permet une continuité dans la transmission de l'information et donc sa réactualisation permanente. A l'instar des modèles double voie concernant la lecture, la vision d'une double voie est donc aussi envisagée en ce qui concerne l'écriture.

Concernant la langue française spécifiquement, ce sont Perry et al. (2014) qui ont testé ce modèle double voie de la lecture dans sa version la plus récente, le CDP++ (*Connectionist Dual Process*). L'objectif était de vérifier si le modèle prenait en compte les spécificités du français, notamment la présence des nombreuses lettres muettes en fin de mot. Le modèle CDP++ propose donc deux voies distinctes, comme le modèle DRC : une voie lexicale, et une voie sublexicale. La voie lexicale mobilise le lexique orthographique pour activer le mot en entier, puis le lexique phonologique ; la voie sublexicale utilise la sélection de graphèmes à partir des lettres présentées en input, et, par un processus d'appariement statistique entre les lettres et le graphème le plus probable, propose une activation phonologique. C'est finalement la combinaison entre les dernières étapes de chacune des voies qui va permettre de produire la prononciation du mot lu. L'implémentation de ce modèle en français semble bien prendre en considération la spécificité du français, notamment concernant les lettres muettes finales, mais les auteurs précisent que l'efficacité du modèle est réduite en ce qui concerne par exemple le voisinage orthographique.

Il y a également d'autres modèles qui simulent la lecture experte mais n'impliquent pas l'existence d'une double voie. Par exemple, le modèle PDP (*Parallel Distributed Processing model*, Seidenberg & McClelland, 1989) est un modèle connexionniste qui propose trois couches d'unités élémentaires : orthographique, phonologique, et sémantique, entre lesquelles viennent agir en tant que médiatrices trois couches d'unités cachées. Un entraînement sur plus de 2000 mots anglais monosyllabiques permet à ce modèle connexionniste de donner des performances correctes concernant la prononciation de plus de 97% des mots appris. Le modèle initial ayant toutefois des difficultés dans la lecture de pseudomots, de nouvelles versions du modèle ont été proposées, dont l'objectif était de tendre vers une amélioration de la lecture de ce type d'items.

Le modèle MTM (*Multiple Trace Memory*, Ans, Carbonnel & Valdois, 1998) a quant à lui pour objectif de simuler la lecture de n'importe quel type de mots sans utiliser de système de règles explicites ou implicites de conversion, qui correspondrait à la voie non lexicale des modèles à double voie. Ce modèle MTM propose, à travers quatre couches, deux procédures successives : une procédure globale (qui concerne la connaissance des mots entiers), et une procédure analytique (qui suppose un découpage des mots en syllabes). Le réseau, entraîné sur des mots français mono et polysyllabiques, est efficace à la fois pour les mots et les pseudomots, qu'ils soient réguliers ou irréguliers. L'originalité de ce modèle réside principalement dans sa simulation d'un traitement visuo-attentionnel modulable (sur le mot entier ou sur une syllabe à la fois) au niveau de la couche orthographique.

Perfetti (Perfetti & Hart, 2002 ; Perfetti, 2007) s'est également intéressé au lien entre la compréhension, c'est-à-dire le but final recherché dans l'activité de lecture et la qualité des représentations lexicales associées aux mots lus. Si l'hypothèse de qualité lexicale a pour but l'explication de la compréhension en lecture, cette théorie met en avant l'importance des représentations lexicales (orthographe, phonologie, morphologie...) ainsi que leur interconnexion. Pour lui, si le décodage est indispensable à la lecture, les connaissances lexicales jouent également un rôle majeur dans la capacité du lecteur à comprendre ce qu'il lit. Il propose ainsi l'hypothèse de qualité lexicale (*lexical quality hypothesis*, ou LQH) selon laquelle la transformation d'une forme orthographique en une représentation lexicale est la clé pour une compréhension efficace et pertinente. En effet, une représentation lexicale de qualité aura pour lui deux aspects : une représentation orthographique précise et spécifique, et des représentations phonologiques redondantes (une à partir du langage oral, l'autre récupérable

depuis le passage de l'orthographe à la phonologie). Si la représentation lexicale comprend ces deux aspects, alors la récupération en mémoire, lors d'une activité de lecture, sera plus pertinente. La qualité lexicale peut ainsi varier selon plusieurs critères : l'orthographe (l'écriture correcte du mot), la récupération des prononciations possibles du mot, et les différents sens possibles du mot. Il y a donc trois différents niveaux sur lesquels la qualité lexicale peut être impactée. Tout d'abord, la polysémie : dans cette situation, à une orthographe donnée correspondent plusieurs sens (ou morphèmes). ; par exemple, *souris* peut désigner à la fois un animal et un outil informatique. Ensuite, l'homographie : ici, il y a également une seule orthographe mais plusieurs prononciations possibles, qui correspondent à plusieurs sens ; par exemple, *fil*s, désignant un nombre d'enfants garçons, et *fil*s, désignant des morceaux de tissu agglomérés. Enfin, l'homophonie : dans ce cas de figure, il y a plusieurs orthographes possibles pour une même prononciation, qui conduisent à plusieurs sens différents, comme avec *ver*re/*ver*s/*ver*, qui se prononcent tous de manière identique, mais désignent respectivement une matière, une unité qui compose une rime en poésie, ou encore un lombric. Dans ces trois configurations, où une orthographe ne correspond pas forcément à une seule phonologie ni à un seul sens, la qualité lexicale peut être impactée ; c'est donc principalement le contexte, s'il y en a, qui permettra au lecteur de choisir la représentation qui lui semble être la plus adaptée. Pour Perfetti et Hart (2002), la fréquence d'occurrence du mot dans la langue a également un impact majeur en cas de choix multiple : en effet, plus un mot est fréquent, plus il a de chance d'avoir une qualité lexicale haute, et donc de permettre une meilleure discrimination entre les items proposés. Concrètement, lors de tâches orthographiques par exemple, une représentation orthographique incomplète (donc, avec une qualité lexicale faible ou moyenne) pourrait suffire à permettre à l'élève de trouver la réponse correcte dans une tâche de reconnaissance orthographique, où plusieurs choix sont proposés. Par contre, cette même représentation orthographique incomplète ne serait pas suffisante pour aider l'élève à faire un choix sûr et correct dans une tâche de production de type dictée.

Selon Chaves et collaborateurs (2012), ces modèles experts, s'ils présentent un fonctionnement différent et intéressant dans la façon dont le réseau connexionniste apprend puis est capable de lire des nouveaux mots, ne permettent toutefois pas une correspondance réaliste avec la phase d'apprentissage de l'enfant. Le modèle double voie tend à rendre compte du processus de lecture experte, et n'explique donc pas la manière dont cette lecture experte s'acquiert ; certains modèles développementaux présentés ci-après se sont basés sur ce modèle

double voie pour essayer d'expliquer comment l'apprentissage vers ce processus de lecture experte peut se produire.

2.2 Développement et acquisition de la lecture et de l'orthographe

2.2.1 Universalité de la lecture et de l'orthographe

Les modèles experts présentés précédemment étant majoritairement alphabético-centrés, certains auteurs ont tenté de dégager les habiletés communes à l'ensemble des apprentissages des différentes langues présentes à travers le monde. Pour ce faire, ils ont étudié et comparé un grand nombre de langues, provenant de systèmes d'écritures divers.

Frost (2012) propose l'idée d'un modèle universel qui prendrait en compte l'ensemble des systèmes d'écriture. Ce modèle serait donc plus général, voire abstrait, puisqu'il devrait pouvoir s'appliquer à tous ces systèmes d'écriture, pour lequel il utilise le terme d'universaux de la lecture (*reading universals*). Concrètement, il propose par exemple que si deux langues montrent un fonctionnement différent voire opposé dans le mécanisme de lecture, le modèle universel devrait chercher l'habileté commune de plus haut niveau qui régit ce mécanisme, même si celui-ci peut fonctionner différemment selon la langue concernée. Il évoque également le concept de plausibilité linguistique, selon lequel les mots ont des caractéristiques morphologiques, phonologiques, sémantiques et orthographiques que le modèle universel devrait prendre en compte sans que l'explication de l'une de ces caractéristiques ne vienne contredire ou empêcher l'explication des autres. En s'appuyant sur cinq langues de cinq systèmes d'écriture différents (chinois, japonais, finnois, anglais et hébreu), il propose de s'intéresser aux propriétés statistiques des langues pour expliquer leur fonctionnement, notamment sur le plan cognitif : ainsi, chaque langue utiliserait la stratégie la plus efficace pour distinguer les mots lors du processus de lecture. Selon ce point de vue, l'inadéquation entre certains modèles théoriques et certains systèmes d'écriture serait donc contournée.

Certaines caractéristiques peuvent également se retrouver dans plusieurs langues, la plupart du temps issues de la même famille : Verhoeven et Perfetti (2022) ont récemment comparé dix-sept langues, provenant de différentes familles de langues ainsi que de différents systèmes d'écriture, afin de mettre en évidence leurs caractéristiques communes. Ils ont pour cela choisi des langues provenant de cinq systèmes d'écriture : syllabiques (kana), morpho-

syllabiques (kanji), alphasyllabiques (kannada), abjad (arabe, hébreu) et alphabétique (majorité des langues européennes). Ils considèrent deux composantes dans les systèmes d'écriture dont la variation est à prendre en compte : la complexité graphique et la configuration des principes du langage écrit. Selon eux, la complexité graphique aurait une influence sur l'acquisition d'une bonne qualité de représentation orthographique, avec l'exemple d'un continuum composé d'un côté de l'hébreu qui aurait la complexité graphique la plus simple, et de l'autre du chinois traditionnel pour la plus complexe. La configuration du langage écrit correspondrait quant à elle à la manière dont l'écriture encode le langage oral à différents niveaux (logographique, syllabique et alphabétique) de manière indépendante mais aussi interconnectée en fonction des caractéristiques des systèmes d'écriture concernés. Enfin, en ce qui concerne l'apprentissage de la lecture, ils considèrent que les principes généraux applicables à l'ensemble des systèmes d'écriture se basent essentiellement sur le fait que le développement de la lecture implique l'apprentissage de la manière dont un système d'écriture donné encode le langage. Ainsi, il s'agit de considérer de manière générale les trois aspects principaux de l'apprentissage de la lecture : la conscience linguistique, l'acquisition de la capacité d'identification des mots et le développement de la compréhension. Dans cette perspective, Li et collaborateurs (2022) ont proposé une synthèse sur les mécanismes de lecture dans différents systèmes d'écriture. Les auteurs, s'ils mettent en avant les distinctions entre les différents systèmes d'écriture, montrent aussi leurs bases communes, à savoir la transformation d'un texte écrit en un code utilisé par le système de traitement du langage, l'utilisation d'informations *top-down* et *bottom-up* pour donner du sens, et le traitement du texte en fonction du mouvement oculaire. Si les modèles basés sur les systèmes alphabétiques mettent en avant trois catégories nécessaires au traitement de l'information, que cela soit en lecture ou en écriture (c'est-à-dire le lien entre sémantique, phonologie et orthographe), ce lien sera moins direct dans les systèmes logographiques, nécessitant par exemple une phase transitoire d'identification de caractère entre l'information sémantique et phonologique.

La vision universelle de la lecture et de l'écriture, si elle met en évidence trois grands principes liés à la langue orale (conscience linguistique, identification des mots et compréhension) ne permet cependant pas d'explicitier le développement complexe de ces compétences. Cette perspective « macro » de la lecture et de l'écriture permettrait alors de proposer un cadre très général dans lequel chaque langue pourrait s'inscrire. Une perspective « micro », complémentaire, semble dans ce cas indispensable pour permettre d'étudier les spécificités de chacune des langues, afin de pouvoir mieux comprendre leur fonctionnement et

donc d'améliorer leur enseignement. En ce qui concerne les langues alphabétiques plus spécifiquement, des modèles développementaux ont donc cherché à investiguer plus profondément la construction des habiletés que sont la lecture et l'écriture.

2.2.2 Les modèles développementaux

Frith (1985), dans la perspective des modèles à double voie, présente un modèle développemental, précisant que pour accéder à ces voies, trois étapes sont indispensables : il s'agit des procédures logographiques, alphabétiques, et orthographiques. Cette vision en stades d'acquisition des capacités de lecture requiert tout d'abord que le sujet soit capable de reconnaître des formes abstraites comme étant des mots, sans faire appel à des capacités autres que les capacités visuelles : la procédure logographique s'acquiert généralement par association (e.g., « reconnaître » son prénom pour un enfant, ou encore le nom d'une marque dont le logo est visuellement marqué et familier pour l'enfant comme l'écriture de la marque *Coca-Cola*...). Une fois cette procédure logographique acquise, l'enfant accède peu à peu à la procédure alphabétique : avec l'apprentissage explicite de la lecture à l'école, l'enfant apprend les correspondances entre graphèmes et phonèmes. Ce n'est en effet qu'à la procédure orthographique, soit le dernier stade d'acquisition, que l'enfant pourra pleinement accéder à la voie directe, ou voie d'adressage, qui lui permettra de lire les mots sans transiter par l'étape alphabétique qui reviendrait à du décodage. Selon ce modèle, l'orthographe lexicale n'apparaît donc qu'au moment de la procédure orthographique.

Cette vision stadiste de l'acquisition de la lecture a été critiquée pour la rigidité dans l'accès aux stades ultérieurs, qui ne pourrait se faire que via la stricte maîtrise du stade précédent. Frith (1986) a donc proposé une version plus détaillée de ce modèle, précisant deux sous-stades pour chacun des trois stades d'acquisition, pendant lesquels un chevauchement du deuxième sous-stade avec le premier sous-stade du stade suivant permettrait un accès partiel au stade suivant, alors que le stade précédent n'est pas encore totalement maîtrisé. Elle présente ainsi une forme d'alternance entre la lecture (reconnaissance des mots) et l'écriture (production de mots).

Un autre modèle développemental, proposé par Ehri (2005), différencie trois phases, correspondant à trois stratégies utilisées majoritairement par l'enfant au moment de la lecture : les indices visuels, les indices phonétiques et le recodage phonologique. Le traitement par

indices visuels correspond peu ou prou à la phase logographique présentée par Frith (1985) ; ces trois stratégies par indices et recodage correspondent selon le modèle à quatre phases du développement de l'enfant : la phase pré-alphabétique, où le traitement par indices visuels est le plus prégnant ; la phase alphabétique partielle, qui sert de transition au début de l'apprentissage explicite de la lecture et pour laquelle l'enfant se base sur des indices phonétiques partiels dans différentes parties du mot ; la phase alphabétique complète, où l'enfant maîtrise le décodage via la correspondance graphème-phonème ; et enfin la phase alphabétique consolidée où l'enfant est capable de traiter les informations écrites en unités plus grandes que la lettre, et peut utiliser une stratégie de *chunking* (assemblage des lettres proches en une unité composée de plusieurs items). Cette dernière phase dite alphabétique correspond au stade orthographique de Frith (1985), permettant l'accès à la voie directe pour une lecture des mots qui ne se base plus sur une activité de décodage. Là encore, l'accès à l'orthographe lexicale est donc envisagé uniquement à partir du dernier stade développemental.

Si ces modèles développementaux ont permis de décrire certaines stratégies de lecture lors de son apprentissage chez l'enfant, ils ont cependant été critiqués notamment pour leur caractère peu généralisable : ils n'expliquent en effet que peu les différences interindividuelles existantes. De plus, la dépendance d'une acquisition de compétence pour passer d'un stade à l'autre les rend très inflexibles, et pas toujours réalistes en ce qui concerne le processus d'acquisition de la lecture et les difficultés qui peuvent y être liées.

2.2.3 Les modèles interactifs

Les modèles interactifs ont vocation à tenter de se détacher de la rigidité des stades des modèles développementaux en permettant aux différentes parties qui les composent d'évoluer de manière parallèle si cela est spécifié dans le modèle. Seymour (1997) propose le *dual foundation model*, ou modèle à fondation duale du développement orthographique, avec pour objectif d'éclairer le développement d'un système alphabétique. Également, il lui semble important que le modèle présenté tienne compte de trois aspects mis en avant par Goswami et Bryant (1990; Goswami, 1999) : premièrement, le fait que dans la langue anglaise, l'orthographe code à la fois les informations au niveau lexical, morphologique, mais aussi alphabétique ; deuxièmement, que le modèle permette l'existence de déficiences orthographiques de différentes formes (lexicales ou non lexicales) ; enfin troisièmement, que l'effet potentiel de la conscience linguistique soit visible non seulement au niveau attaque-rime,

mais aussi au niveau phonémique. Tenant compte de ces caractéristiques, il propose un modèle à cinq composants majeurs. Les deux premiers composants constituent la base ou fondation du modèle ; ils ne jouent pas seulement un rôle au début de l'apprentissage, mais continuent tout au long du développement ; il s'agit du processeur logographique et du processeur alphabétique. Nous précisons ici toutefois que le terme logographique utilisé par l'auteur n'a pas le même sens qu'évoqué dans le modèle de Frith (1985) : ici, le terme logographique renvoie à la reconnaissance et au stockage du mot entier, contrairement aux autres modèles pour qui ce terme renvoie à des indices visuels ou à des parties de mot mais non au mot entier. Le troisième composant du modèle est la conscience linguistique ; elle a ici un rôle interactif, de relai entre les autres composants. Enfin, les deux derniers composants sont la structure orthographique et la structure morphographique : ces deux structures, centrales dans le modèle, ont un rôle de représentation : ce sont à l'intérieur de ces structures que se fait le stockage des connaissances abstraites de l'individu.

Gombert, Bryant et Warrick (1997) mettent l'accent sur l'utilisation des analogies au cours de l'apprentissage de la lecture. Ils insistent en effet sur l'importance de l'effet de voisinage orthographique lors de la lecture, c'est-à-dire que la prononciation d'une lettre ou d'un groupe de lettres va dépendre des lettres environnantes. Ainsi, grâce aux connaissances préalablement acquises, le lecteur va pouvoir choisir une prononciation qu'il considérera comme la plus adaptée lorsqu'il rencontrera des mots nouveaux (ou des pseudomots s'il se trouve dans une situation d'expérimentation pour une recherche). Les auteurs proposent une apparition plus précoce de cette utilisation des analogies ; contrairement au modèle de Frith par exemple, qui ne laisse la place aux analogies qu'au deuxième stade, le stade alphabétique, les auteurs considèrent que la capacité à se baser sur des sons similaires est déjà présente chez les enfants de six à huit ans (Gombert, 1986, cité dans Gombert, Bryant & Warrick, 1997). Dans cette étude, les enfants étaient déjà capables de lire des mots trop difficiles pour leur âge, en tenant compte de la rime et de la terminaison orthographique des mots proposés (voisinage phonologique et orthographique). Ainsi, les auteurs considèrent que l'enfant acquiert une base de connaissances non seulement visuelle au début de l'apprentissage, mais également phonologique grâce à ses expériences préalables concernant la manipulation du langage oral.

Share (1995 ; 1999) a proposé une théorie majeure pour expliquer l'acquisition de la lecture mais aussi le développement du lexique orthographique : il s'agit de l'hypothèse de l'auto-apprentissage (*self-teaching hypothesis*). Plus l'enfant va être exposé à des mots écrits,

via des activités de lecture, plus il va rencontrer les formes écrites correctes des mots. La théorie de l'auto-apprentissage postule qu'à force de rencontrer ces mots et de les lire de manière correcte, le décodage étant partie intégrante de cet auto-apprentissage, le lecteur va pouvoir stocker dans son lexique orthographique la forme correcte du mot, de manière implicite, c'est-à-dire sans avoir eu à l'apprendre de manière volontaire. Ainsi, l'exposition répétée à la forme écrite du mot associée à sa lecture permettrait une mémorisation efficace de l'orthographe du mot. Pour l'instant, à notre connaissance, le nombre d'expositions nécessaire à un stockage durable et correct ainsi que la durée de rétention ne sont toutefois pas clairement définis : les résultats trouvés par Share par exemple diffèrent de ceux trouvés, en anglais cette fois-ci, par Nation et collaborateurs (2007). Cette hypothèse d'auto-apprentissage est composée de construits similaires à ceux proposés dans les modèles précédents : en effet, l'auto-apprentissage passe par une première étape d'assemblage qui lui permet ensuite de créer du lien avec la voie lexicale (ou d'adressage). Ce lien, renforcé par la récurrence de présentation des items, va pouvoir permettre de récupérer ensuite directement l'item vu par la voie lexicale, en se basant sur l'orthographe lexicale du mot. Nous retrouvons donc ici également l'idée d'une voie indirecte et d'une voie directe.

Ziegler et collaborateurs (2014) ont modélisé cet apprentissage de manière connexionniste en utilisant leur modèle CDP pour vérifier la faisabilité de cet auto-apprentissage. Ils ont ainsi entraîné un réseau de neurones sur un petit échantillon de correspondances graphèmes-phonèmes, équivalent aux contenus des méthodes phoniques de lecture anglaise. Ensuite, ils ont présenté des mots écrits nouveaux à ce réseau de neurones, dans le but de les lui faire apprendre. Le réseau a été capable d'apprendre et de mémoriser plus de 25000 mots, et a également appris à en lire de nouveaux : les auteurs en ont donc conclu que l'auto-apprentissage était une théorie plausible et correspondant bien à la réalité d'un processus développemental. Toutefois, si l'auto-apprentissage semble validé dans de nombreux systèmes d'écriture, il ne permet cependant pas d'expliquer entièrement des profils de type « bon lecteur – faible orthographeur », renforçant l'idée selon laquelle l'auto-apprentissage n'est pas le seul contributeur à l'apprentissage orthographique.

2.2.4 Théories spécifiques au développement orthographique

Contrairement aux modèles à étapes que nous avons présentés précédemment, Treiman (2017) a proposé un cadre de travail dans lequel la sensibilité aux régularités graphotactiques est présente même avant l'apprentissage formel de la lecture. Dans un premier temps elle présente, de manière développementale, les différentes périodes par lesquelles l'enfant va passer pour arriver jusqu'à une production orthographique pertinente et réaliste, c'est-à-dire correspondant au but de l'activité d'écriture, à savoir le fait d'utiliser l'écrit comme un outil pour communiquer selon des codes prédéterminés par le système d'écriture auquel l'enfant appartient. Ainsi, dès l'âge de deux ans environ, l'enfant peut commencer à produire du contenu se rapprochant des caractéristiques de l'écrit, notamment via l'incorporation dans le dessin de « gribouillages ». Cela le prépare, d'un point de vue moteur, au geste de l'écriture, différent de celui du dessin classique. Dans l'étape suivante, que Treiman (2017) appelle écriture préphonologique, l'enfant utilise les régularités graphotactiques ainsi que les digrammes pour produire des écrits sans lien avec la phonologie correspondante. Vient ensuite l'étape des écritures inventées, dans laquelle l'enfant va pouvoir incorporer la phonologie ; puis finalement, plus tard dans le développement, l'enfant apprendra à faire correspondre un graphème au phonème donné. Toutefois, la capacité à faire correspondre graphèmes et phonèmes n'est pas suffisante, spécifiquement dans les orthographes opaques, pour maîtriser l'écrit dans toutes ses possibilités. Ainsi, l'enfant devra intégrer d'autres sources d'informations pour lui permettre de choisir la forme écrite correcte, comme par exemple les informations morphologiques ou encore le contexte.

Dans la suite de sa synthèse, Treiman (2017) présente les différents courants théoriques sur le développement de l'orthographe. Ainsi, elle développe brièvement les modèles stadistes du développement orthographique en lien avec celui de la lecture, comme le modèle de Frith ou celui de Ehri que nous avons présentés plus tôt dans le Chapitre 2. Ces modèles, s'ils mettent sur la voie de l'importance de la phonologie et des connaissances sur la phonologie dans le développement de la lecture et de l'écriture, n'évoquent la connaissance non phonologique que tardivement dans le développement. Les modèles double voie évoqués précédemment incluent la phonologie, en proposant notamment une voie non lexicale, mais ne permettent pas d'expliquer le rôle du graphotactique et de la morphologie. Treiman propose donc un autre cadre théorique dont l'objectif est de décrire le développement de l'écriture tout en prenant en compte une diversité de facteurs impactant ce développement (*Integration of Multiple Patterns,*

Treiman & Kessler, 2014, cité dans Treiman, 2017). Ce modèle propose, en plus de la manière classique d'envisager l'apprentissage de l'écrit comme la forme orthographique du mot, une seconde manière qui intégrerait des liens entre l'unité écrite et les unités linguistiques telles que la phonologie ou la morphologie par exemple. L'idée dans cette proposition serait que les enfants seraient plus à même de choisir une orthographe si plusieurs indices viennent favoriser ce choix ; dans ces circonstances, le rôle de l'apprentissage statistique est donc augmenté, comparativement aux autres modèles. Cette vision permet aussi de mettre en évidence l'émergence de ces différents indices dans l'apprentissage de l'écriture bien avant le début de l'apprentissage formel, incluant ainsi les premières formes d'écriture précoces apparaissant vers l'âge de deux ou trois ans. Ce cadre de travail est en cohérence avec des résultats expérimentaux concernant l'impact de la sensibilité à la régularité graphotactique et de la morphologie sur l'écriture de pseudomots chez des enfants et des adultes (Pacton, Fayol & Perruchet, 2005) : dans une tâche de production de pseudomots, avec ou sans indice morphologique, les enfants comme les adultes utilisent les connaissances sur les régularités graphotactiques (fréquence de co-occurrence des lettres) ainsi que les connaissances morphologiques en plus des informations phonologiques pour produire un nouveau mot plausible au regard de l'orthographe du français.

Ce deuxième chapitre théorique nous a donc permis de présenter, dans un premier temps, les modèles experts les plus influents dans l'explication de la lecture et de l'écriture. Nous avons ensuite proposé de nous interroger sur les aspects universels de ces compétences, avant de présenter des modèles explicatifs du développement de la lecture et de l'écriture. Enfin, nous avons mis en évidence certains aspects précoces liés à la lecture et à l'écriture, et peu développés dans les modèles vus précédemment. L'ensemble de ce chapitre avait donc pour objectif de poser quelques bases explicatives concernant lecture et écriture, tant d'un point de vue novice qu'expert ; le chapitre suivant aura pour objectif de présenter des facteurs externes à la construction de ces compétences mais qui peuvent toutefois influencer leur acquisition.

Résumé du Chapitre 2

Les modèles experts de la lecture et de l'écriture sont généralement envisagés comme ayant deux voies, une voie directe (correspondant au lexique mental) et une voie indirecte (assemblage des correspondances graphème-phonème). La qualité de la représentation lexicale en mémoire est importante en vue d'une récupération correcte et facilitée du mot.

Trois principes universels, communs à l'ensemble des langues, ont été présentés : il s'agit de la conscience linguistique, de l'identification des mots et de la compréhension. Ces trois principes sont indispensables à l'activité de lecture et d'écriture, et ce, quel que soit le système linguistique.

Plusieurs modèles développementaux de la lecture et de l'écriture ont été présentés ; néanmoins, les stades d'acquisition n'envisagent que tardivement dans le développement la compétence orthographique. Les modèles interactifs permettent le développement en parallèle des différentes acquisitions.

L'hypothèse d'auto-apprentissage a été présentée : elle postule que l'exposition répétée à la lecture via le décodage permet le stockage de la forme orthographique du mot (Share, 1995). Dans cette hypothèse, nous retrouvons également l'idée d'une voie directe et d'une voie indirecte.

Enfin, les modèles orthographiques présentés incluaient les compétences graphotactiques et morphologiques, avec l'idée selon laquelle le développement de l'écriture commence bien avant son enseignement formel.

Chapitre 3 : les facteurs de l'environnement scolaire impliqués dans l'apprentissage de la lecture et de l'écriture

Le chapitre précédent proposait une présentation synthétique de grandes théories explicatives de l'apprentissage de la lecture et de l'écriture. Dans ce troisième chapitre, nous allons nous attacher à présenter différents facteurs externes intervenant dans les performances des élèves en lecture et en écriture : nous allons pour cela considérer trois parties. Tout d'abord, les facteurs liés à l'enseignement, c'est-à-dire à l'environnement scolaire dans lequel l'élève est intégré. Nous présenterons ensuite des facteurs liés à la phonologie en début de scolarité, puis à l'habileté orthographique.

3.1 Des facteurs liés au type d'enseignement

S'il semble essentiel de chercher à comprendre comment l'enfant acquiert et développe ses compétences de lecture et d'écriture d'un point de vue cognitif, il est également nécessaire de prendre en compte les facteurs externes jouant un rôle important dans le développement de ces compétences. Il est par exemple communément admis que l'origine sociale des élèves est un facteur déterminant dans leur réussite scolaire, et particulièrement en France (MENJ-DEPP, 2019). Dans ce travail de thèse, nous ne nous centrerons pas sur cet aspect déjà largement documenté, mais sur d'autres facteurs qui concernent plus directement la classe et l'environnement scolaire de l'élève.

3.1.1 Effet-maitre et effet-classe

Lorsque l'on évalue les compétences des élèves, il est en effet utile de s'interroger sur les facteurs, dans la classe, qui pourraient impacter les résultats. Ainsi, des auteurs (e.g. Bressoux, 2012 ; Bianco et al., 2012) ont pu mettre en évidence le caractère non neutre de l'enseignant dans les progrès des élèves : c'est ce qu'on appelle l'effet-maitre. L'effet-classe, quant à lui, correspond aux facteurs qui peuvent également impacter les performances des élèves, mais qui ne relèvent ni des élèves, ni de l'enseignant : on y retrouvera, par exemple, la

composition de la classe, le niveau de classe (simple ou multiple) ou encore les caractéristiques qui composent le groupe classe dans son ensemble (hétérogène ou homogène par exemple).

La prise en compte de ces effets dans les expérimentations peut se faire notamment via l'utilisation de modèles statistiques multiniveaux : ces modèles vont ainsi pouvoir distinguer plusieurs niveaux dans l'analyse, afin de déterminer si une part significative de la variance est expliquée non seulement au niveau des élèves, mais également au niveau de la classe à laquelle l'élève appartient, soit en tant que groupe classe (pour mettre en évidence un effet-classe), soit en tant que conséquence des pratiques pédagogiques de l'enseignant (pour mettre en évidence un effet-maitre). Bianco et collaborateurs (2011), par exemple, ont utilisé des modèles multiniveaux de croissance pour mettre en évidence un effet significatif de la variance entre les élèves, en plus de la variance propre aux progrès de chacun des élèves dans une étude longitudinale de trois ans portant sur un entraînement à la compréhension en français. Notons que cette distinction entre effet-maitre et effet-classe est plus nette au collège, car il y a à ce moment-là plusieurs enseignants pour un même groupe classe, tandis qu'à l'école élémentaire, dans la majorité des cas, il y a un seul enseignant pour un même groupe classe, ce qui rend la distinction entre ces deux effets plus difficile au niveau élémentaire. L'impact que cet effet généré par le maitre ou la classe peut avoir sur les performances des élèves peut toutefois varier suivant la matière concernée (e.g. Goigoux, 2016, pour ces différences en fonction du type de compétences évaluées).

3.1.2 Effet des modalités d'enseignement

L'enseignant peut donc avoir un rôle déterminant dans le développement des apprentissages des élèves ; il s'agit alors d'étudier les différentes méthodes d'enseignement pour comparer leur efficacité. Sur ce point, l'enseignement explicite fait partie des modalités d'enseignement qui semblent particulièrement efficaces. En effet, de nombreux travaux montrent que l'enseignement explicite est une pratique pédagogique favorisant les apprentissages des élèves, et ce, quelle que soit la matière scolaire concernée (compréhension de textes, mathématiques, orthographe...). Ces travaux (voir pour exemple la méga-analyse de Bissonnette et al., 2010), initiés aux Etats-Unis il y a plusieurs décennies déjà, mettent en avant l'utilité de la pédagogie explicite, tant pour les apprentissages simples que pour les apprentissages plus complexes.

L'enseignement explicite fonctionne sur la base d'un étayage continu dans le processus d'apprentissage, en utilisant des énoncés clairs, des exemples effectués par l'enseignant, et le suivi des élèves grâce à des feedbacks jusqu'à la maîtrise du contenu. Plus spécifiquement, une séance construite selon un enseignement explicite comporte différentes étapes qui permettent de la structurer (Bressoux, 2022 ; voir également Bocquillon, 2020 pour une présentation détaillée) :

- La mise en route de la séance, qui permet de mettre en évidence les objectifs de la séance, en présentant les notions essentielles, et en reprenant les notions vues précédemment ;
- Le modelage, phase dans laquelle c'est l'enseignant qui présente la notion à apprendre, en la manipulant devant la classe, et en détaillant toutes les étapes nécessaires pour l'assimilation de cette notion à haute voix ;
- La pratique guidée, où les élèves font le travail tandis que l'enseignant les accompagne dans le travail en le faisant avec eux ;
- La pratique autonome, où les élèves réalisent la tâche seuls ;
- La clôture de la séance, où l'enseignant fait avec les élèves une synthèse des éléments essentiels à retenir vus au cours de la séance.

Des procédures en cinq ou six étapes (Hughes et al., 2017) sont généralement proposées pour expliquer le déroulement des séances pour ce type de pédagogie. L'enseignement explicite est donc une manière de construire et de mener l'enseignement dans laquelle les élèves sont activement impliqués dans la transmission du contenu afin qu'ils apprennent à le comprendre et le manipuler pour mieux l'assimiler : c'est le principe de l'auto-régulation, qui permet alors aux élèves d'avoir conscience des procédures à mettre en œuvre dans la production de l'écrit. Cette régulation faite par l'élève sur sa propre production lui permet de la questionner et donc de devenir acteur de son raisonnement au regard de ce qui aura été mis en place au cours des séances.

Dans leur méga-analyse, Bissonnette et collaborateurs (2010) ont ainsi permis de mettre en évidence l'amélioration des performances des élèves apportée par l'utilisation d'une pédagogie explicite, sur des méta-analyses essentiellement réalisées aux Etats-Unis avant les années 2000. La pertinence de cette pratique pédagogique quelle que soit la matière étudiée (lecture, écriture, mathématiques pour les études mentionnées) permet de justifier l'intérêt de la prise en compte,

lors des analyses statistiques, du potentiel effet exercé par l'enseignant et par les pratiques pédagogiques qu'il met en place au sein de sa classe.

3.2 Les facteurs liés aux méthodes d'enseignement du code en début d'apprentissage

Après avoir présenté différentes manières d'envisager la construction de l'apprentissage de la lecture et de l'écriture, puis les paramètres externes à l'apprenant qui entrent en jeu dans cet apprentissage, nous allons à présent présenter plusieurs études, internationales et francophones, qui se sont intéressées à l'enseignement spécifique de la méthode phonique en début d'apprentissage, et à son effet sur les performances ultérieures des élèves, notamment dans l'acquisition des compétences de lecture et d'écriture. La méthode dite phonique correspond à l'enseignement du code, c'est-à-dire des correspondances grapho-phonologiques, de manière explicite et exhaustive.

Une méta-analyse conduite par Ehri et collaborateurs (2001b) pour le *National Reading Panel* aux Etats-Unis a regroupé 38 études afin de tester les effets d'une instruction phonique systématique comparativement à une instruction non systématique (incluant une instruction phonique et non phonique) ou non phonique. Les données ont été recueillies auprès d'élèves allant de l'équivalent de la maternelle à l'équivalent de la sixième. Cette méta-analyse a mis en évidence un effet modéré ($d=0.41$) de la méthode phonique systématique, qui s'est montrée efficace sur le décodage, la lecture de mots, la compréhension de textes et l'écriture. Les auteurs ont montré que si la méthode était mise en place dès la maternelle, sa taille d'effet augmentait. Par ailleurs, cette méthode semble bénéfique pour les élèves venant de milieu socio-économique faible ou modeste, ainsi que pour les élèves considérés comme étant à risque de développer des troubles des apprentissages ; cependant, cette méthode n'a pas été efficace pour les élèves présentant déjà un profil de faibles lecteurs. Il semblerait donc que la méthode phonique soit plus efficace dès la maternelle, c'est-à-dire avant que les élèves aient déjà mis en place des stratégies pour apprendre à lire, d'où la présence d'un effet plus important au début (de l'équivalent de la maternelle ou du CP) qu'à la fin des études (de l'équivalent du CE2 à l'équivalent de la sixième). Concernant l'écriture, il semblerait que l'effet de l'instruction phonique soit visible principalement en début d'apprentissage, mais ne soit plus efficace par la suite, même chez les élèves plus âgés faibles lecteurs.

Sprenger-Charolles et Colé (2013) se sont intéressées aux résultats de la méta-analyse anglo-saxonne proposée par Ehri et collaborateurs (2001b), en ajoutant à leur synthèse les méta-analyses réalisées sur les méthodes d'enseignement de la lecture en français. Ainsi, elles présentent trois études francophones sur la supériorité de la méthode phonique sur les autres méthodes d'enseignement de la lecture. La première étude est celle de Braibant et Gérard (1996, dans Sprenger-Charolles & Colé, 2013), qui a comparé les effets d'une méthode phonique par rapport à une méthode idéo-visuelle en Belgique, dans des classes équivalentes au CE1 français sur les performances en décodage et en compréhension écrite. Les résultats ont montré que l'approche phonique favorisait les performances à la fois en décodage et en compréhension écrite, sans disparité majeure entre les deux tâches, contrairement à la méthode idéo-visuelle ; également, l'approche phonique permettait aux élèves issus de milieu socio-économique plus faibles d'améliorer leurs performances dans les deux tâches de manière plus importante que l'approche idéo-visuelle. Goigoux (2000, dans Sprenger-Charolles & Colé, 2013) a proposé une étude similaire pour des enfants suivis de la grande section de maternelle jusqu'à l'équivalent du CE2. Il a comparé les scores obtenus par les élèves pour chacune des méthodes à la moyenne nationale, et a mis en évidence la supériorité de la méthode phonique pour les performances en compréhension écrite, mais surtout le fait que les performances en compréhension écrite des élèves ayant suivi une méthode idéo-visuelle étaient inférieures à la moyenne nationale en CE2. De la même manière que pour l'étude de Braibant et Gérard (1996), les résultats ont montré un effet bénéfique de la méthode phonique pour les élèves issus de milieux socio-économiques défavorisés ou peu favorisés. Une troisième étude menée par Gentaz et collaborateurs (2013, dans Sprenger-Charolles & Colé, 2013) est également présentée. Cette étude ne s'est pas concentrée directement sur les méthodes d'enseignement, mais a proposé un entraînement à un groupe d'élèves sur trois points : la correspondance graphème-phonème, la fluence de lecture de texte et la compréhension, à des élèves en fin de CE1 d'écoles situées en zone d'éducation prioritaire. Les résultats ne vont pas dans le sens des deux études présentées précédemment : en effet, il n'y a pas d'amélioration significative des performances des élèves qui ont suivi l'entraînement. Cependant, les auteurs expliquent cette absence de résultats par différents facteurs : le court laps de temps en amont de l'étude, ou encore le fait que les enseignants aient conservé leur matériel habituel d'enseignement, qui peut déjà parfois inclure des exercices sur les habiletés travaillées théoriquement uniquement dans le groupe ayant suivi l'entraînement.

Les résultats présentés par Sprenger-Charolles et Colé (2013) viennent donc globalement corroborer les effets mis en évidence dans la méta-analyse présentée par Ehri et collaborateurs (2001b). L'apport de la méthode phonique en tout début d'apprentissage semble donc aussi bénéfique pour les élèves anglophones que pour les élèves francophones. Cette méthode semble également favoriser le développement des capacités de décodage pour tous les enfants, et notamment pour les enfants issus de milieux plutôt défavorisés. Ces éclairages scientifiques sur l'efficacité d'une approche phonique pour l'enseignement de la lecture et de l'écriture ont été corroborés récemment par Treiman (2018), qui avance également l'idée selon laquelle l'approche phonique permet de faire comprendre aux élèves apprenants le fonctionnement de leur système orthographique. Elle nuance toutefois ce bénéfice au regard de la complexité orthographique de certaines langues : dans ces cas-là, l'explication par la méthode phonique ne serait pas suffisante pour comprendre le fonctionnement de la langue.

3.3 Les facteurs liés à l'orthographe

3.3.1 Les orthographe approchées en période préscolaire

Un courant de recherche sur l'étude du développement des compétences orthographiques s'est intéressé aux pratiques précoces liées à l'apprentissage de l'écriture, c'est-à-dire, avant le début de l'école élémentaire, notamment avec les orthographe approchées. Cette appellation regroupe les pratiques d'écriture préscolaires et correspond à la sensibilité précoce des enfants pour l'écriture ; les orthographe dites « inventées » (mais qui ne correspondent toutefois pas à une création dépourvue de sens) seraient donc des indices « de la façon dont les enfants « pensent » une orthographe » (Jaffré & Morin, 2008, p.191).

McBride-Chang (1998) a testé des élèves de maternelle étatsuniens pendant un an et demi, à quatre temps de mesure, afin de vérifier le poids de ce qu'elle appelle les orthographe inventées dans les apprentissages ultérieurs. Ses résultats ont montré que les orthographe inventées sont une mesure stable dans le temps, et qui est fortement liée aux tâches de conscience phonologique. En plus de cela, les performances obtenues par les jeunes élèves à cette tâche d'orthographe inventées était significativement liée aux performances en décodage de mots et de pseudomots, ainsi qu'aux résultats obtenus aux tests standardisés d'orthographe.

Les orthographes approchées ont aussi été mesurées via des entraînements pour vérifier leur potentiel effet bénéfique pour des élèves « à risque » en ce qui concerne les habiletés préscolaires de lecture et d'écriture. Ainsi, Morin et Montésinos-Gelet (2007) ont proposé à 90 élèves de dernière année de maternelle un programme d'entraînement grâce à la mise en place d'une pédagogie basée sur les principes didactiques des orthographes approchées. Les résultats de ce programme ont montré des effets positifs et significatifs sur les habiletés de décodage pour les élèves considérés à risque sur un ou plusieurs facteurs en maternelle seulement, et plus en CP. De plus, les effets de l'entraînement ont également été significatifs concernant les performances orthographiques (mesure d'extraction phonologique) à la fois en maternelle et en CP, et ce, que les enfants aient un seul ou plusieurs risques préalablement identifiés. Pulido et Morin (2016) ont proposé une méta-analyse sur quatorze recherches s'intéressant à l'accompagnement des premières écritures en maternelle ; ces recherches étaient basées sur des systèmes d'écriture différents, dont la consistance de l'orthographe variait. Cependant, elles ont mis en avant un effet bénéfique de ce type d'accompagnement sur les habiletés phonologiques des élèves (notamment sur la capacité à segmenter les mots en syllabes). Un effet positif a également été démontré sur la capacité des élèves à orthographier des mots entendus.

Les études sur les orthographes approchées en milieu préscolaire permettent ainsi de mettre en évidence l'importance de la mise en place des compétences précoces, avant même le début de l'instruction formelle au CP. Ces compétences sont en lien avec les habiletés phonologiques utilisées ultérieurement en milieu scolaire.

3.3.2 Le cas de l'écriture sous la dictée

Si les performances des élèves en orthographe ont diminué avec le temps (voir Introduction), la question est de savoir si certaines pratiques enseignantes pourraient au contraire favoriser les progrès des élèves. Traditionnellement, l'évaluation de l'orthographe en classe se faisait sous forme d'une dictée, pour laquelle la correction pouvait donner des informations sur le type d'erreur commise par l'élève ; or la question se pose dorénavant de la pertinence de ce type d'exercice, (voir par exemple Brissaud et al., 2011, pour des pistes de réflexion sur l'enseignement de l'orthographe) qui mobiliserait davantage des capacités mnésiques que des raisonnements sur les spécificités orthographiques des mots. Cette question

est centrale dans l'apprentissage du français pour lequel l'orthographe lexicale et grammaticale restent complexes.

Le Brun et collaborateurs (2016) ont testé l'effet d'un entraînement à la dictée utilisant un calque sur les performances en dictée classique d'élèves de sixième. Au total, 121 élèves ont participé à cette étude, composée d'un groupe contrôle et d'un groupe expérimental. Ces groupes étaient constitués par les classes de chacun des deux établissements participant à l'étude : la moitié constituait le groupe contrôle d'une école, l'autre le groupe expérimental. Les élèves étaient également associés à trois groupes selon leur score obtenu en orthographe à une dictée pré-test : le groupe des très faibles orthographes, le groupe des faibles et moyens orthographes, et le groupe des bons orthographes. Les élèves suivaient une phase d'entraînement pendant huit semaines ; les performances étaient mesurées grâce à trois dictées proposées avant le début de l'entraînement, après quatre semaines d'entraînement puis après la dernière semaine d'entraînement. Durant la phase d'entraînement, une dictée était proposée par l'enseignant chaque semaine : une dictée classique pour le groupe contrôle, et une dictée avec calque pour le groupe expérimental. Le calque permet de donner des indications sur le nombre de lettres qui composent le mot dicté ainsi que sur leur forme globale ; cela permet d'aider les élèves à produire un travail réflexif sur l'orthographe qu'ils produisent pendant la dictée, notamment en leur permettant de voir s'ils ont oublié de prendre en considération certains accords (comme le pluriel d'un mot ou d'un verbe). En effectuant cet exercice une fois par semaine pendant les huit semaines d'entraînement, le groupe expérimental a eu des résultats significativement meilleurs au post-test comparativement au groupe contrôle équivalent. Les chercheurs notent toutefois que ce progrès n'est avéré que pour le groupe des élèves moyens-faibles en orthographe, les résultats n'ayant pas montré d'effet significatif pour le groupe des élèves forts ainsi que pour celui des élèves très faibles. Le calque pourrait donc permettre la réactivation de certains mécanismes de réflexion sur le contenu (métacognition) principalement chez les élèves qui ont déjà acquis certaines connaissances à ce sujet.

Au Québec, Fisher et Nadeau (2014) ont mis en place une évaluation dont le but était d'étudier les effets de la pratique régulière de dictées dites innovantes sur les compétences orthographiques des élèves de 26 classes d'école primaire. Deux types de dictées innovantes ont été proposées : la dictée zéro faute et la phrase dictée du jour. La dictée zéro faute consiste en une dictée de taille moyenne de plusieurs phrases. A la suite de chaque phrase, les élèves peuvent poser toutes leurs questions concernant les doutes orthographiques qu'ils peuvent avoir

sur la phrase énoncée. Le rôle de l'enseignant est alors de guider au besoin les réflexions des élèves afin de les amener à résoudre leurs problèmes collectivement. Une fois la réflexion collective terminée, l'enseignant résume le raisonnement généré par l'ensemble de la classe et valide la graphie attendue. La phrase dictée du jour se base sur la dictée d'une phrase à l'ensemble de la classe, puis à l'affichage au tableau des différentes graphies proposées par les élèves pour chaque mot la composant. S'ensuit une discussion collective de la classe pour choisir, pour chacun des mots, l'orthographe correcte, en justifiant les choix et en arrivant à un accord unanime sur la graphie finale à valider. L'enseignant, à la fin de la discussion, valide le résultat et les élèves peuvent recopier la phrase correcte.

Le principe des dictées innovantes est donc centré sur la verbalisation par les élèves de leurs réflexions concernant les choix orthographiques à faire, et la confrontation avec l'ensemble du groupe. L'interaction avec le groupe va pouvoir permettre aux élèves de prendre conscience de leur raisonnement et de pouvoir le modifier si besoin, via la confrontation avec le raisonnement des autres et l'actualisation des représentations qu'ils en avaient si elles étaient incorrectes ou incomplètes et ne leur permettait pas d'accéder au choix orthographique attendu. Une fois cet exercice ritualisé, l'élève sera capable de faire ce raisonnement dans sa tête pour la suite de sa scolarité : cela lui permet alors d'être en capacité de réfléchir aux procédures qu'il a utilisées, et de contrôler ses apprentissages en ayant conscience (principe d'auto-régulation). Selon les autrices, l'avantage pour les enseignants de ce type de dictée est qu'il leur permet d'avoir accès aux représentations mentales des élèves, ce qu'ils ne peuvent pas faire lors d'une dictée classique : cela leur donne donc des indications concrètes sur le raisonnement des élèves, et leur permet de pouvoir les travailler avec eux en classe. Cette étude menée par Fisher et Nadeau (2014) a mis en évidence l'effet bénéfique au-delà de la norme en ce qui concerne les performances en orthographe grammaticale de toutes les classes. En l'absence d'un groupe contrôle, les progrès des élèves ont en effet été comparés à ceux obtenus par les élèves à la dictée de Manesse et Cogis (2007), qui servaient alors de « norme » de progression.

3.3.3 L'effet du temps passé sur les activités liées à l'écrit

La recherche Lire et écrire au CP (Goigoux, 2016, voir aussi Brissaud & Fayol, 2018 pour une synthèse sur l'écrit) s'est intéressée aux progrès des élèves réalisés pendant le CP et le CE1 sur de nombreuses habiletés de lecture et d'écriture. En ce qui concerne la production orthographique, différentes épreuves, en début et en fin de CP, ont été utilisées pour tester les

performances des élèves : dictée du prénom, dictée de mots (*lapin, rat, éléphant*), dictée de phrases (*Tom joue avec le rat, les lapins courent vite*, cette dernière ayant été ajoutée en fin de CP), ou encore production écrite d'un texte à partir d'images. Des paramètres extérieurs ont également été mesurés : niveau socio-économique, langues parlées à la maison, genre, date de naissance, etc., ainsi que des variables liées à la classe ou à l'enseignant, comme le temps passé en moyenne sur chaque compétence. Les résultats montrent, concernant l'effet de l'écriture sur la lecture, que plus le temps consacré à l'écriture sous la dictée augmente, plus les performances en lecture-décodage de l'élève augmentent également. Ces résultats sont semblables en ce qui concerne le temps consacré à l'encodage autonome. On observe toutefois un effet quadratique du temps sur ces performances : au-delà d'un certain temps, il n'y a plus de progression observée chez les élèves. L'écriture sous la dictée a aussi des effets bénéfiques sur les performances en écriture, de la même manière que la tâche de production autonome et que les tâches de planification et de révision des écrits produits : plus les élèves y consacrent du temps, meilleures sont leurs performances en fin de CP. De manière générale, cette recherche montre que le temps passé à faire travailler les élèves sur des tâches mobilisant l'écriture et la morphologie (pratiques réflexives autour de l'écrit) est bénéfique pour la progression des élèves dans tous les domaines évalués, à l'exception de la compréhension orale.

A travers ce troisième chapitre théorique nous avons donc pu présenter, de manière plus concrète, les facteurs qui entrent également en jeu dans les performances des élèves et dans leur progression, mais qui ne sont pas directement liés au fonctionnement cognitif des élèves : il s'agit des effets liés à la classe, à l'enseignant, ainsi que les pratiques pédagogiques utilisées par ce dernier au sein de sa classe. Nous avons également présenté brièvement des études expérimentales sur l'efficacité d'une méthode d'enseignement en début d'apprentissage formel, à savoir l'approche phonique. Plusieurs méta-analyses (e.g. Bissonnette et al., 2010 ; Ehri et al., 2001b), qu'elles soient anglophones ou francophones, ont ainsi permis de mettre en évidence l'apport non négligeable de la méthode phonique très tôt dans la scolarisation des enfants, qui permet ainsi l'amélioration des performances en lecture et en compréhension, et ce, quel que soit le milieu socio-économique dont les élèves sont issus. Nous nous sommes ensuite centrés sur l'enseignement de l'orthographe de manière plus spécifique, afin de mettre en évidence l'intérêt de certaines pratiques pédagogiques dites innovantes en ce qui concerne l'exercice

d'écriture sous la dictée. Nous avons finalement terminé en rappelant l'aspect bénéfique du temps consacré aux activités liées à l'écrit en classe, notamment en début de scolarité. Ainsi, au-delà des capacités des élèves et de leur apprentissage direct, de nombreux autres paramètres sont à prendre en considération lors de l'évaluation des compétences en lecture et en écriture.

Une fois ces facteurs externes présentés, quelles habiletés, directement liées aux élèves et à leurs apprentissages, peuvent favoriser le développement des compétences de lecture et d'écriture tout au long de l'école élémentaire ? Le prochain chapitre proposera une esquisse de réponse à travers la présentation d'habiletés connues pour leur impact sur les performances en lecture et en écriture. Nous nous intéresserons ensuite davantage à l'aspect longitudinal sous-tendant ces apprentissages à l'école élémentaire.

Résumé du Chapitre 3

Dans ce chapitre, nous avons tenté de mettre en évidence l'importance des facteurs liés à l'environnement de la classe et à l'enseignant dans l'acquisition de la lecture et de l'orthographe. L'enseignement explicite par exemple est une pratique pédagogique qui favorise les apprentissages scolaires. Également, les effets bénéfiques de la méthode phonique en début d'apprentissage, en français comme en anglais, ont été montrés.

Au niveau préscolaire, le travail sur les orthographes approchées semble bénéfique pour les performances ultérieures en décodage et en orthographe. Les dictées innovantes, via l'utilisation de calque ou de type zéro faute, favorisent la réflexivité sur l'écrit et permettent également l'amélioration des performances orthographiques ultérieures. Enfin, le temps consacré au travail sur la compétence orthographique est également un facteur à prendre en considération.

Chapitre 4 : les liens entre lecture et écriture

Nous avons pu voir dans les chapitres précédents que certaines capacités étaient fondamentales pour l'acquisition et le développement de la lecture et de l'écriture, et largement mises en avant dans les modèles théoriques proposés par les chercheurs. Nous allons dans ce dernier chapitre théorique présenter les principaux prédicteurs connus de la lecture et de l'écriture dans un premier temps ; nous introduirons ensuite une synthèse d'études longitudinales sur ce lien lecture-écriture, sur le temps de l'école élémentaire.

4.1 Les prédicteurs de la lecture et de l'écriture à l'école élémentaire

Castles et collaborateurs (2018) proposent une synthèse sur les connaissances actuelles sur l'acquisition de la lecture : le décodage étant à la base de l'apprentissage de la lecture pour des systèmes alphabétiques tels que le nôtre, son importance dans l'évolution vers l'expertise du lecteur va laisser place, au fil du temps, à des capacités de plus haut niveau permettant une lecture plus rapide et automatisée (Castles & Nation, 2006). On retrouve ici notamment les théories principales sur le développement des habiletés de fluence de lecture de mots, telles que l'hypothèse d'auto-apprentissage de Share (1995, voir Chapitre 2), ainsi que l'hypothèse de qualité lexicale (Perfetti, 1992, 2007, voir Chapitre 2) en lien avec l'expérience de l'écrit. Trois types d'habiletés sont mises en avant pour leur impact sur les performances en lecture et en écriture des élèves en début d'apprentissage : il s'agit des habiletés liées au code, des habiletés métalinguistiques, et des habiletés cognitives générales.

Les habiletés liées au code regroupent les habiletés directement liées à la lecture, comme le décodage, la fluence de lecture ou encore la connaissance des lettres. Le décodage, comme nous l'avons vu précédemment avec le modèle de l'auto-apprentissage, est un support essentiel au développement de la lecture (e.g. Share, 1995), de la compréhension (modèle simple de la lecture, Gough & Tunmer, 1986) et est également présent dans de nombreux modèles théoriques comme support au développement de l'écriture via la voie non lexicale (e.g. modèles double voie, voir Chapitre 2). Juel (1988) a mené une étude longitudinale sur les apprentissages

des élèves en lecture et en écriture et s'est intéressée aux élèves qui, après la quatrième année d'enseignement formel, présentaient un schéma de faible lecteur ou faible scripteur. Les faibles lecteurs, en quatrième année, avaient eu des faibles performances en décodage au début de l'étude ; même si d'autres facteurs entraient également en jeu, l'objet principal de l'étude était de tester les élèves sur les habiletés mises en avant dans le modèle simple de la lecture et le modèle simple de l'écriture. Cet effet du décodage précoce sur les performances ultérieures en lecture a aussi été mis en évidence à travers une étude sur le décodage CP et ses conséquences sur les performances en lecture CM1 (Sprenger-Charolles et al., 2003, cité dans Sprenger-Charolles, 2017).

La connaissance du nom des lettres est admise comme étant un prédicteur préscolaire majeur pour les acquisitions ultérieures en lecture (voir Foulin, 2005, pour une revue). Foulin et Pacton (2006) précisent que cela peut s'expliquer par le fait que connaître le nom d'une lettre donne déjà à l'enfant des indications sur sa prononciation et donc sur le son contenu. La connaissance préscolaire du nom des lettres participe donc à la fois à préparer les bases du décodage grâce au lien qu'il commence à créer entre la lettre reconnue et le son qui lui est associé. Dans leur étude, Foulin et Pacton (2006) ont testé des enfants de moyenne section de maternelle avec l'hypothèse selon laquelle les enfants qui ont un score élevé en connaissance du nom des lettres auraient de meilleures performances en identification de pseudomots, en se basant sur l'idée qu'une bonne connaissance du nom des lettres favoriserait l'accès à la prononciation de pseudomots grâce aux indices fournis par le son des lettres connues. Leur hypothèse est bien validée, et permet ainsi de mettre en évidence l'intérêt de certaines habiletés préscolaires sur les acquisitions ultérieures. Dans cette perspective, Caravolas et collaborateurs (2001) ont mené une étude longitudinale sur les acquisitions des élèves entre la première et la troisième année d'enseignement formel. Les auteurs ont mis en évidence un lien entre les performances en connaissance du nom et du son des lettres en début de scolarité et les performances en lecture de mots isolés (mots fréquents dans les manuels scolaires pour moitié, et mots fréquents pour l'âge de référence pour autre moitié) au temps ultérieur ; cet effet est d'ailleurs bidirectionnel, les performances en lecture au temps 1 prédisant les performances en connaissance du nom et du son des lettres au temps 2. Cependant, l'effet prédictif de la connaissance du nom et du son des lettres n'était plus significatif aux temps ultérieurs.

La fluence de lecture correspond à l'automatisation du processus de décodage et à sa qualité. Cette habileté se mesure généralement via le nombre de mots correctement lus par

minute, et peut se faire à haute voix ou en lecture silencieuse (voir Fuchs et al., 2001 pour une synthèse sur la fluence de lecture à haute voix, par exemple). Goigoux et collaborateurs (2016) ont mesuré les effets de plusieurs facteurs considérés comme prédicteurs des performances en lecture en début d'apprentissage, dont la fluence, via une mesure de lecture de mots chronométrée en contexte chez les jeunes enfants (CP). La corrélation entre la fluence de lecture en contexte et la lecture de mots isolés était de 0.65, montrant donc l'intérêt d'étudier ces deux variables de manière séparée dans les modèles statistiques. Après avoir testé un premier modèle qui correspondait strictement au modèle simple de la lecture, avec donc comme variables le décodage et la compréhension orale, les auteurs ont testé un second modèle avec des variables supplémentaires : fluence de lecture en contexte, phonologie, vocabulaire, raisonnement non verbal. Ce second modèle avait un meilleur pouvoir prédictif que le premier (il expliquait 43.5% des performances en lecture contre 37.5% pour le premier modèle). Également, la fluence avait un effet significatif sur les performances en lecture en CP. Ainsi, même si les deux prédicteurs du modèle simple (décodage et compréhension orale) avaient le plus fort pouvoir explicatif, d'autres prédicteurs, dont la fluence en contexte, permettaient aussi d'expliquer une part significative des performances des élèves de CP. Plus tard dans le parcours scolaire, Tilstra et collaborateurs (2009) ont investigué les prédicteurs de la compréhension écrite à différents temps : 271 élèves étatsuniens de Grade 4, Grade 7 et Grade 9 (respectivement l'équivalent du CM1, cinquième et troisième) ont participé à l'étude. Les résultats ont montré que la fluence de lecture (épreuve de type *maze test*) expliquait une part significative des performances en compréhension écrite. Cette étude venait infirmer les résultats trouvés par Adlof et collaborateurs (2006), qui avaient également testé une version du modèle simple en y ajoutant un facteur fluence, mais n'avaient trouvé aucun effet (en Grade 2, Grade 4 et Grade 8). Toutefois, la mesure utilisée dans cette étude pour la fluence était de la fluence de mots isolés (lecture de mots chronométrée), et non de la fluence en contexte comme c'était le cas pour Tilstra et collaborateurs (2009).

La mise en place de ces habiletés précoces, particulièrement le décodage et la connaissance du nom et du son des lettres va permettre, lors de l'entrée à l'école élémentaire, de favoriser le développement d'autres habiletés, et notamment des habiletés métacognitives, à l'instar de la conscience phonologique. La conscience phonologique consiste en la capacité à reconnaître et manipuler les différentes unités sonores d'une langue, à savoir la rime, la syllabe et le phonème. Pour les langues utilisant un système d'écriture alphabétique, cette

compréhension des sons du langage est une compétence nécessaire pour l'acquisition de la lecture. Ainsi, une conscience phonologique développée permettra l'amélioration des performances en lecture (Sprenger-Charolles, 2017). Ecalte et collaborateurs (2022) ont présenté les résultats d'une étude récente sur les compétences en littéracie et en lecture chez des enfants suivis entre la moyenne section de maternelle et le CP. Pour évaluer l'effet des habiletés phonémiques, une tâche de détection d'intrus et une tâche de syllabe étaient proposées en moyenne section, et une tâche de segmentation phonémique en CP. Les résultats montrent que les habiletés en littéracie en maternelle (habiletés phonémiques, connaissance du nom des lettres, vocabulaire) prédisent largement les habiletés liées à la lecture en CP. Desrochers et collaborateurs (2009) précisent que la conscience phonologique est une habileté composée de plusieurs niveaux de difficulté ; en effet, travailler sur la manipulation de la syllabe ou du phonème n'est pas de la même complexité que d'identifier une syllabe ou un phonème dans un mot ; c'est pourquoi ils font la distinction entre les habiletés épilinguistiques (concernant la reconnaissance des unités sonores) et les habiletés métalinguistiques (plus complexes, car elles concernent la manipulation consciente de ces unités). Les auteurs résumant également dans cette synthèse l'importance de la conscience phonologique pour l'amélioration des performances en lecture, en anglais comme en français, et affirment également qu'une bonne connaissance des lettres de l'alphabet avant la scolarisation favorise d'une part les acquisitions en lecture, et d'autre part le développement de la conscience phonologique. Ils montrent également un effet bidirectionnel, entre conscience phonologique et connaissance des lettres, ces deux compétences favorisant donc le développement l'une de l'autre. Ehri et collaborateurs (2001a) ont conduit une méta-analyse sur 52 études internationales évaluant l'effet d'un enseignement plus spécifiquement centré sur la conscience phonémique en début d'enseignement formel sur des enfants de l'équivalent de la maternelle à l'équivalent de la sixième. Plusieurs langues étaient étudiées, comprenant donc divers degrés de transparence. Les chercheurs ont ainsi mis en évidence plusieurs paramètres pour lesquels les entraînements à la conscience phonémique ont eu un effet significatif. Tout d'abord, l'effet de ces entraînements à la conscience phonémique a un impact fort ($d= 0.86$) sur les progrès des élèves dans cette habileté spécifiquement ; ce travail sur la conscience phonémique, comparé grâce à des groupes contrôles, a permis également des progrès dans les performances en lecture de mots et en compréhension. Des effets plus modérés ont été aussi notés sur les performances en écriture.

Les habiletés cognitives générales ont également de l'importance dans l'apprentissage de la lecture et de l'écriture. Desrochers et collaborateurs (2012) ont recensé différents facteurs de risque influant sur le développement de l'acquisition de la lecture chez les enfants. Sur le plan cognitif, et en dehors de tout trouble ou handicap connu affectant les fonctions cognitives liées à la lecture, ils distinguent plusieurs facteurs : la mémoire et la dénomination rapide. En effet, l'activité de lecture, tout comme celle d'écriture, ne mobilisent pas que les habiletés liées à l'écrit et à la compréhension : pour comprendre un texte ou pour lire, l'enfant doit être capable de se concentrer sur le contenu, de mobiliser ses connaissances (voir Eme & Rouet, 2002, pour une synthèse). Cela est coûteux en terme de ressources cognitives, et la variation de la mobilisation (ou de l'automatisation) de ces ressources peut donc impacter directement l'activité de lecture ou d'écriture. Plusieurs auteurs ont ainsi cherché à comprendre l'impact de ces capacités cognitives pour estimer leur poids dans les apprentissages de la lecture et de l'écriture. La dénomination rapide (*Rapid Automated Naming*, ou RAN) est une habileté liée à la reconnaissance rapide d'items vus par le sujet. Or, la lecture est par nature une activité basée sur le traitement d'une information écrite, et donc, visuelle. Envisager la dénomination rapide comme prédicteur de l'apprentissage de la lecture s'appuie donc sur l'idée que la vitesse à laquelle le lecteur va identifier l'information qui lui est présentée va avoir un effet sur ses capacités de lecture. Wolf, Bally et Morris (1986) ont testé le pouvoir prédictif du RAN dans une étude longitudinale, de la dernière année de maternelle au CE2 (équivalents), sur 83 enfants. Les auteurs ont comparé le RAN alphanumérique (dénomination de chiffres et de lettres) et le RAN non-alphanumérique (dénomination de couleurs et d'objets). En début d'apprentissage, les deux types de RAN prédisaient les performances de l'ensemble des épreuves proposées pour évaluer les performances en lecture ; cependant, avec le temps, seul le RAN alphanumérique influençait les performances des élèves, et uniquement sur les tâches de lecture orale et lecture de mots isolés. Ziegler et collègues (2008) ont mené une étude pour vérifier le lien entre RAN et lecture selon plusieurs hypothèses. Ils ont mis en avant trois facteurs pouvant expliquer l'impact du RAN sur les performances en lecture chez les jeunes enfants normo-lecteurs et dyslexiques : la fenêtre visuo-attentionnelle (dont l'effet sur les performances en lecture a depuis été démontré par Bosse et Valdois, 2009), les facteurs phonologiques et l'automatisation du RAN notamment alphanumérique.

Enfin, la mémoire, et plus spécifiquement la mémoire de travail, a été considérée comme un facteur important prédictif de la réussite en lecture. En effet, la mémoire de travail, soit la capacité à garder une information de manière temporaire en mémoire, dans le but de la manipuler, est fortement sollicitée pour toute activité de lecture dont l'objectif est la compréhension d'un texte écrit, par exemple. Des études ont été réalisées afin de vérifier ce lien notamment avec les travaux de Daneman et Carpenter (1983), qui avaient testé le lien, chez les adultes, entre mémoire de travail et compréhension. Dans deux études, les auteurs ont testé l'empan de mémoire de travail et la détection d'inconsistances en lecture à voix haute d'une part, et en lecture silencieuse d'autre part. Les résultats ont montré que les lecteurs dont l'empan de mémoire de travail était large avaient un meilleur pourcentage d'interprétation correcte du texte lu, et d'autant plus s'il était lu à haute voix. Gathercole et Pickering (2000) ont testé sur 87 enfants de niveau équivalent au CP l'effet de la mémoire de travail sur les performances scolaires. Trois catégories de mémoire ont été évaluées (boucle phonologique, exécutif central, mémoire visuo-spatiale) en lien avec le vocabulaire, les mathématiques, la reconnaissance des mots et la compréhension écrite. Les scores obtenus aux épreuves correspondant à la boucle phonologique ont montré des liens significatifs avec les performances sur les trois types d'habiletés testées en CP, et sur le vocabulaire en CE1. Mais seul le lien avec le vocabulaire (en CP et CE1) reste prédictif une fois l'exécutif central et l'âge pris en compte. L'exécutif central quant à lui restait prédictif des trois types d'habiletés au premier temps de mesure.

Toutefois, si l'importance de ces habiletés a bien été démontrée en début d'acquisition voire avant l'entrée à l'école élémentaire et le début de l'enseignement formel de la lecture et de l'écriture, il semble y avoir un effet du temps, au fur et à mesure que l'enfant avance vers l'expertise en lecture. En effet, si l'objectif de la lecture est la compréhension, le poids des habiletés évoquées précédemment peut évoluer en fonction de leur automatisation et de l'évolution vers de nouvelles compétences, de plus haut niveau (voir Bianco, 2015, p.90 pour une illustration). Des habiletés de langage oral, telles que le vocabulaire ou la compréhension orale semblent également avoir une influence sur les performances.

Une étude longitudinale a été proposée par Storch et Whitehurst (2002) pour évaluer l'évolution des habiletés prédictives, et notamment les habiletés de langage oral, de l'équivalent de la maternelle au CM1 (*preschool* à Grade 4). Leur objectif était triple : d'une part, étudier le lien entre les habiletés liées au code, les habiletés langagières et la compétence ultérieure en

lecture ; d'autre part, évaluer les effets de plusieurs habiletés émergentes sur les performances en lecture ; enfin, étudier le potentiel changement entre les facteurs au fur et à mesure de l'apprentissage, en ayant par exemple un seul facteur en CP mais deux à la fin de l'étude. Les habiletés de langage oral, qui nous intéressent particulièrement ici, consistaient en une épreuve de vocabulaire réceptif, une épreuve de vocabulaire expressif et une épreuve de connaissance des concepts. Les mesures de lecture étaient réparties en deux catégories : compréhension en lecture et acuité de la lecture. Les résultats ont montré que le lien entre les habiletés liées au code et les habiletés langagières était fort en début d'étude mais semblait s'affaiblir avec le temps ; au début de l'école élémentaire, les performances en lecture sont directement liées aux habiletés liées au code de maternelle ; sur les deux dernières années de l'étude (équivalent CE2 et CM1), trois prédicteurs se détachaient : les performances préalables en lecture, l'acuité de la lecture, les habiletés langagières. Ces résultats sont donc cohérents avec l'idée selon laquelle il y a, au fur et à mesure de l'apprentissage, un changement dans le poids des habiletés impliquées dans les compétences en lecture ; ils mettent également en évidence l'importance des habiletés langagières dans l'acquisition de ces compétences.

Bianco et collaborateurs (2012) ont étudié, sur une durée de 3 ans, l'impact des habiletés précoces liées au code ainsi que la compréhension orale. L'objectif était de tester le poids précoce de ces deux types d'habiletés sur les performances ultérieures en lecture d'une part pour confirmer l'intérêt des habiletés de code, d'autre part pour tenter de préciser quelle habileté langagière de plus haut niveau participe, de manière précoce dans l'apprentissage, au développement des compétences en lecture. La compréhension orale et la conscience phonologique ont ainsi été testées avec des enfants de 4 ans, ainsi que leurs acquisitions en lecture deux ans plus tard. En maternelle, les habiletés mesurées étaient donc la conscience phonologique, la compréhension orale (de phrases et de textes), et le vocabulaire réceptif. En CP, les élèves ont été testés sur de la lecture de pseudomots, de l'écriture (pseudomots, mots, phrases courtes), de la compréhension écrite. Les résultats montrent que la conscience phonologique précoce et la compréhension orale prédisent toutes deux l'identification de mots et la compréhension écrite ultérieure. Plus spécifiquement, la conscience phonologique prédisait particulièrement les compétences de lecture alphabétique, et la compréhension orale, la compréhension écrite. Ainsi, si les habiletés précoces liées au code sont déterminantes dans l'apprentissage de la lecture, les habiletés de langage oral précoce jouent également un rôle dans le développement des compétences ultérieures en lecture et notamment en compréhension.

Dans cette première partie de chapitre, nous avons donc présenté un certain nombre d'habiletés connues pour leur implication dans l'acquisition de la lecture. En effet, les habiletés liées au code, les habiletés cognitives générales ou encore les habiletés de langage oral participent à l'accroissement des performances en lecture. Maintenant que les prédicteurs de cet apprentissage majeur ont été esquissés, nous allons nous intéresser plus spécifiquement à l'évolution du lien entre les habiletés de lecture et d'écriture de manière longitudinale.

4.2 Liens entre lecture et écriture pendant l'apprentissage formel

Cette seconde partie de chapitre s'intéresse aux études longitudinales qui se sont centrées sur les liens entre lecture et écriture à l'école élémentaire. En effet, cette méthode d'investigation permet de mettre en évidence l'évolution de ces liens sur plusieurs années, et malgré les nombreuses recherches sur la lecture et sur l'écriture et leurs développements respectifs, il n'y a pas encore de consensus établi concernant les relations entretenues par ces deux habiletés, notamment sur le plan développemental en lien avec l'école. Cette partie va donc présenter une synthèse de plusieurs études longitudinales, centrées sur la relation entre la lecture et l'orthographe, pendant l'école élémentaire.

Pour Fitzgerald et Shanahan (2000), la plupart des études dans la littérature sur ce sujet se sont attachées à mesurer des corrélations entre lecture et écriture. Cependant, ces corrélations étaient généralement plutôt modérées (entre .20 et .50 selon les études), laissant donc une part importante de variance non expliquée par les habiletés étudiées. Ils relèvent également que la plupart des études sont de relativement faible ampleur (moins de 50 sujets selon leur revue), et sont transversales plutôt que longitudinales, ce qui *in fine* n'apporte que peu d'informations sur la complexité du lien lecture-écriture. Les études présentées dans les paragraphes ci-après sont quant à elles des études longitudinales, nous permettant ainsi de mettre en évidence une brève synthèse des connaissances apportées par ce type d'études. Ces études ont été sélectionnées selon plusieurs critères : les variables étudiées, l'empan temporel (le nombre d'années scolaires évaluées) ainsi que les méthodes d'analyses. L'objectif étant de pouvoir comparer nos résultats à des études déjà existantes, nous avons choisi de présenter des études qui s'étendaient sur au moins trois années scolaires, qui mesuraient des variables de lecture et d'écriture similaires à

celles que nous avons choisies, et qui proposaient des modèles statistiques permettant de voir l'évolution conjointe de ces habiletés dans le temps. La majorité des études sur le sujet a été menée en langue anglaise ; nous trouvons toutefois quelques études en français, finnois, néerlandais, allemand ou grec. Également, il semble que ce soit la période temporelle du début d'apprentissage formel (particulièrement les deux premières années d'enseignement formel, qui correspondent au CP et CE1 pour la France) qui soit la plus étudiée ; nous notons en effet un nombre plus réduit d'études portant sur la fin de la scolarité élémentaire (ce que certains considèrent comme le début de l'expertise en lecture-écriture). Concernant les habiletés étudiées, nous retrouvons toutes celles ayant montré leur impact sur les habiletés liées à la lecture et à l'écriture : conscience phonologique, RAN, encodage et décodage, écriture et reconnaissance de mots et pseudomots, compréhension en lecture.

Shanahan et Lomax (1986 ; 1988) ont cherché à étudier le lien entre lecture et écriture en testant des relations unidirectionnelles et bidirectionnelles. En utilisant pour mesures les variables connues comme prédictrices de ces deux habiletés, ils ont mis en évidence une influence mutuelle de la lecture et de l'écriture, qui, comparativement aux influences unilatérales d'une habileté envers l'autre, présentait le modèle qui s'ajustait le mieux aux données. Les auteurs montrent également qu'au début de l'étude, le lien entre la reconnaissance de mots et l'orthographe est plus important comparativement aux autres variables, mais qu'avec l'avancée dans la scolarité ou dans l'expertise, ce sont les aspects plus généraux liés à la connaissance du texte qui deviennent plus prégnants. Ainsi, si les corrélations permettent de distinguer lecture et écriture comme étant deux activités distinguables l'une de l'autre, leurs études ont également montré qu'un aspect développemental était à prendre en considération : en fonction du développement, le poids des différentes habiletés varie, qu'elles soient de bas niveau ou de plus haut niveau.

Abbott et collaborateurs (2010) proposent la théorie des niveaux de langue pour investiguer les relations entre lecture et écriture de manière longitudinale. En effet, ils postulent que l'échelle à laquelle on se situe pour étudier ce lien va influencer les résultats ; ainsi, les conclusions pourraient différer selon que l'on se place au niveau infra-mot (syllabe, phonème...), mot (décodage, épellation), phrase ou encore texte (compréhension de texte, production écrite). De la même manière, si l'on croise différents niveaux de langue lors d'une expérimentation, cela aurait un impact sur les résultats obtenus. Ils avancent l'idée selon laquelle la complexité des mécanismes cognitifs impliqués dans l'activité de lecture ou d'écriture va différer en fonction du niveau de langue étudié. Ils ont proposé une étude

longitudinale de grande ampleur (Abbott et al., 2010) menée sur deux cohortes (la première allant du Grade 1 au Grade 5, la seconde du Grade 3 au Grade 7). La première était composée de 128 élèves, et la seconde de 113 élèves. Sur cette étude menée en langue anglaise, les mesures utilisées étaient extraites de tests standardisés (*Wechsler Individual Achievement Test*, 2^{ème} édition). Nous ne présenterons ici qu'une partie des résultats proposés par les chercheurs. Les mesures étudiées portaient sur l'écriture (production de mots de difficulté croissante), la lecture de mots (de difficulté croissante également), la compréhension écrite (pour laquelle suite à la lecture d'un texte des questions littérales et inférentielles étaient posées), la compréhension orale, ainsi qu'une épreuve d'écriture des lettres de l'alphabet. Les résultats concernant les liens entre lecture et écriture ont montré tout d'abord l'importance des liens autorégressifs pour la lecture (poids entre .56 et .72) comme pour l'écriture (poids entre .61 et .83), comparativement aux liens entre les habiletés de lecture et d'écriture, mettant en avant leur distinctivité. En ce qui concerne les relations longitudinales entre ces deux habiletés, les résultats ont montré un lien bidirectionnel entre la lecture de mots et l'écriture du Grade 2 au Grade 7, avec un poids plutôt faible (compris entre .14 et .28). En revanche, les auteurs ont mis en évidence un lien unidirectionnel en Grade 1 de l'écriture à la lecture. Selon eux, le niveau de langue utilisé comme mesure pour tester l'influence des habiletés liées à la lecture et à l'écriture peut donc avoir un impact sur les résultats obtenus.

Les relations lecture-écriture dépendraient donc du niveau de langue auquel on se réfère dans la construction de l'expérimentation : pour Ahmed et collaborateurs (2014), il y a un lien bilatéral entre la lecture et l'écriture lorsque l'on utilise des variables au niveau de la phrase ; toutefois, si l'on se place au niveau du mot ou du texte, le lien est unilatéral de la lecture vers l'écriture. Leur étude, en langue anglaise, se basait sur un échantillon d'élèves (N=316) allant du Grade 1 au Grade 4, soit l'équivalent du CP au CM1. Cela est cohérent avec la théorie des niveaux de langage présentée précédemment (Abbott et al, 2010), et a depuis été répliqué par une étude d'Abbott et collaborateurs (2016), qui comparaient des performances d'élèves français et étatsuniens de Grade 2 et Grade 5 (correspondant respectivement au CE1 et CM2 français) : les mesures situées au niveau du mot étaient mieux corrélées les unes aux autres que les mesures situées au niveau du mot en lecture ou écriture avec celles situées au niveau du texte, par exemple.

Les relations lecture-écriture dépendraient également de la période temporelle à laquelle les études ont lieu. Pour Ahmed et collaborateurs (2014), du Grade 1 au Grade 4 en anglais le

lien est réciproque entre lecture et écriture seulement si l'on se place au niveau de la phrase ; pour le niveau du mot et du texte, le lien est unilatéral de la lecture vers l'écriture. Les auteurs ont testés pendant quatre ans 316 élèves du Grade 1 au Grade 4, sur des habiletés de décodage (lecture de mots isolés, lecture de pseudomots), lecture de phrases, fluence de lecture orale, compréhension en lecture. Concernant l'écriture, les élèves ont été testés sur les subtests liés à l'orthographe et à l'expression écrite de la WIAT (3^{ème} édition) ainsi que sur de l'écriture de phrases. Les auteurs ont ainsi trouvé, en plus de l'impact des niveaux de langue évoqué précédemment, une très grande corrélation entre les habiletés de lecture et d'écriture en Grade 1, corrélation qui, par conséquent, n'était plus aussi importante aux temps suivants.

Dans la même perspective, Chung et collaborateurs (2022) ont trouvé des différences importantes de tailles d'effets en fonction de la période temporelle. Les auteurs ont testé 73 élèves anglophones, en immersion dans une école francophone pendant trois ans, du Grade 1 au Grade 3. Les élèves ont été testés en lecture sur de la lecture de mots (lecture de lettres et de mots en anglais, lecture à voix haute de mots isolés en français), de la conscience phonologique en anglais, des matrices, et de l'écriture (écriture de mots d'après des tests standardisés en français et en anglais). Ils étaient évalués sur de la reconnaissance orthographique en variable dépendante. Les modèles *cross-lagged* ont montré que lecture et écriture prédisaient toutes deux l'orthographe ultérieure dans les deux langues, même après prise en compte des autorégressions. Par contre, il n'y avait pas d'effet de la reconnaissance orthographique sur la lecture ou l'écriture de mots ultérieurs. Cette étude a également montré des effets plus grands du Grade 1 au Grade 2 que du Grade 2 au Grade 3 de manière générale ; les auteurs en concluent que ceci est consistant avec l'auto-apprentissage proposé par Share (1995), qui postule l'importance du décodage pour la construction des représentations lexicales : les activités de lecture et d'écriture permettraient de consolider les représentations orthographiques en mémoire, en mobilisant l'attention du lecteur-scripteur sur les lettres qui constituent le mot lu ou écrit.

Toujours en anglais, Deacon et collaborateurs (2012) ont suivi une cohorte de 123 élèves pendant trois ans (du Grade 1 au Grade 3) et ont également trouvé un lien unilatéral de la lecture vers l'écriture. En testant la lecture de mots isolés, le vocabulaire en réception, la conscience phonologique et les matrices ainsi que l'orthographe lexicale (reconnaissance orthographique) et sublexicale (tâche de choix forcé), les auteurs se sont également intéressés à ce lien en début d'école élémentaire. Leurs résultats de l'analyse *cross-lagged* ont tout d'abord permis de mettre en évidence la distinction entre orthographe lexicale et sublexicale, mais ils ont également

montré que la lecture de mots préalable prédisait bien l'orthographe ultérieure (qu'elle soit lexicale ou sublexicale), mais que l'inverse n'était pas valable.

En néerlandais (Schaars et al., 2017) le schéma semble similaire à celui de l'anglais : 487 enfants ont été testés en maternelle (habiletés de littéracie précoce), Grade 1 et Grade 2 (RAN, mémoire, lecture de mots, dictée de mots isolés, de phrases et tâche de décision lexicale). Les résultats des modèles cross-lagged ont montré un lien unidirectionnel de la lecture vers l'écriture. Là encore, les liens autorégressifs des habiletés sont forts, et plus particulièrement à l'intérieur d'un même domaine d'habileté. Ces résultats sont donc en cohérence avec les études énoncées précédemment, mettant en évidence un lien fort de la lecture vers l'écriture dans les orthographe non transparentes en début de scolarité formelle.

Il semble y avoir un nombre plus faible d'études sur la fin de l'école élémentaire, comparativement au début de l'apprentissage formel. Shanahan (1984) a testé des élèves en Grade 2 et Grade 5, pour une comparaison novices/experts au niveau du mot. L'auteur a montré que la relation entre lecture et écriture change avec le développement de l'expertise de l'enfant : les performances des novices présentaient un poids plus important de la production de mots et de la reconnaissance de mots ; du côté des lecteurs experts, c'est le vocabulaire qui semble avoir plus d'importance comparativement aux performances liées à la production écrite.

La langue étudiée peut également avoir un impact majeur sur les résultats, notamment sur les habiletés dont le poids est très important en début d'acquisition. Georgiou et collaborateurs (2012) ont voulu comparer cela sur trois langues d'opacité différente : le finnois, le grec et l'anglais (respectivement de la langue la moins opaque à la langue la plus opaque). Les élèves (respectivement 82 pour l'échantillon anglais, 88 pour l'échantillon finlandais et 70 pour l'échantillon grec) étaient prétestés en maternelle puis post-testés en Grade 2, pour vérifier l'effet de la connaissance des lettres, du RAN et de la conscience phonologique, dont on sait qu'ils sont importants en début d'apprentissage. En fin de Grade 2, ils étaient testés sur du décodage de pseudomots, de la fluence de lecture de texte, et de l'écriture sous la dictée (pseudomots pour le finnois, dictée à trous pour l'anglais et dictée de mots pour le grec). Les résultats ont montré que la connaissance des lettres avait un impact très significatif sur les performances ultérieures en lecture comme en écriture, et ce, dans les trois langues. Toutefois, le RAN ne semblait avoir d'effet que sur l'anglais pour les performances ultérieures en lecture, et sur l'anglais et le finnois pour les performances ultérieures en écriture. Les auteurs notent

toutefois l'absence de mesure en Grade 1 qui aurait permis de voir un palier intermédiaire dans la progression des habiletés, mais nous notons surtout la difficulté de comparer les performances aux épreuves de dictée, puisqu'elles étaient différentes dans chacune des langues étudiées.

Notre travail s'inscrit donc dans cette perspective longitudinale, qui n'a que peu été étudiée en langue française à notre connaissance. L'originalité de notre travail va donc résider principalement dans son amplitude, que cela soit dans le temps ou dans la taille de l'échantillon. Les recueils de donnée se sont en effet déroulés sur cinq années scolaires, soit l'ensemble de l'école élémentaire (du CP au CM2), et plus de 500 élèves ont participé sur l'ensemble de l'étude. De plus, de nombreuses habiletés connues pour être impliquées dans l'apprentissage de la lecture et de l'écriture ont été utilisées tout au long de cette étude, afin d'essayer de mettre en évidence sur l'ensemble de l'école élémentaire l'évolution du lien entre lecture et écriture, ainsi que les prédicteurs impliqués dans la construction de la compétence orthographique.

Résumé du Chapitre 4

Les prédicteurs principaux de la lecture et de l'orthographe à l'école élémentaire ont été présentés dans une première partie : il s'agit des habiletés liées au code (décodage, fluence de lecture, connaissance des lettres), des habiletés métalinguistiques (conscience phonologique), des habiletés cognitives générales (RAN, mémoire de travail), et du langage oral (vocabulaire, compréhension orale).

Dans une seconde partie ont été présentées certaines études évaluant les liens entre lecture et orthographe pendant l'école élémentaire. Les études de notre synthèse ont majoritairement mis en évidence un lien unilatéral de la lecture vers l'orthographe ; les quelques études ayant montré un lien bilatéral montrent également un poids nettement plus fort de la lecture vers l'orthographe. Enfin, les poids entre les différentes habiletés semblent plus forts en début d'école élémentaire, pour s'atténuer par la suite.

Synthèse théorique

A travers la mise en contexte théorique présentée pendant ces quatre premiers chapitres, nous avons donc pu voir tout d'abord les différents systèmes d'écriture. Cette différence dans la façon de construire la communication écrite implique un fonctionnement qui n'est parfois pas comparable, et en tout cas qui n'est pas identique ; au sein d'un même système d'écriture, des différences majeures peuvent d'ailleurs exister. Par exemple, dans les systèmes alphabétiques, l'apprentissage de la lecture pourra prendre plus ou moins de temps selon le degré d'opacité de la langue. En effet, pour les langues transparentes comme le finnois, la régularité dans les correspondances entre phonèmes et graphèmes permettra un apprentissage plus rapide de la lecture. *A contrario*, pour les langues opaques comme l'anglais ou le français, l'apprentissage sera plus long étant donné les nombreuses irrégularités dans les correspondances entre phonèmes et graphèmes. Plus spécifiquement, la langue française, outre sa grande disparité entre les possibles graphèmes et phonèmes, comporte également une morphologie très particulière, notamment avec la présence nombreuse de lettres muettes en fin de mots, ou encore les accords en genre, nombre et grammaticaux très nombreux, ce qui complexifie davantage la lecture et l'écriture du français.

Dans un second temps, nous avons présenté différents modèles théoriques de la lecture et de l'écriture : qu'ils soient basés sur un niveau expert, ou développemental, nous avons suivi l'évolution des modèles jusqu'aux modèles interactionnistes. Nous nous sommes ensuite concentrés sur l'orthographe et ses particularités : après avoir défini le terme d'orthographe, nous avons présenté plusieurs modèles théoriques concernant son acquisition et son expertise. Dans ce travail, nous accordons de l'importance à la distinction entre orthographe lexicale et sublexicale : nous avons donc présenté plus en détails les spécificités de l'orthographe sublexicale en français. Nous avons également discuté du caractère universel que certains auteurs considèrent comme essentiel pour justifier de la pertinence théorique d'un modèle.

Ensuite, nous avons proposé quelques facteurs environnementaux qui peuvent impacter les performances des élèves : ces facteurs peuvent dépendre de la classe à laquelle appartiennent les élèves, de l'enseignant, de la pratique pédagogique mise en place pour la transmission du savoir, ou encore du contenu en lui-même.

Cette partie théorique s'est achevée sur les liens entre lecture et écriture : après avoir présenté les prédicteurs communs aux deux habiletés et le lien qui les sous-tend notamment en début d'apprentissage, nous avons présenté une synthèse de travaux portant sur l'étude longitudinale de l'évolution des liens entre lecture et écriture tout au long de l'école élémentaire. Les travaux que nous avons effectués portent en partie sur l'étude longitudinale de ce lien.

Notre travail a en effet pour objectif d'essayer de mieux comprendre le développement des capacités orthographiques, ou d'écriture, pendant la période charnière qu'est l'école élémentaire. Nous avons choisi pour cela deux approches : une approche de type « macro », dans une perspective développementale, avec tout d'abord une présentation descriptive de l'évolution des habiletés de lecture et d'écriture sur un temps long (ici, du CP au CM2). Dans un second temps, nous étudierons, à l'aide d'un modèle *cross-lagged*, l'évolution du lien entre ces habiletés de lecture et d'écriture du CE1 au CM2. Ensuite, nous proposerons une approche de type « micro », en interrogeant la distinction entre orthographe lexicale et sublexicale, sur les trois premières années de l'école élémentaire.

Concernant notre approche plus globale, si plusieurs études longitudinales ont mis en avant un lien unidirectionnel de la lecture vers l'écriture en début d'apprentissage (e.g. Ahmed et al., 2014 ; Deacon et al., 2012) en langue anglaise et en néerlandais (Schaars et al., 2017), nous avons pour objectif d'étudier ce lien en français sur quatre années scolaires. Nous souhaitons également vérifier s'il est possible de distinguer la lecture isolée de la lecture en contexte dans un modèle à trois voies, composé de la lecture isolée, de la fluence de lecture en contexte et de l'orthographe d'usage afin d'étudier leur évolution sur un temps long en français.

Ensuite, dans une perspective plus ciblée, nous souhaitons focaliser nos travaux sur le lien plus spécifique entre l'orthographe lexicale et sublexicale. Dans la littérature, il ne semble pas y avoir de consensus concernant cette distinction possible entre orthographe lexicale et sublexicale, certains les considérant comme deux façons de mesurer une même habileté. D'autres, en revanche, considèrent l'orthographe lexicale comme un indicateur de la capacité à mesurer des unités orthographiques spécifiques, et l'orthographe sublexicale comme un indicateur des capacités d'apprentissage statistique des régularités graphotactiques de la langue. Il s'agit alors pour nous d'étudier les relations entre ces deux habiletés, à travers la mise en

évidence de prédictors communs mais également de prédictors propres à chacune du CP au CE2.

La partie empirique de ce travail de thèse s'attachera donc à présenter trois études portant sur l'acquisition et le développement des compétences orthographiques. La première étude sera proposée selon une approche globale, centrée sur l'évolution sur un temps long des habiletés d'orthographe ainsi que de lecture, et sur les liens qui peuvent les unir ; les deux autres études seront ciblées plus spécifiquement sur l'évolution de la composante orthographique, en début de scolarité élémentaire.

- L'étude 1 porte sur l'évolution du lien lecture-écriture du CP au CM2. Après avoir décrit les évolutions des différentes habiletés étudiées et présenté leurs corrélations, nous tâcherons de répondre aux hypothèses suivantes :
 - Hypothèse 1 : nous observons un lien unilatéral de la lecture vers l'écriture, particulièrement en début de scolarité.
 - Hypothèse 2 : nous observons une distinction entre lecture isolée et lecture en contexte dans les liens avec l'écriture tout au long de l'école élémentaire.
- Les études 2 et 3 portent sur la distinction entre orthographe lexicale et sublexicale du CP au CE2. Nous tenterons, via ces études, de répondre à l'hypothèse suivante :
 - Hypothèse 3 : orthographe lexicale et sublexicale sont deux habiletés distinctes et dissociables de la composante orthographique, car elles ont des prédictors en commun, mais également des prédictors propres à chacune.

SECONDE PARTIE : EMPIRIE

Chapitre 5 : Méthode générale

Ce travail de thèse s'inscrit dans le cadre du projet ACQUISLONGIT. Ce projet de recherche est issu de l'Institut Carnot de l'Éducation Auvergne Rhône-Alpes (ICÉ), lui-même intégré dans l'Institut Français de l'Éducation de l'ENS Lyon ; la DEPP (Direction de l'Évaluation, de la Prospective et de la Performance, Ministère de l'Éducation Nationale) ainsi que l'équipe de la circonscription du Haut-Grésivaudan (Isère) se sont également impliqués dans l'élaboration de ce travail. L'objectif du projet ACQUISLONGIT était de pouvoir étudier sur un temps long (cinq années scolaires, soit du CP au CM2) les habiletés liées à l'acquisition de la lecture et de l'écriture ainsi que leur interaction. Outre ces données quantitatives, des données qualitatives ont aussi été récoltées dans le cadre de ce projet, notamment concernant la motivation et le sentiment de compétence en lecture et en écriture. Les recueils de données ont commencé lors de l'année scolaire 2014-2015 et se sont achevés en 2018-2019.

Nous allons, dans un premier temps, présenter la méthode générale pour l'ensemble des études. Nous commencerons par détailler les participants et leurs caractéristiques, puis nous présenterons le matériel utilisé, et pour finir la procédure.

5.1 Méthode

5.1.1 Participants

561 élèves de CP ont participé aux premiers recueils de données, dont 52% de garçons ; l'âge moyen au début de l'étude était de 6.01 ans (0.35 SD). Préalablement au début de l'expérimentation, les parents ou tuteurs légaux avaient été informés de l'étude et de ses objectifs généraux ; ils avaient également rempli une fiche d'informations ainsi qu'une autorisation parentale ; ils avaient de plus été informés qu'ils pouvaient désinscrire leur enfant de l'étude à n'importe quel moment durant les cinq années de passation.

Les élèves appartenaient à 23 écoles de la circonscription du Haut-Grésivaudan, en Isère. Les classes pouvaient être de composition variée, notamment en fonction de la taille de l'école : cours simple ou cours multiple (double, triple voire quadruple niveau). Le récapitulatif de la

composition des classes (nombre moyen d'élèves dans le niveau concerné par l'étude, ainsi que l'écart-type moyen) est présenté dans le Tableau 2.

Tableau 2. Nombre moyen d'élèves par classe et écart-type pour chaque niveau de classe étudié

	CP	CE1	CE2	CM1	CM2
Moyenne	17,6	15,61	14,88	14,71	17,52
Ecart-type	7,88	8,11	8,04	5,95	8,33

L'échantillon n'est toutefois pas représentatif de l'ensemble de la population française ; le Grésivaudan présente en effet un taux d'activité légèrement supérieur à la moyenne nationale (respectivement 78.2 contre 74.1%, INSEE, RP, 2019), ainsi qu'un taux de chômage presque deux fois moins élevé que la moyenne nationale (respectivement 7.5% contre 13.4%, INSEE, RP, 2019). De plus, le territoire du Grésivaudan compte 21% de cadres supérieurs, contre 14% pour le territoire de référence (INSEE, 2011) même si des disparités importantes existent à l'intérieur du territoire.

Les variations du nombre d'élèves ayant participé à l'étude pour chaque temps de mesure (Tableau 3) peuvent avoir plusieurs explications, notamment, en ce qui concerne directement les élèves, comme un déménagement ou une absence imprévue au moment des passations. En effet, les contraintes temporelles et organisationnelles ne nous permettaient pas, la plupart du temps, de revenir dans les écoles après les jours prévus de passation.

Tableau 3. Nombre d'élèves ayant participé à l'étude à chaque temps de mesure, du CP au CM2

	CP	CE1	CE2	CM1	CM2
Nombre d'élèves	561	599	620	577	591

Sur l'ensemble de cet échantillon, 26 élèves évoluent dans un environnement bilingue à la maison. En plus du français, nous retrouvons sept élèves pour qui l'anglais est également parlé à la maison, six pour l'arabe, trois pour le portugais, deux pour l'italien, l'espagnol et le russe, et un pour le polonais, le turc, le tchèque et la langue des signes française.

5.1.2 Matériel

Tableau 4. Synthèse des épreuves utilisées en fonction des années scolaires pour l'ensemble du suivi longitudinal

Compétence	Epreuve	T1 - Début CP	T2 - Fin CP	T3 - Fin CE1	T4 - Fin CE2	T5 - Fin CM1	T6 - Fin CM2
Compétences générales	Matrices WPPSI 3	X					
	RAN Chiffres - Lettres - Couleurs	X					
	RAN Chiffres					X	X
	Mémoire des chiffres WISC 4				X	X	X
Langage oral	EVIP	X (forme A)	X	X	X	X	X
	Vocabulaire WISC 4	X (WPPSI 3)		X	X	X	X
	Comp. orale - textes	X	X	X	X	X	X
	Comp. orale - phrases ECOSSE 2	X (ECOSSE 1)		X	X	X	X
	Synonymes					X	
Code oral	Comptage syllabe	X					
	Suppression syllabe initiale	X					
	Suppression syllabe finale	X					
	Détection phonème	X					
	Suppression phonème	X	X			X	X
	Addition phonème		X				
	Contrepèteries					X	
Code écrit - Lecture	Lecture mots	X	X	X	X	X	X
	Lecture pseudomots	X	X	X	X	X	X
	Lecture lettres	X					
Code écrit - Orthographe	Dictée lettres	X					
	Dictée pseudomots	X	X	X	X	X	X
	Dictée mots	X	X	X			
	Dictée phrases		X	X			
	Dictée Le Corbeau			X	X	X	X
	Reconnaissance orthographique				X		
Code écrit	Fluence de lecture (1min)		X	X	X	X	X
	Sensibilité graphotactique			X	X	X	X
	Vitesse d'écriture			X			
	Morphologie flexionnelle					X	X
	Conscience morphologique						X
	Conscience syntaxique						X
Compréhension écrite	Histoires		X	X	X	X	X
	Maze test			X	X	X	X
	Inférences			X	X	X	X
	Informations		X				
	Phrases						X
Echelles	Motivation	X	X	X	X	X	X
	Sentiment de compétence	X	X	X	X	X	X
Questionnaires enseignants	Avis sur les élèves	X		X	X	X	X
	Pratiques enseignantes	X		X	X	X	X

Le Tableau 4 répertorie l'ensemble des mesures ou épreuves utilisées pendant les cinq années de suivi, et proposées aux élèves ou aux enseignants. Dans le cadre de ce travail de thèse, nous ne présenterons toutefois que les épreuves utilisées pour nos études expérimentales ; ces épreuves sont répertoriées dans le Tableau 5.

Tableau 5. Présentation des épreuves utilisées pour les études 1, 2 et 3 et temps de mesure auxquelles elles ont été testées

	T1 (début CP)	T2 (fin CP)	T3 (fin CE1)	T4 (fin CE2)	T5 (fin CM1)	T6 (fin CM2)
Matrices	x					
RAN AN	x					
Lecture mots		x	x	x	x	x
Lecture pseudomots		x	x			
Fluence textes		x	x	x	x	x
Dictée mots		x	x			
Dictée pseudomots		x	x			
Dictée phrases		x				
Dictée Corbeau (usage)			x	x	x	x
Reconnaissance ortho				x		
Sensibilité graphotactique			x	x		
Vocabulaire		x	x			
Syntaxe		x	x			
Compréhension orale		x	x			
Compréhension écrite			x			

Nous allons maintenant présenter en détails chacune des épreuves utilisées dans nos trois études ; les épreuves sont présentées de manière longitudinale. Pour chacune d'entre elles, nous précisons la modalité de passation (individuelle ou collective) ainsi que la description de l'épreuve.

5.1.3 Compétences générales

Début CP

Logique non verbale (MAT, passation individuelle). L'épreuve est celle du subtest Matrices de la batterie WPPSI 3 (Wechsler, 2004). Elle est constituée de 29 items de difficulté croissante. Pour chaque item, l'enfant observe une matrice incomplète et choisit l'élément manquant pour la compléter, parmi quatre ou cinq propositions. Après quatre erreurs consécutives, l'épreuve s'arrête. Le score relevé est le nombre de réponses correctes (max = 29).

Dénomination rapide (RAN, passation individuelle). Chaque épreuve comprend deux séries de 40 items, chaque série est constituée de 5 lignes de 8 items. L'enfant doit dénommer

l'ensemble des items le plus rapidement possible. Pour chaque épreuve, on relève le temps mis pour nommer correctement chaque série.

Dénomination rapide de chiffres (RANch). Les items sont les chiffres de 1 à 5 (une série est présentée en Annexe A).

Dénomination rapide de lettres (RANI). Les items sont 5 lettres voyelles présentées en majuscules, police Colibri en taille 14 (une série est présentée en Annexe B).

5.1.4 Langage oral

Fin CP

Vocabulaire - étendue (EVIP, passation collective). L'expérimentateur dit un mot à haute voix, et les enfants doivent choisir parmi quatre images celle qui correspond au mot entendu. Respectivement 45 et 32 mots sur l'ensemble des mots constituant l'EVIP (Dunn et al, 2013) étaient proposés début et fin CP, puis 45 autres du CE1 au CM1 (l'ensemble des listes est présenté en Annexe C).

Vocabulaire - profondeur (WISC, passation individuelle) : subtest Vocabulaire du WISC 4 (Wechsler, 2005). L'enfant doit définir des mots de difficulté croissante. Après cinq réponses erronées consécutives, l'épreuve s'arrête. Début CP, la version utilisée est celle de la WPPSI-III (Wechsler, 2004).

Compréhension de textes (COMPo, passation individuelle) : l'enfant écoute une histoire courte lue par l'expérimentateur et doit ensuite répondre à des questions de compréhension. Deux histoires différentes sont présentées à l'enfant à chaque temps de mesure. Plusieurs niveaux de compréhension sont évalués (e.g., littéral ou inférentiel). L'Annexe D présente les différentes histoires ainsi que les questions correspondantes pour chaque temps de mesure.

Compréhension de phrases (ECOSSE, passation collective) : l'expérimentateur lit une phrase à haute voix, et l'enfant doit choisir parmi quatre images celle qui correspond à la phrase énoncée. L'Annexe E présente les listes de phrases extraites de l'ECOSSE (Lecocq, 2011) et proposées pour chaque temps de mesure.

CE1

Vocabulaire – étendue (EVIP, épreuve identique à celle du CP).

Vocabulaire – profondeur (WISC, épreuve identique à celle du CP).

Compréhension de textes (COMPo, épreuve identique à celle du CP).

Compréhension de phrases (ECOSSE, épreuve identique à celle du CP).

5.1.5 Code écrit

Début CP

Lecture de mots (LEC_MOT, passation individuelle). Durée de l'épreuve : une minute (chronométrée). Sur une feuille recto-verso, 150 mots sont imprimés (police Colibri, taille 12). L'enfant doit en lire le plus possible en une minute, sans se tromper. Le score pour cette épreuve est le nombre total de mots correctement lus en une minute. Les mots sont de longueur et de difficulté croissante (voir Annexe F). Pour le premier temps de mesure (début CP), la liste était réduite aux cinquante premiers mots. Cette épreuve est tirée de la batterie EVALEC (Sprenger-Charolles et al., 2010).

Lecture de pseudomots (LEC_PSEUDO, passation individuelle). Durée de l'épreuve : une minute (chronométrée). Sur une feuille recto-verso, 150 pseudomots sont imprimés (police Colibri, taille 12). L'enfant doit en lire le plus possible en une minute, sans se tromper. Le score pour cette épreuve est le nombre total de pseudomots correctement lus en une minute. Les pseudomots sont de longueur et de difficulté croissante (voir Annexe G). Tout comme l'épreuve précédente, la lecture de pseudomots pour les recueils de type pré-tests (début CP) a été réduite à cinquante pseudomots. Cette épreuve est tirée de la batterie EVALEC (Sprenger-Charolles et al., 2010).

Fin CP

Lecture de mots (LEC_MOT). Epreuve identique à celle du début CP.

Lecture de pseudomots (LEC_PSEUDO). Epreuve identique à celle du début CP.

Fluence de lecture (FLUENCE, passation individuelle). L'élève doit lire un texte à voix haute, le plus rapidement possible pendant une minute (épreuve chronométrée). Le score pour cette épreuve correspond au nombre de mots correctement lus dans le temps imparti. La longueur des textes était variable, et liée à la difficulté du texte en question ; elle est comprise entre 221 et 467 mots pour les textes utilisés sur l'ensemble des temps de mesure. Deux textes sont présentés ainsi à l'élève (voir en Annexe H pour l'ensemble des textes).

CE1

Lecture de mots (LEC_MOT). Epreuve identique à celle du CP.

Lecture de pseudomots (LEC_PSEUDO). Epreuve identique à celle du CP.

Fluence de lecture (FLUENCE). Epreuve identique à celle du CP.

CE2

Lecture de mots (LEC_MOT). Epreuve identique à celle du CP.

Lecture de pseudomots (LEC_PSEUDO). Epreuve identique à celle du CP.

Fluence de lecture (FLUENCE). Epreuve identique à celle du CP.

CM1

Lecture de mots (LEC_MOT). Epreuve identique à celle du CP.

Fluence de lecture (FLUENCE). Epreuve identique à celle du CP.

CM2

Lecture de mots (LEC_MOT). Epreuve identique à celle du début CP.

Fluence de lecture (FLUENCE). Epreuve identique à celle du début CP.

5.1.6 Orthographe

Début CP

Dictée de mots (DICT_MOTS, passation collective). Les enfants doivent d'abord écrire leur prénom, puis sept mots sous dictée (*maman, nez, robe, à, tu, le, sur*). Le score total correspond au nombre de mots correctement écrits (max=7). Les mots dictés ont été choisis selon leur fréquence dans les manuels scolaires ; des mots-outils ont aussi été présentés (*le, sur*) ainsi que des mots dont l'orthographe est irrégulière (*nez, par exemple*) afin d'éviter un effet plafond potentiel en recourant à l'utilisation de mots dont l'orthographe lexicale est plus irrégulière.

Dictée de pseudomots (DICT_PSEUDO, passation collective). Les enfants doivent écrire sous dictée sept pseudomots (*PA, TO, LI, NAL, PIVA, DARO, SILU*). Chaque graphème est considéré comme valide si le phonème correspondant est plausible. Le score total correspond au nombre de graphèmes plausibles (max=172).

Fin CP

Dictée de mots (DICT_MOTS). Epreuve identique à celle du début CP, mais constituée de douze mots (*tu, nez, robe, à, porte, arbre, marche, pied, attention, sable, ruche, scie*). Le score total correspond au nombre de mots correctement écrits (max=12).

Dictée de pseudomots (DICT_PSEUDO). Epreuve identique à celle du début CP, mais constituée de douze pseudomots (*TO, NAL, PIVA, DARO, SILU, NUTON, BIVOU, GASSI, FLOPAN, CHULPIN, CRATAN, BRUTO*). Chaque graphème est considéré comme valide si le phonème correspondant est plausible. Le score total correspond au nombre de graphèmes plausibles (max=56).

Dictée de phrases (DICT_PHRASES, passation collective). Les enfants doivent écrire deux phrases sous dictée (*Maman met un moule à gâteau dans le four ; Il compte sept livres sur la table*). Le score total correspond au nombre de mots correctement écrits sur l'ensemble des phrases (max=17).

Dictée de mots (DICT_MOTS). Epreuve identique à celle du début CP, mais constituée de 23 mots (*tu, nez, robe, à porte, marche, pied, attention, sable, ruche, scie, minute, ouvre, sourire, femme, tomate, lavabo, poudre, écharpe, punition, nœud, sept, album*). Le score total correspond au nombre de mots correctement écrits (max=23).

Dictée de pseudomots (DICT_PSEUDO). Epreuve identique à celle du début CP, mais constituée de seize pseudomots (*NAL, PIVA, DARO, SILU, NUTON, BIVOU, GASSI, FLOPAN, CHULPIN, CRATAN, BRUTO, NAGULE, GISOVE, LUBSIR, BADUSSE, GUIGNON*). Chaque graphème est considéré comme valide si le phonème correspondant est plausible. Le score total correspond au nombre de graphèmes plausibles (max=172).

Dictée de phrases (DICT_PHRASES). Epreuve identique à celle de fin CP.

Dictée *Le corbeau* (DICT_CORBEAU, passation collective). Les enfants écrivent sous dictée un court texte, dont chaque segment de phrase est répété deux fois. Cette épreuve standardisée (*L2MA*, Chevrie-Muller, 1997) propose trois niveaux de cotation correspondants aux réponses correctes des enfants pour les scores phonologie, orthographe d'usage, et grammaire, ainsi qu'un score total de l'épreuve. Dans le cadre de ce travail de thèse, nous présenterons uniquement les sous-scores de l'orthographe d'usage, correspondant à l'orthographe lexicale. En CE1, le score total de l'orthographe d'usage est calculé en additionnant les mots écrits selon une orthographe plausible (max=22).

Sensibilité graphotactique (GRAPHO, passation collective). Epreuve basée sur une tâche de Pacton et al. (2013). Les enfants doivent choisir, entre deux pseudomots, celui qui ressemble le plus à un mot du français. Les paires de pseudomots sont réparties en trois groupes, chacun correspondant à une règle implicite d'écriture de la langue française : une consonne en position initiale d'un mot ne peut être doublée (e.g. *ffurois – furrois*) ; une consonne ne peut être doublée si elle suit déjà une consonne simple (e.g. *acclomir – acllomir*) ; toutes les consonnes ne peuvent être doublées (e.g. *boxxit – bottit*). L'épreuve comporte au total 21 paires de pseudomots ; le score pour cette épreuve correspond au nombre total de pseudomots cibles choisi par les élèves (max=21). L'ensemble des paires de pseudomots utilisés pour cette épreuve est présenté en Annexe I.

CE2

Dictée *Le corbeau* (DICT_CORBEAU). Epreuve identique à celle du CE1.

Reconnaissance orthographique (RECO_ORTHO, passation collective). Les enfants doivent, sur leur cahier, cocher la proposition parmi trois qui correspond à l'orthographe correcte d'un mot énoncé par l'expérimentateur. Les items proposés consistent en trois différentes orthographes d'un même phonème : une cible et deux homophones. Les deux orthographes incorrectes sont chacune l'orthographe cible d'un autre item (i.e. le *ête* de *quête* se retrouve comme proposition pour *lunette*). Le score total pour cette épreuve correspondait à la somme de tous les items pour lesquels trois mots étaient présentés (max=42). L'ensemble des items proposés pour cette épreuve est présenté en Annexe J.

Sensibilité graphotactique (GRAPHO). Epreuve identique à celle du CE1.

CM1

Dictée *Le corbeau* (DICT_CORBEAU). Epreuve identique à celle du CE1.

CM2

Dictée *Le corbeau* (DICT_CORBEAU). Epreuve identique à celle du CE1. En CM2, la dictée était majorée de deux phrases, ce qui portait le score maximum pour l'orthographe d'usage à 30.

5.1.7 Compréhension écrite

Fin CP

Compréhension de textes (COMP_ECRITE, passation collective). Quatre courts textes (*la boîte, le chaton, l'oiseau, le pêcheur*) sont présentés aux élèves, qui doivent les lire puis répondre à des questions de compréhension. Trois types de questions étaient posées : les questions littérales (la réponse est explicitement donnée dans le texte), les questions *text-based* (la réponse implique une inférence fondée sur les informations données dans le texte), et les questions *knowledge-based* (la réponse implique une inférence de connaissance). Tous les textes présentés ainsi que les questions correspondantes sont présentés en Annexe K. Les textes et questions choisis sont issus d'une étude précédente (DEVCOMP), pour laquelle des analyses des TRI (théories de réponse à l'item) ont été effectuées pour l'ensemble des textes et des

questions. Cela a permis de vérifier la catégorisation des différentes questions, et de calibrer la difficulté des textes pour chaque temps de mesure.

CE1

Compréhension de textes (COMP_ECRITE). Epreuve identique à celle de fin CP ; deux textes sont présentés aux élèves (*le pêcheur*, *Bouddha*). Le texte *le pêcheur* reprend les trois types de questions évoquées précédemment (littérale, *text-based* et *knowledge-based*). Le texte *Bouddha* quant à lui comporte également trois types de questions : littérale, basée sur la cohérence locale du texte, et basée sur la cohérence globale du texte. Le texte ainsi que les questions sont présentés en Annexe K.

5.2 Procédure

Les passations se déroulaient en deux temps. Tout d'abord, les élèves remplissaient deux cahiers de passations collectives ; le cahier de la passation individuelle était quant à lui rempli par l'expérimentateur pour chaque élève. Les passations collectives se déroulaient dans la mesure du possible sur une matinée. Pour chaque passation collective, en classe entière, au moins deux expérimentateurs étaient présents : l'un présentait les consignes, l'autre circulait dans les rangées pour s'assurer de la bonne compréhension des consignes par les élèves.

Après s'être présentés et avoir présenté globalement l'étude, les expérimentateurs faisaient passer un premier cahier d'épreuves collectives avant la récréation, et un second après. Ces passations avaient lieu dans la salle de classe, en présence de l'enseignant, excepté pour les classes à cours multiples. Pour la passation individuelle, l'élève et l'expérimentateur s'installaient au calme, dans une salle prévue pour les expérimentateurs si les locaux le permettaient, ou bien dans le couloir. Pour chacune des épreuves, un ou plusieurs exemples, selon la complexité de la tâche, était systématiquement présenté aux élèves une fois la consigne donnée, afin de s'assurer de la bonne compréhension de l'exercice. La durée moyenne de chaque passation (Tableau 6), et particulièrement de la passation individuelle, devait durer moins de 45min dans la mesure du possible.

Tableau 6. Durée moyenne des épreuves collectives et individuelles à chaque temps de mesure.

	Début CP	Fin CP	Fin CE1	Fin CE2	Fin CM1	Fin CM2
Collectif 1	30 min	40 min	45 min	45 min	45 min	45 min
Collectif 2	30 min	35 min	45 min	45 min	45 min	45 min
Individuel 1	25 min	30 min	35 min	40 min	50 min	45 min
Individuel 2	30 min					

Chapitre 6 : Etude 1 : évolution longitudinale du lien lecture-écriture

Cette première étude concerne l'évolution du lien entre les habiletés liées à la lecture et à l'écriture. Nous avons choisi pour cela de suivre les relations de trois variables sur quatre années scolaires, du CE1 au CM2. Ce chapitre va donc s'attacher à présenter les hypothèses ayant guidé les analyses statistiques, le matériel et la méthode utilisés, ainsi que les résultats obtenus. Nous terminerons ce chapitre par une brève conclusion, puisqu'une conclusion générale sera proposée à la fin des différentes études pour synthétiser l'ensemble des résultats.

6.1 Hypothèses

Deux hypothèses ont été formulées concernant l'évolution du lien entre lecture et écriture sur un temps long. Tout d'abord, la première hypothèse correspond au lien entre lecture et orthographe : s'il n'y a pour le moment pas de consensus sur le sens de ce lien dans la littérature, nous postulons, comme certaines études et comme le propose Share (1995) avec l'hypothèse d'auto-apprentissage, un lien plutôt unilatéral de la lecture vers l'orthographe : ainsi, nous nous attendons à voir un effet des deux habiletés de lecture sur l'habileté orthographique utilisés dans le modèle. Cet effet devrait également être plus fort en début d'apprentissage.

Notre seconde hypothèse pour cette première étude concerne les deux variables utilisées pour mesurer la lecture : nous supposons que ces deux variables (lecture de mots isolés et fluence de lecture de texte en contexte) sont distinguables et montreront ainsi des effets propres à chacune. Cette hypothèse fait suite à des études ayant interrogé l'apport unique de la fluence de lecture en contexte et celui de la fluence de lecture de mots isolés (e.g. Goigoux et al., 2016).

6.2 Matériel et méthode

Afin d'étudier l'évolution longitudinale du lien lecture-écriture, une analyse *cross-lagged* a été effectuée. Nous avons pour cela utilisé le logiciel MPlus (version 8.4).

Ce type d'analyse permet, sur un temps long, d'étudier l'évolution de plusieurs variables de trois manières : d'une part, de manière autorégressive, c'est-à-dire en étudiant l'évolution d'une même variable au fil du temps ; d'autre part, de manière transversale, c'est-à-dire en étudiant les relations entre deux variables à un temps donné ; enfin, de manière croisée, c'est-à-dire en étudiant l'effet d'une variable à un temps donné sur une autre variable à un temps ultérieur. L'intégration du temps dans ce type d'analyse nous permet donc de prendre en compte les potentielles évolutions et différences entre les habiletés étudiées à chaque temps de mesure, qui correspond ici à une année scolaire. Nous avons ainsi choisi d'étudier l'évolution de trois variables à travers cette analyse : la lecture isolée de mots, la fluence de lecture (lecture en contexte) ainsi que l'orthographe d'usage.

Le choix d'utiliser deux variables liées à la lecture nous donne la possibilité de voir si le fait que la lecture se fasse de manière isolée (lecture de mots) ou en contexte (fluence de lecture de texte) a un impact sur les liens entretenus avec la variable orthographique. En effet, cela nous permet de distinguer les potentiels liens entre lecture et écriture d'une part dans le cas d'une lecture simple de type décodage (lecture de mots), et d'autre part dans le cas d'une lecture en contexte (lecture d'un texte). Les épreuves utilisées pour mesurer ces variables consistaient en une épreuve de lecture chronométrée en une minute de mots pour la lecture de mots ; de lecture chronométrée en une minute d'un texte pour la fluence de texte ; et du sous-score d'orthographe d'usage de la dictée *Le corbeau* (L2MA, Chevrie-Muller, 1997) pour la mesure d'orthographe lexicale.

Concernant les variables contrôles, nous avons choisi deux habiletés non verbales d'importance dans l'acquisition de la lecture et de l'écriture : la logique non verbale (Matrices de la WPPSI-III, Wechsler, 2004) ainsi que la dénomination rapide alphanumérique. L'objectif est ici d'éviter autant que possible de contaminer le modèle avec de la variance qui serait en réalité attribuable à des facteurs externes aux variables cibles. Ces deux variables contrôles ont été mesurées en début de CP (année scolaire 2014-2015). Les mesures des variables expérimentales ont été récoltées quant à elles du CE1 au CM2, en fin d'année scolaire. Pour cette analyse, l'échantillon se composait de 509 participants.

6.3 Résultats

Nous allons présenter, dans un premier temps, les résultats descriptifs concernant l'évolution des performances en lecture et en écriture du CP au CM2. Il s'agit de présenter l'évolution des trois variables expérimentales choisies sur cinq temps de mesure. Nous présenterons ensuite les corrélations entre ces différentes variables, avant de terminer par le modèle *cross-lagged* et ses conclusions sur l'évolution des liens entre les habiletés liées à la lecture et à l'écriture.

6.3.1 Résultats descriptifs de l'évolution des performances en lecture et écriture du CP au CM2

Tableau 7. Nombre de mots correctement lus en une minute à l'épreuve de lecture de mots et de fluence de texte : moyennes et écart-types par année scolaire.

	CP	CE1	CE2	CM1	CM2
Lec_Mots	37,42 (15,87)	61,17 (15,54)	70,12 (15,20)	78,61 (16,61)	86,22 (17,31)
Fluence	38,50 (25,01)	83,01 (35,31)	105,82 (37,04)	124,34 (38,33)	142,05 (38,23)

Note. Lec_Mots correspond à la lecture de mots isolés.

Le Tableau 7 présente l'évolution des performances moyennes en lecture de mots et fluence de texte du CP au CM2. Nous constatons une augmentation régulière de ces performances moyennes, qui correspondent au nombre de mots correctement lus en une minute, soit en lecture isolée pour la lecture de mots, soit en lecture en contexte pour la fluence de texte. Entre la fin du CP et la fin du CM2, le nombre moyen de mots correctement lus en une minute a plus que doublé (respectivement 37.42 mots correctement lus en une minute et 86.22). Si la difficulté des mots à lire augmente au fur et à mesure de l'exercice, on remarque la bonne progression des élèves sur les cinq années de l'école élémentaire, avec un progrès très important entre le CP et le CE1, puis une augmentation plus modeste mais régulière entre le CE1 et le CM2. Concernant la fluence de lecture de texte, nous constatons une forte progression entre le CP et le CE1 (de 38.50 mots lus en moyenne à 83.01 mots). On observe, de manière générale, une augmentation des scores jusqu'en CM2, montrant la progression des élèves en fluence de lecture de textes jusqu'à la fin de l'école élémentaire.

Les écart-types varient aussi tout au long des années scolaires. En lecture de mots, les écart-types restent stables au début et en milieu de scolarité, et semblent très légèrement augmenter aux Cours Moyens (16.61 en CM1 et 17.31 en CM2). En lecture de texte, les écart-

types semblent s'accroître sur l'ensemble de la période étudiée : de 25.01 mots en moyenne au CP, l'écart-type augmente à 35.31 mots en CE1, puis se stabilise jusqu'en CM2 (38.23 mots). Cette augmentation des écart-types montre que les écarts entre les élèves les plus et les moins performants se creusent de plus en plus au cours des niveaux scolaires.

Enfin, si les performances des deux activités de lecture sont similaires en CP (respectivement 37.42 mots lus de manière isolée, et 38.50 de manière contextuelle), nous remarquons une différence notable dans leur évolution : la lecture de mots isolés a bien augmenté jusqu'en CM2 (pour une moyenne de 86.22 mots), mais la lecture de mots en contexte a augmenté nettement plus (142.05 mots lus en contexte). Cette distinction dans l'évolution des performances pour la même activité de lecture pourrait appuyer l'importance du contexte dans la lecture, mais est à relativiser compte tenu des nombreux mots de liaison inclus dans les textes (déterminants, etc.).

Tableau 8. Scores moyens obtenus à l'épreuve de dictée *Le corbeau* : score phonologique, score d'usage et score grammatical, du CE1 au CM2.

	CE1	CE2	CM1	CM2
Phonologie Score max CE1 à CM1=15, CM2=17	9,89 (3,78)	12,06 (3,24)	13,64 (2,14)	15,32 (2,49)
Usage Score max CE1 à CM1=22, CM2=30	6,82 (4,44)	10,94 (5,07)	13,70 (4,65)	20,21 (5,25)
Grammaire Score max CE1 à CM1=13, CM2=23	3,32 (1,92)	4,93 (2,54)	6,45 (2,77)	13,71 (4,70)

L'épreuve de dictée *Le corbeau* était composée de trois catégories : score phonologique, score d'orthographe d'usage, score d'orthographe grammaticale ; les scores moyens obtenus pour chaque catégorie ainsi que les écart-types sont présentés dans le Tableau 8. Le sous-score « phonologie » codait pour l'orthographe phonologiquement plausible d'un mot. Le sous-score « usage » codait pour l'orthographe lexicale correcte d'un mot, sans tenir compte des accords en nombre et en genre. Le sous-score « grammaire » codait pour l'orthographe grammaticale du mot. Nous précisons à nouveau ici que, selon le protocole de l'épreuve, la dictée en CM2 contenait une phrase supplémentaire, majorant le total de 25 mots, et augmentant ainsi les possibilités d'erreurs des élèves. Concernant le sous-score phonologique, nous remarquons

qu'il s'accroît très rapidement en début d'apprentissage, puis semble se stabiliser dans le temps, proche d'un effet plafond de l'épreuve (9.89 mots phonologiquement corrects en CE1 en moyenne sur un maximum de 15, contre 15.32 en CM2 sur un maximum de 17). L'orthographe d'usage semble également progresser régulièrement tout au long de la scolarité élémentaire : les élèves écrivaient correctement seulement environ un mot sur trois en CE1 (6.82 mots correctement orthographiés en moyenne sur 22), puis augmentaient assez fortement leur performance en CE2 (10.94), pour ensuite stabiliser leur progression sur la fin de la scolarité élémentaire (respectivement 13.70 mots sur 22 en CM1 et 20.21 mots sur 30 en CM2), doublant ainsi leur taux d'écriture correcte des mots entre le CE1 et le CM2. Le score grammatical est quant à lui le plus lent à progresser, avec 3.32 mots grammaticalement corrects sur 13 en CE1 ; ce score moyen progresse en CE2 pour atteindre 4.93, et se stabilise en fin de scolarité, avec 6.45 mots grammaticalement corrects, soit presque un mot sur deux en fin de CM1. Ce taux reste stable en CM2, avec 13.71 mots grammaticalement corrects en moyenne, sur un total maximal possible de 23 mots.

Concernant les écart-types dans cette épreuve de dictée, l'écart-type du score phonologique semble relativement stable au fil des années, il a même tendance à décroître légèrement (de 3.78 en CE1 à 2.49 en CM2). L'écart-type de l'orthographe d'usage est le plus important comparativement aux deux autres scores, ce qui est à relativiser compte tenu du fait que cela soit le score avec le maximum le plus élevé (22 du CE1 au CM1 contre 15 pour le maximum en phonologie et 13 en grammaire). Cet écart-type de l'usage reste également relativement stable avec le temps (de 4.44 en CE1, à 5.25 en CM2). Enfin, en ce qui concerne l'orthographe grammaticale, l'écart-type moyen semble également stable avec le temps : s'il croît en CM2, c'est sans doute car le score maximum augmente fortement sur cette année-là : l'écart-type était jusqu'alors compris entre 1.92 (CE1) et 2.77 (CM1), et augmente à 4.70 (CM2), là où le score maximal possible pour cette épreuve passe de 13 à 23. Si nous présentons ici l'ensemble des sous-scores obtenus à cette épreuve de dictée, nous ne présenterons dans la partie suivante que les corrélations correspondant au sous-score d'orthographe lexicale, qui est le sous-score retenu pour le modèle *cross-lagged* par la suite.

6.3.2 Corrélations entre les habiletés de lecture et d'orthographe

Après avoir décrit l'évolution des habiletés de lecture et d'orthographe choisies pour cette étude, nous allons maintenant présenter les corrélations entre ces habiletés. Nous

présenterons d'abord les corrélations entre les différents temps de mesure de la même habileté. Nous présenterons ensuite les corrélations entre les trois habiletés testées à un même niveau scolaire.

Concernant la lecture de mots isolés (Tableau 9), les corrélations sont dans l'ensemble fortes entre tous les temps de mesure, ce qui semble tout à fait cohérent, compte tenu de la nature inchangée de l'épreuve pendant les cinq années d'expérimentation. Nous remarquons également que la force de cette corrélation décroît avec l'écart dans le temps des temps de mesures. Ainsi, par exemple, la lecture de mots CP corréle très fortement avec la lecture de mots CE1 (.805), mais cette corrélation s'atténue avec le temps (.713 avec le CE2, .660 avec le CM1 et .644 avec le CM2).

Tableau 9. Corrélations de l'épreuve de lecture de mots isolés du CP au CM2

	Lec_Mots CP	Lec_Mots CE1	Lec_Mots CE2	Lec_Mots CM1	Lec_Mots CM2
Lec_Mots CP	1	.805	.713	.660	.644
Lec_Mots CE1	.805	1	.833	.785	.744
Lec_Mots CE2	.713	.833	1	.835	.805
Lec_Mots CM1	.660	.785	.835	1	.851
Lec_Mots CM2	.644	.744	.805	.851	1

Note. Lec_Mots correspond à la lecture de mots isolés.

En ce qui concerne la fluence de textes (Tableau 10), nous observons sensiblement la même chose : les corrélations sont très fortes, comme attendu, entre deux mesures de la fluence de textes qui se succèdent à un an d'intervalle. Elles décroissent légèrement quand le temps qui sépare les deux mesures augmente.

Tableau 10. Corrélations de la fluence de textes du CP au CM2

	Fluence CP	Fluence CE1	Fluence CE2	Fluence CM1	Fluence CM2
Fluence CP	1	.814	.752	.726	.688
Fluence CE1	.814	1	.875	.851	.810
Fluence CE2	.752	.875	1	.918	.883
Fluence CM1	.726	.851	.918	1	.925
Fluence CM2	.688	.810	.883	.925	1

Enfin, concernant l'orthographe d'usage (Tableau 11), les corrélations sont également fortes et décroissantes avec l'écart dans les temps de mesure.

Tableau 11. Corrélations de l'orthographe d'usage du CE1 au CM2

	Usage CE1	Usage CE2	Usage CM1	Usage CM2
Usage CE1	1	.770	.695	.637
Usage CE2	.770	1	.761	.712
Usage CM1	.695	.761	1	.767
Usage CM2	.637	.712	.767	1

L'observation la plus intéressante ici concerne toutefois les corrélations entre ces trois habiletés par temps de mesure. En CE1 (Tableau 12), les deux habiletés de lecture sont logiquement très fortement corrélées (.880) et corrélées de manière sensiblement identiques à l'orthographe d'usage (respectivement .668 pour la lecture de mots et .673 pour la fluence de texte).

Tableau 12. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CE1

	Lec Mots	Fluence	Usage
Lec_Mots	1	.880	.668
Fluence	.880	1	.673
Usage	.668	.673	1

Note. Lec_Mots correspond à la lecture de mots isolés.

En CE2 (Tableau 13), nous observons le même fonctionnement : une corrélation très forte pour les deux habiletés de lecture (.844) même si elle baisse très légèrement par rapport au CE1, et une corrélation forte pour ces deux habiletés avec l'orthographe d'usage, même si l'on observe une corrélation stable en CE2 entre fluence et orthographe d'usage (.686 contre .673 en CE1) tandis que la corrélation entre lecture de mots et orthographe d'usage décroît très légèrement par rapport au CE1 (.642 contre .668 en CE1).

Tableau 13. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CE2

	Lec Mots	Fluence	Usage
Lec_Mots	1	.844	.642
Fluence	.844	1	.686
Usage	.642	.686	1

Note. Lec_Mots correspond à la lecture de mots isolés.

Ces corrélations se retrouvent de la même manière en CM1 (Tableau 14) : on observe une corrélation très forte entre les habiletés de lecture, et une corrélation forte entre chacune de ces habiletés et l'orthographe d'usage. La corrélation entre fluence et orthographe d'usage reste sensiblement identique au CE2 (.683), de même que la corrélation entre lecture de mots et orthographe d'usage (.621).

Tableau 14. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CM1

	Lec Mots	Fluence	Usage
Lec_Mots	1	.868	.621
Fluence	.868	1	.683
Usage	.621	.683	1

Note. Lec_Mots correspond à la lecture de mots isolés.

Enfin, en CM2 (Tableau 15), la corrélation entre lecture de mots et fluence est toujours très forte (.847), et l'orthographe d'usage corrèle avec la même force avec la lecture de mots (.640) et la fluence de texte (.680).

Tableau 15. Corrélations entre la lecture de mots, la fluence de texte et l'orthographe d'usage en CM2

	Lec Mots	Fluence	Usage
Lec_Mots	1	.847	.640
Fluence	.847	1	.680
Usage	.640	.680	1

Note. Lec_Mots correspond à la lecture de mots isolés.

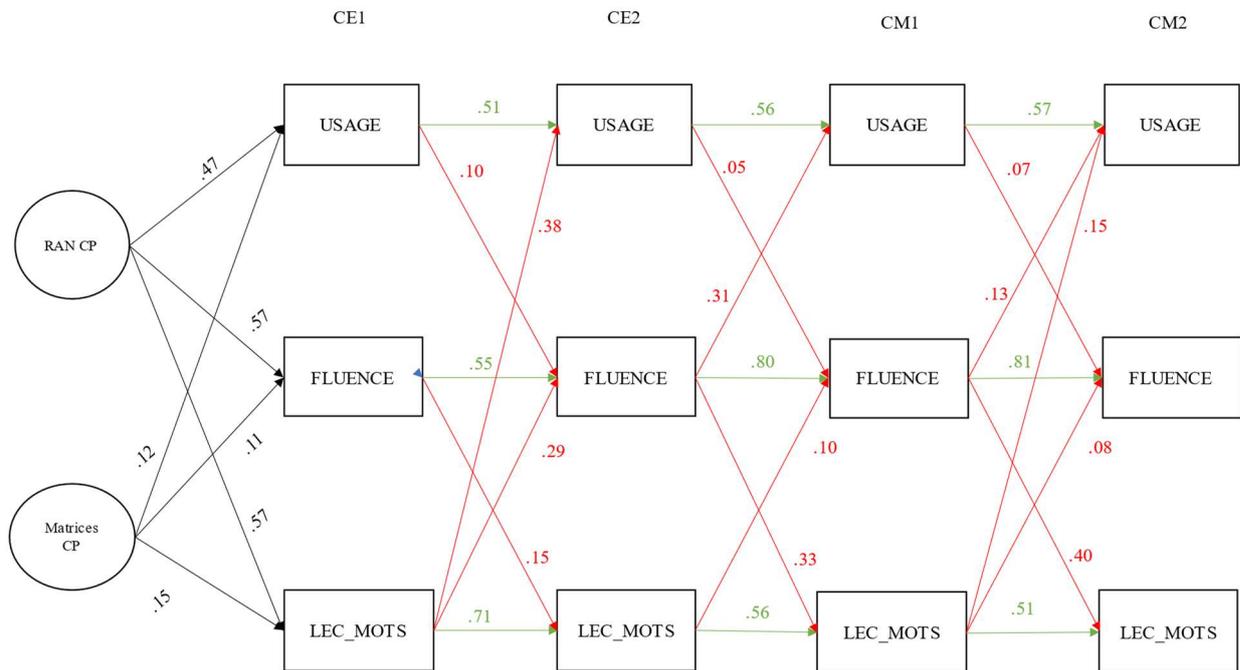
Pour conclure sur ces corrélations, elles font apparaître très clairement une relation très forte entre les deux habiletés liées à la lecture (lecture de mots et fluence de texte) alors que l'orthographe d'usage est une habileté un peu à part et ce, à chaque année scolaire. Cependant, il est clair aussi que lecture et orthographe d'usage sont significativement corrélés à tous les niveaux scolaires. Nous allons maintenant présenter le modèle *cross-lagged* qui investigate l'évolution des liens entre ces trois habiletés du CE1 au CM2.

6.3.3 Résultats du modèle *cross-lagged* sur l'évolution de la lecture et de l'écriture du CE1 au CM2

Notre objectif principal est d'essayer de voir si les progrès en lecture améliorent l'orthographe, et vice-versa. Grâce à notre suivi longitudinal sur quatre années scolaires, nous pourrions également observer l'évolution de ces liens lecture-orthographe entre le début et la fin de l'école élémentaire. Pour répondre à cet objectif, nous avons fait le choix d'un *cross-lagged panel model* (CLPM). L'avantage de ce type de modèle est de pouvoir mettre en évidence les différentes pistes causales au cours du temps, dans un même modèle, en testant les directions des relations d'une habileté vers l'autre. Ces modèles *cross-lagged* présentent aussi certaines limites, notamment l'indistinction des variances inter- et intra-individuelles. Toutefois, les méthodologies élaborées pour résoudre cette limite sont aujourd'hui largement discutées (pour une discussion récente sur cette question, voir Núñez-Regueiro et al., 2022).

Le modèle a été construit de la façon suivante. Deux variables de contrôle ont été introduites dans le modèle afin de ne pas confondre, dans l'évolution des habiletés spécifiques de lecture-écriture, ce qui relève en réalité d'habiletés générales autres : à savoir, la dénomination rapide et la logique non verbale (matrices). Ces deux variables, mesurées au début du CP, ont donc été intégrées comme prédicteurs des habiletés de lecture-écriture au premier temps de mesure retenu dans le modèle (CE1). Le modèle a ensuite été spécifié de telle sorte que chaque temps de mesure prédise le suivant (et seulement le suivant) : autorégression d'une habileté sur-elle-même au temps précédent, et relation croisée entre chaque type d'habileté (toutes les combinaisons de croisement potentiels entre les habiletés d'un temps sur l'autre ont été spécifiées). A partir de ce modèle initial considéré comme « complet », toutes les pistes non significatives ont été supprimées. Les indices d'ajustement montrent que ce modèle s'ajuste bien aux données (CFI=.982 ; TLI=.968 ; RMSEA=.066 (IC .055-.077) ; SRMR=.031). Le modèle a été estimé avec une méthode de maximum de vraisemblance robuste (MLR), qui permet de traiter les cas de non-normalité : elle permet d'estimer les erreurs-types qui sont robustes à la non-normalité. Le modèle obtenu est présenté en Figure 1. Dans un souci de clarté, les covariances et résidus ont été enlevés du modèle présenté en Figure 1, mais la figure complète (Figure 2) est proposée en Annexe L.

Figure 1. Modèle *cross-lagged* sur l'évolution des liens entre orthographe d'usage (USAGE), fluence en contexte (FLUENCE) et lecture de mots isolés (LEC_MOTS) du CE1 au CM2. Les variables prédictrices début CP sont la dénomination rapide alphanumérique (RAN) et les matrices (MATRICES). En vert sont présentés les liens autorégressifs ; en rouge les liens croisés entre les variables au fil du temps. Seuls les liens significatifs sont présentés.



Tout d'abord, concernant les variables contrôles utilisées, nous remarquons que les matrices ont un poids faible sur les trois habiletés étudiées (de .11 à .15, voir Figure 1). A contrario, la dénomination rapide alphanumérique en début de CP pèse fortement sur les acquisitions en fin de CE1, dans chacune des trois habiletés : elle est respectivement de .47 pour l'orthographe d'usage, de .57 pour la fluence de lecture et de .57 également pour la lecture de mots. Nous allons tout d'abord présenter les trajectoires de chacune des trois habiletés étudiées dans le temps, puis nous présenterons les trajectoires croisées entre ces habiletés sur les différents temps de mesure.

L'orthographe d'usage présente un effet autorégressif stable dans le temps : son poids reste sensiblement identique à chaque temps de mesure, avec tout de même un léger accroissement au fil du temps, notamment entre le CE2 et le CM1 : en effet, en CE2, le poids de l'orthographe d'usage de CE1 était de .51 ; le poids de cette même habileté en CE2 est de .56 sur les performances en CM1, et de .57 pour les performances de CM1 sur les performances de CM2. Les performances en orthographe d'usage révèlent donc un effet autorégressif fort et apparaissent ainsi comme grandement dépendantes de leur propre niveau préalable.

La fluence de lecture, quant à elle, est très fortement prédite par ses performances préalables, et ce phénomène semble également s'amplifier avec le temps. Nous observons en effet que les performances initiales en CE1 des élèves expliquent déjà une grande part des performances ultérieures en CE2 (.55), mais cette part s'accroît encore avec le temps (.80 pour le CM1 et .81 en CM2). Ainsi, la fluence de lecture en contexte a un poids très important dans sa propre évolution ; si ce poids est déjà fort en début de scolarité, il se renforce encore avec le temps.

Les performances initiales en lecture de mots (CE1) ont un fort impact sur le temps suivant en début d'apprentissage, puisqu'elles ont un poids très fort dans l'explication des performances ultérieures (.71 en CE2), et diminuent au fil du temps, même si elles restent importantes jusqu'en fin de scolarité élémentaire (respectivement .56 du CE2 au CM1, et .51 du CM1 au CM2). Les capacités préalables en lecture de mots isolés restent donc importantes tout au long de la scolarité pour expliquer les performances ultérieures des élèves dans cette habileté. Nous remarquons par ailleurs que le poids des performances préalables en lecture de mots isolés semble aller à l'inverse du poids des performances en fluence de lecture en contexte : quand les performances préalables en lecture de mots isolés pèsent fortement sur les performances ultérieures en début d'apprentissage puis décroissent, les performances en lecture de mots en contexte (fluence) gagnent en poids en cours de scolarité.

En ce qui concerne les influences de ces trois habiletés les unes par rapport aux autres (*cross-lagged*) sur différents temps de mesure, nous remarquons dans un premier temps plusieurs différences entre les deux habiletés de lecture choisies et l'habileté d'écriture : effectivement, si la lecture de mots et la fluence de lecture influencent toutes deux à un moment dans la scolarité l'orthographe d'usage, celle-ci semble n'influencer que la fluence de lecture, et non la lecture de mots isolés. Le modèle ne montre en effet aucun lien significatif direct entre l'orthographe d'usage et la lecture de mots isolés sur les quatre temps de mesure.

L'orthographe d'usage n'a qu'un poids faible, qui plus est uniquement sur la fluence ultérieure, et ce poids diminue avec le temps : l'orthographe d'usage en CE1 a un poids de .10 sur la fluence de lecture en CE2 ; en CE2, son poids n'est plus que de .05 sur la fluence de lecture CM1, et, de même, la fluence de lecture CM2 n'est que très faiblement influencée par l'orthographe d'usage CM1 (.07). L'orthographe d'usage ne semble donc intervenir que très faiblement, mais néanmoins significativement, dans les performances ultérieures en fluence de

lecture en contexte. Il est toutefois intéressant de noter que ce poids, s'il reste très faible comme nous l'avons évoqué, n'est cependant pas inexistant, contrairement au poids de l'orthographe d'usage sur la lecture de mots isolés.

La lecture de mots isolés en début de scolarité semble avoir un poids relativement important à la fois dans les acquisitions orthographiques (.38) ainsi qu'en fluence (.29) en CE2. L'année suivante, ce poids semble n'avoir plus qu'un faible impact, uniquement sur la fluence (.10 sur la fluence CM1), puis, en fin de scolarité, les performances en lecture de mots CM1 semblent n'avoir qu'un faible impact à la fois sur les performances ultérieures en fluence de texte (.08) et sur les performances orthographiques en CM2 (.15). Nous remarquons cependant que le moment où le poids de la lecture de mots préalable est le plus fort sur les acquisitions orthographiques ultérieures correspond au moment où la fluence de lecture de texte n'a pas d'effet significatif sur les performances orthographiques ultérieures (du CE1 au CE2). Nous pouvons également noter que le poids des performances préalables en lecture de mots isolés, qu'il soit sur la fluence de lecture de texte ultérieure ou bien sur l'orthographe d'usage ultérieure, décroît avec le temps.

La fluence de lecture de texte, ou lecture en contexte, ne présente quant à elle pas le même schéma que la lecture de mots isolés. La fluence de lecture préalable a un poids sur la lecture de mots ultérieure qui s'accroît au fil de la scolarité : ainsi, la fluence CE1 a un poids très faible (.15) sur la lecture de mots CE2 ; ce poids augmente l'année suivante (.33 sur lecture de mots CM1), pour arriver à un poids assez fort en CM2 (.40). La fluence de lecture semble également jouer un rôle dans le développement de l'orthographe d'usage : si la fluence de lecture CE1 n'a pas de lien significatif sur l'orthographe d'usage CE2, elle acquiert de l'importance avec le temps (.31 sur le CM1), mais semble cependant décroître en fin de scolarité élémentaire (.13), moment où la lecture de mots a également un poids significatif dans l'explication des performances orthographiques de CM2.

Concernant la part de variance restant à expliquer à chaque temps de mesure, nous remarquons des différences selon que l'on s'intéresse à l'habileté d'orthographe ou aux habiletés de lecture. Nous présentons en Annexe L la Figure 2, qui présente le modèle avec les covariances et la variance non expliquée, afin de ne pas surcharger le modèle présenté plus tôt (Figure 1). Dans l'ensemble, le pourcentage de variance résiduelle en CE1 est élevé pour toutes les habiletés, puisqu'à ce stade-là il n'y a pas d'effet autorégressif. Entre le CE2 et le CM2, l'ensemble des variables présente un fort pouvoir explicatif, même si l'orthographe d'usage reste la moins fortement prédite. En effet, pour l'orthographe d'usage, à la fin du CE1 il reste

75% de variance non expliquée ; ce pourcentage diminue fortement pour ne représenter plus que 35% à la fin du CE2, puis 35% également à la fin du CM1, pour finir à 38% à la fin du CM2. En ce qui concerne les habiletés de lecture, la variance restant à expliquer pour la lecture de mots isolés passe de 63% à la fin du CE1 et diminue jusqu'à 22% seulement à la fin du CM2. On observe une diminution similaire pour la fluence de lecture en contexte (lecture de texte), avec 65% de variance restant à expliquer à la fin du CE1, et seulement 13% à la fin du CM2.

6.4 Conclusion

Cette première étude avait pour objectif d'étudier les relations entre l'orthographe et la lecture sur un temps long, soit ici quatre années scolaires, correspondant au temps de l'école élémentaire (du CE1 au CM2). Deux habiletés ont été contrôlées en début de CP (la dénomination rapide alphanumérique et les matrices), puis trois habiletés ont été mesurées en fin de CE1, CE2, CM1 et CM2. Pour la lecture, nous avons choisi d'une part la lecture de mots isolés ainsi que la fluence de lecture en contexte, ou fluence de texte. Pour l'écriture, il s'agissait de l'orthographe d'usage dans un exercice d'écriture sous la dictée.

Le modèle *cross-lagged* a permis de mettre en évidence les différentes relations entre ces trois habiletés et l'évolution de ces relations du CE1 au CM2. Chaque habileté à un temps donné a un poids très important dans son développement au temps suivant ; toutefois, ce poids autorégressif varie selon l'habileté étudiée. En effet, pour l'orthographe d'usage ce poids augmente petit à petit à chaque temps de mesure ; pour la fluence de texte, il y a une nette augmentation entre le CE2 et le CM1, puis une augmentation très légère en CM2 ; enfin, pour la lecture de mots isolés, le poids est à l'inverse très fort entre le CE1 et le CE2 puis diminue jusqu'au CM2, même s'il reste fort.

Si la fluence de texte semble impacter à la fois les performances en orthographe d'usage et en lecture de mots, l'orthographe d'usage n'a d'effet significatif que sur la fluence de texte ; ce poids reste toutefois faible. Une piste d'explication possible pour cette particularité serait les niveaux de langue évoqués par Abbott et collaborateurs (2010), avec comme argument que l'épreuve utilisée pour mesurer l'orthographe d'usage et celle utilisée pour mesurer la fluence de texte concernent le niveau texte de la langue, alors que l'épreuve utilisée pour mesurer la lecture de mots se situe au niveau mot de la langue. Une autre explication pourrait résider dans

la différence des deux habiletés utilisées pour l'évaluation des performances en lecture : la fluence de texte (lecture en contexte) prend de plus en plus de poids dans l'explication des performances ultérieures en lecture de mots isolés au fil du temps tandis que nous observons le fonctionnement inverse pour la lecture de mots isolés ; or la lecture de mots isolés, ou décodage, est une habileté nécessaire pour le développement de la lecture notamment en tout début d'apprentissage (selon le modèle simple de la lecture, Gough & Tunmer, 1986). Ainsi, le changement de poids des habiletés correspond à la modification de l'importance de ces habiletés dans l'acquisition de la lecture : le décodage, devenu automatisé, permettrait à une lecture plus globale du texte de prendre davantage d'importance dans l'explication des performances ultérieures.

Le changement le plus important dans l'impact des habiletés les unes par rapport aux autres se situe entre le CE2 et le CM1 : à ce temps de mesure, le poids de la fluence de texte de CE2 par rapport aux deux autres habiletés ultérieures, même s'il reste modéré, est le plus important, et les impacte de manière similaire (.31 vers l'orthographe CM1, .33 vers la lecture de mots CM1). Cela correspond également au moment où le poids de la fluence préalable augmente le plus (.65 du CE1 au CE2 contre .80 du CE2 au CM1), et où le poids des deux autres habiletés préalables est le plus faible dans l'explication des progrès de fluence de texte CM1 (respectivement .05 de l'usage CE2 et .10 de la lecture de mots CE2).

Ce modèle ne permet donc pas de valider notre première hypothèse, selon laquelle un lien unilatéral allant de la lecture vers l'écriture serait visible, et d'autant plus fort en début de scolarité. En effet, les résultats montrent un lien bilatéral entre lecture et écriture, tout au long de la scolarité élémentaire. Ce lien n'est toutefois pas égal : l'orthographe ne semble que prédire faiblement, mais de manière non nulle et significative, les performances ultérieures en lecture en contexte. Ce lien de l'écriture vers la lecture décroît avec le temps. En revanche, le lien de la lecture vers l'écriture semble se modifier avec le temps : le décodage a un pouvoir explicatif modéré des performances ultérieures en écriture en début de scolarité ; à ce lien vient s'ajouter l'impact de la fluence en contexte au fur et à mesure que l'apprentissage progresse. Ces résultats tendent donc à supposer l'existence d'un lien bilatéral dès le CE1, donc en début de scolarité, avec un impact plus fort de la lecture vers l'écriture, confortant les résultats trouvés par Abbott et collaborateurs (2010), et *a contrario* du lien unilatéral trouvé dans d'autres études longitudinales (e.g. Deacon et al., 2012).

Notre deuxième hypothèse est quant à elle validée : les résultats montrent une distinguabilité entre la lecture de mots isolés et la lecture en contexte (fluence de texte). Nous

constatons en effet que la lecture en contexte prend de plus en plus de poids au fil du temps dans l'explication des performances ultérieures en lecture de mots isolés, cette dernière perdant année après année de l'influence sur les performances ultérieures en lecture en contexte. En plus de ces observations croisées entre fluence en contexte et lecture de mots isolés, nous observons le même fonctionnement sur l'évolution dans le temps de chacune de ces habiletés : la lecture de mots isolés a un poids de moins en moins important dans la prédiction des performances ultérieures en lecture de mots isolés ; à l'inverse, la fluence de texte est, avec le temps, de plus en plus importante dans la prédiction des performances ultérieures dans cette habileté.

Les résultats du modèle *cross-lagged* ont donc permis de mettre en avant le lien bilatéral entre les habiletés de lecture et d'orthographe, tout au long de l'école élémentaire, bien que le lien venant de l'orthographe soit faible. Les résultats montrent en effet un déséquilibre fort entre les poids des liens entre orthographe et lecture, cette dernière impactant de manière beaucoup plus importante l'orthographe. Si la part de variance restant à expliquer en fin d'étude (CM2) est faible pour les variables liées à la lecture, elle est néanmoins plus importante en ce qui concerne l'orthographe d'usage. Nous allons donc nous intéresser maintenant de plus près à cette variable orthographique, et notamment à sa composition et aux variables qui peuvent intervenir dans l'évolution des performances orthographiques.

Chapitre 7 : Etudes 2 et 3 : études du lien entre orthographe lexicale et sublexicale en début d'apprentissage formel

L'objectif de ces deux études était de s'intéresser de manière plus spécifique au lien entre l'orthographe lexicale et sublexicale en début de scolarité élémentaire, c'est-à-dire, d'étudier le lien potentiellement bidirectionnel entre orthographe lexicale et sublexicale, ainsi que son évolution dans le temps. Il s'agissait donc de voir si ces deux habiletés, communes à l'apprentissage orthographique, ont des prédicteurs communs mais également des prédicteurs propres à chacune. Pour ce faire, deux analyses de régression ont été menées. L'une considère les mesures prises en fin de CP et étudie leur pouvoir prédictif sur des performances en orthographe lexicale et sublexicale en fin de CE2. La suivante étudie le pouvoir prédictif des mesures prises en fin de CE1 sur ces mêmes performances orthographiques de fin de CE2.

7.1 Hypothèse

L'hypothèse est donc la suivante : orthographe lexicale et sublexicale ont des prédicteurs communs en début de scolarité, mais également des prédicteurs propres à chacune de ces habiletés puisqu'elles sont dissociables ; si certaines études les considèrent comme faisant partie d'une seule habileté orthographique (e.g. Cunningham, 2006), d'autres les distinguent dès le début de la scolarité (e.g. Binamé & Poncelet, 2016). Notre hypothèse se situe donc dans la perspective d'une dissociation de ces deux composants de l'orthographe.

7.2 Matériel et méthode

Concernant l'étude du CP au CE2, l'échantillon était composé de 458 élèves. Pour la seconde étude, du CE1 au CE2, l'échantillon était composé de 591 élèves. La différence du nombre d'élèves entre les deux échantillons s'explique par le nombre d'élèves absent sur un des deux temps de mesure (CP et CE2, ou CE1 et CE2) sur au moins une des épreuves. Outre

le fait que le CE2 soit une classe située à la fin d'un cycle d'apprentissage scolaire en France (le cycle 2, commencé en CP), il s'agissait également dans notre étude du niveau scolaire pour lequel deux mesures d'orthographe lexicale étaient proposées, via deux types de tâches différentes : une tâche de production et une tâche de choix forcé. C'est pour ces raisons que les études 2 et 3 ont été effectuées sur un temps allant du CP au CE2.

Pour ces deux études, les variables dépendantes étaient des mesures orthographiques CE2 : il s'agissait, pour la mesure de la sensibilité graphotactique, d'une tâche de choix de pseudomots ; concernant les mesures d'orthographe lexicale, nous avons deux épreuves : une tâche de production sous la dictée (dictée Le corbeau) et une tâche de choix orthographique.

Les variables indépendantes étaient les suivantes :

En CP, pour la première étude, les mesures utilisées comprenaient le décodage (correspondant à la somme des épreuves de lecture de mots, de lecture de pseudomots et de fluence de texte sur deux textes), la dictée lexicale (somme des épreuves de dictée de mots et dictée de phrases), la dictée de pseudomots, le vocabulaire (somme des épreuves de vocabulaire en étendue mesurée avec l'EVIP et syntaxe mesurée avec l'ECOSSE), ainsi que la compréhension orale.

En CE1, pour la seconde étude, les mesures utilisées comprenaient pour la lecture : la lecture de mots, de pseudomots ainsi que la fluence de textes. Pour les mesures de l'orthographe étaient testées la dictée de mots et la dictée de pseudomots, ainsi que la sensibilité graphotactique. Enfin, le vocabulaire en profondeur (subtest vocabulaire du WISC IV), la compréhension orale et la compréhension écrite étaient également mesurées.

7.3 Résultats

Les modèles statistiques ont été effectués grâce au logiciel SPSS (version 25). Une analyse factorielle sur les items des épreuves d'orthographe lexicale et sublexicale a été réalisée préalablement pour vérifier la distinction entre les épreuves d'orthographe lexicale et sublexicale. L'analyse en composantes principales présentait un indice KMO très correct (.826), et mettait en évidence deux facteurs distincts pour l'ensemble des items intégrés à l'analyse. Ces facteurs comprenaient respectivement les items de l'épreuve de choix orthographique pour

la tâche d'orthographe lexicale pour le premier, et les items de l'épreuve de choix de pseudomots pour la tâche d'orthographe sublexicale pour le second.

La distinction entre orthographe lexicale et sublexicale se vérifie également à travers les analyses multiniveaux effectuées : la création d'un modèle vide permet en effet de voir s'il est pertinent de continuer dans une analyse multiniveau, ce premier niveau d'analyse permettant de mettre en évidence la présence d'un effet de la classe (dans notre cas) dans laquelle se situent les élèves. Nous remarquons donc une première distinction entre orthographe lexicale et sublexicale à ce niveau-là : le modèle vide de l'orthographe lexicale a montré l'intérêt de poursuivre dans une analyse multiniveau, la classe ayant un effet significatif sur les performances des élèves. A contrario, le modèle vide de l'orthographe sublexicale n'a pas montré d'effet de la classe ; nous avons donc, pour cette habileté, procédé à une analyse de régression.

7.3.1 Prédicteurs de l'orthographe sublexicale

Des analyses de régression ont donc été effectuées afin de mettre en évidence les prédicteurs, en CP et en CE1, de l'orthographe sublexicale en CE2.

Tableau 16. Analyse de régression montrant les prédicteurs en CP de l'orthographe sublexicale en CE2.

	β	t	sig.
Dictée Lexicale CP	.261***	3.686	<.001
Dictée Pseudomots CP	.009	.161	.872
Décodage CP	.011	.173	.863
Vocabulaire & Syntaxe CP	.038	.702	.483
Compréhension Orale CP	-.121*	-2.334	<.05

Note. R^2 global du modèle =.011

* p <.05. ** p <.01. *** p <.001

La sensibilité graphotactique CE2 est prédite de manière très significative par la dictée lexicale CP (β =.261, t =3.686, p <.001, voir Tableau 16), mais pas par la dictée de pseudomots

ni le décodage. De manière générale, les variables choisies comme potentiels prédicteurs ne permettent qu'une très faible explication des performances sur cette épreuve (1.1%).

Tableau 17. Analyse de régression montrant les prédicteurs en CE1 de l'orthographe sublexicale en CE2.

	β	t	sig.
Dictée de mots CE1	.251***	4.358	<.001
Fluence de texte CE1	.017	.181	.856
Compréhension écrite CE1	.058	.963	.336
Sensibilité graphotactique CE1	.292***	6.57	<.001
Dictée de pseudomots CE1	-.078	-1.602	.110
Lecture de pseudomots CE1	-.043	-.507	.612
Compréhension orale CE1	.023	.463	.644
Vocabulaire CE1	-.095	-1.884	.060

Note. R^2 global du modèle = .164

* $p < .05$. ** $p < .01$. *** $p < .001$

En CE1, la dictée de mots est un prédicteur significatif ($\beta = .251$, $t = 4.358$, $p < .001$, voir Tableau 17) de la sensibilité graphotactique ultérieure ; le pouvoir explicatif de la dictée de mots CE1 est sensiblement identique à celui de la sensibilité graphotactique CE1 (autorégression, $\beta = .292$, $t = 6.57$, $p < .001$). La dictée de mots, comprise dans la dictée lexicale mesurée en CP, est donc toujours prédictive en CE1 des performances ultérieures en orthographe sublexicale. Ici, le modèle global n'a qu'un faible pouvoir explicatif de cette habileté (16.4%), même s'il est supérieur à celui des prédicteurs de fin de CP.

7.3.2 Prédicteurs de l'orthographe lexicale

Des analyses multiniveaux ont été effectuées pour étudier les prédicteurs de l'orthographe lexicale ; en effet, la classe dans laquelle se situent les élèves a un impact sur les performances, que ce soit pour les prédicteurs de CP ou de CE1. Ceci se vérifie tout d'abord par la création d'un modèle vide, sans variable explicative, afin de permettre d'estimer la part de variance entre les différents « niveaux » de l'analyse. On considère alors un seuil au-delà

duquel l'analyse multiniveaux permettra une « meilleure » explication de la répartition de la variance, en considérant non seulement les variables explicatives choisies pour l'analyse, mais également les différences entre les niveaux d'analyse. Dans le cas de cette étude, le modèle vide a montré qu'une part significative de la variance s'expliquait par le niveau de la classe à laquelle appartiennent les élèves. Dans le cas de l'étude 2 sur l'orthographe sublexicale, par exemple, le modèle vide n'avait pas montré cet effet significatif du niveau de la classe, c'est pourquoi des analyses de régression ont été réalisées.

En ce qui concerne les prédicteurs de CP, le modèle 2 (Tableau 18) est le modèle qui s'ajuste le mieux aux données. En effet, c'est le modèle 2 pour lequel l'indice BIC était le plus faible (BIC=2733.293).

La dictée lexicale et le décodage sont significativement prédicteurs des performances ultérieures en reconnaissance orthographique (respectivement $b=2.353$, $t=.356$, $p<.001$ pour la dictée lexicale, et $b=.367$, $t=.315$, $p<.001$ pour le décodage). Dans une moindre mesure, la dictée de pseudomots ($b=-.770$, $t=.331$, $p<.05$) ainsi que le vocabulaire et la syntaxe ($b=.758$, $t=.324$, $p<.05$) sont également significatifs. La compréhension orale CP n'a toutefois pas d'effet sur les performances en reconnaissance orthographique ultérieure.

Tableau 18. Analyse multiniveaux montrant les prédicteurs en CP de l'orthographe lexicale en CE2 (tâche de reconnaissance orthographique)

Parameters	Model 0	Model 1	Model 2
Fixed effects			
intercept	61,611 (.445)***	61,274 (.327)***	61,131 (.275)***
Dictée de pseudomots CP		-.291 (.361)	-.770 (.331)*
Dictée lexicale CP		4.659 (.318)***	2.353 (.356)***
Décodage CP			3.367 (.315)***
Vocabulaire et Syntaxe CP			.758 (.324)*
Compréhension orale CP			.201 (.253)
Random effects			
<i>Level 2 (interclass variance)</i>			
intercept variance	4.265 (1.820)*	1.853 (.954)	1.018 (.703)
<i>Level 1 (intraclass variance)</i>			
	49.150 (2.990)***	25.356 (1.741)***	20.569 (1.432)***
-2 log L	3850.163	2834.44	2884.331
BIC	3969.262	2865.13	2733.293

* $p<.05$. ** $p<.01$. *** $p<.001$

En CE1, nous avons proposé deux types de tâches orthographiques : la reconnaissance orthographique (Tableau 19), comme en CP (tâche de choix forcé similaire à la tâche proposée pour la sensibilité graphotactique), et une tâche de production (Tableau 20, dictée). En effet, selon le type de tâche proposé au sein d'une même habileté, les habiletés mobilisées peuvent différer : ici en l'occurrence, la tâche de production requiert davantage d'habiletés de plus haut niveau (mémoire à court terme, capacités attentionnelles) que la tâche de choix forcé, pour laquelle les propositions sont déjà présentes.

Tableau 19. Analyse multiniveaux montrant les prédicteurs en CE1 de l'orthographe lexicale en CE2 (tâche de reconnaissance orthographique)

Parameters	Model 1	Model 2	Model 3
Fixed effects			
intercept	25.33 (.47)***	-1.16 (1.14)	-3.61 (1.36)**
Dictée de mots CE1		.52 (.07)***	.51 (.07)***
Fluence de texte CE1		.19 (.02)***	.19 (.02)***
Compréhension écrite CE1		.26 (.07)***	.25 (.07)***
Sensibilité graphotactique CE1		.28 (.07)***	.24 (.07)***
Sensibilité graphotactique CE2			.19 (.07)**
Random effects			
<i>Level 2 (interclass variance)</i>			
intercept variance	5.46 (2.00)**	.99 (.57)	.89 (.55)
<i>Level 1 (intraclass variance)</i>			
-2 log L	3813.61	2742.87	2696.23

* $p < .05$. ** $p < .01$. *** $p < .001$

Pour la tâche de reconnaissance orthographique (Tableau 19), tous les prédicteurs de CE1 choisis se sont montrés significatifs : dictée de mots ($b=.51$, $t=.07$, $p<.001$), fluence de lecture ($b=.19$, $t=.02$, $p<.001$), compréhension écrite ($b=.25$, $t=.07$, $p<.001$) et sensibilité graphotactique ($b=.24$, $t=.07$, $p<.001$) étaient très significatifs. Nous avons également un effet significatif de la sensibilité graphotactique en CE2 ($b=.19$, $t=.07$, $p<.01$) sur la reconnaissance orthographique de CE2, même lorsque la sensibilité graphotactique de CE1 est prise en compte. C'est avec la prise en compte de la sensibilité graphotactique au même temps de mesure (CE2) que le modèle (modèle 2) s'ajuste le mieux aux données.

Tableau 20. Analyse multiniveaux montrant les prédicteurs en CE1 de l'orthographe lexicale en CE2 (tâche de production sous forme d'écriture sous la dictée)

Parameters	Model 1	Model 2
Fixed effects		
intercept	10.91 (.35)***	-5.62 (.71)***
Dictée de mots CE1		.50 (.05)***
Fluence de texte CE1		.12 (.01)***
Compréhension écrite CE1		.16 (.05)**
Dictée de pseudomots CE1		
Random effects		
<i>Level 2 (interclass variance)</i>		
intercept variance	3.42 (1.18)**	.74 (.36)*
<i>Level 1 (intraclass variance)</i>		
-2 log L	3548.42	2568.76

* $p < .05$. ** $p < .01$. *** $p < .001$

En tâche de production (Tableau 20), la dictée de mots ($b = .50$, $t = .05$, $p < .001$), la fluence de lecture ($b = .12$, $t = .01$, $p < .001$) ainsi que la compréhension écrite de CE1 ($b = .16$, $t = .05$, $p < .01$) prédisent significativement les performances ultérieures en CE2, et ce de manière très significative pour la dictée de mots et la fluence de texte. La dictée de pseudomots n'était quant à elle pas significative.

Le type de tâche proposé pour une même habileté peut donc influencer le poids des prédicteurs potentiels ; en orthographe lexicale, les performances en dictée de mots, fluence de lecture et compréhension écrite préalables expliquent les performances ultérieures sur l'ensemble des tâches. Toutefois, sur une tâche spécifique de reconnaissance orthographique, qui implique donc des propositions orthographiques correctes et incorrectes, la sensibilité graphotactique est mobilisée et devient un prédicteur significatif.

7.4 Conclusion

Si l'orthographe lexicale semble dépendre en partie de l'enseignement effectué en classe, l'orthographe sublexicale quant à elle n'est pas influencée par les situations d'enseignement explicite et correspondrait donc plutôt à un apprentissage implicite des régularités orthographiques. Cela se constate également dans les résultats, puisque l'orthographe lexicale

présente un effet du niveau classe, requérant donc des analyses de type multiniveaux, alors que l'orthographe sublexicale ne présente pas cet effet. Cette caractéristique de l'orthographe sublexicale, non enseignée en classe, peut par ailleurs être une explication potentielle concernant le très faible pouvoir explicatif des modèles qui tentent d'étudier ses prédicteurs en se basant pour cela sur des prédicteurs connus des performances en lecture et en écriture.

Également, les prédicteurs de l'orthographe lexicale et sublexicale ne sont pas tous identiques. En effet, l'orthographe lexicale est prédite dans nos modèles par les dictées (dictée lexicale ou dictée de mots). En début d'apprentissage, d'autres habiletés comme la dictée de pseudomots, le décodage, le vocabulaire et la syntaxe sont aussi des facteurs importants pour les performances ultérieures en écriture. La fluence de texte et la compréhension écrite ont également un rôle sur ces performances ultérieures, quel que soit le type de tâche proposé pour l'évaluation de l'orthographe lexicale. Cependant, si la tâche choisie comme mesure de performance orthographique ultérieure est une tâche de reconnaissance orthographique, la sensibilité aux régularités orthographiques de la langue devient un prédicteur important, tant sur les acquisitions préalables que sur les acquisitions faites en cours d'année.

Concernant l'orthographe sublexicale, la dictée lexicale en début d'apprentissage ainsi que la sensibilité aux régularités orthographiques et la dictée de mots CE1 permettent d'expliquer une petite partie des performances orthographiques sublexicales en fin de CE2. Néanmoins, le faible pouvoir explicatif des modèles laisse penser que cette habileté implicite pourrait ne pas uniquement dépendre, en début de scolarité, d'habiletés spécifiques à la lecture ou à l'écriture : en début d'apprentissage, les habiletés habituellement prédictrices de la lecture et de l'écriture n'expliquent que très peu les performances ultérieures dans cette habileté, qui ensuite s'explique majoritairement par sa propre compétence préalable.

Notre hypothèse sur ces deux études est donc validée : les résultats mettent effectivement en avant une différence entre l'orthographe lexicale et l'orthographe sublexicale. Quelques prédicteurs communs ont pu être mis en évidence en début d'apprentissage, comme la dictée lexicale en CP et la dictée de mots en CE1 ; toutefois, aucune des habiletés classiquement admises comme impactant l'apprentissage de la lecture et de l'écriture ne semblent avoir de l'influence sur l'acquisition de l'orthographe sublexicale. Le second argument en faveur de cette distinction correspond à l'impact de l'apprentissage explicite fait

en classe, apprentissage qui impacte les performances en orthographe lexicale, mais pas en orthographe sublexicale.

Ces études ont, au-delà de notre hypothèse de départ, permis de mettre en évidence l'impact des mesures choisies : en effet, le type de tâche utilisé semble impacter le pouvoir explicatif des variables concernées. Cela se constate, par exemple, dans l'implication de l'orthographe sublexicale mesurée via une tâche de reconnaissance (sensibilité graphotactique), sur l'orthographe lexicale mesurée via une tâche de reconnaissance également (reconnaissance orthographique) mais pas via une tâche de production orthographique au même temps de mesure. Cependant, la mesure de production sous la dictée impacte les performances ultérieures en orthographe sublexicale.

Ainsi, si l'orthographe lexicale et sublexicale semblent mobiliser quelques prédicteurs communs, notamment au niveau de la capacité à restituer correctement des mots (dictées), leur lien ne semble pas bidirectionnel ; en effet, l'orthographe sublexicale ne joue un rôle important dans l'acquisition de l'orthographe lexicale seulement quand la tâche proposée pour mesurer l'orthographe lexicale est une tâche de choix orthographique, c'est-à-dire lorsque l'élève doit choisir, parmi plusieurs propositions, celle qui est correctement orthographiée. Cette tâche est, dans sa forme, similaire à la tâche proposée pour évaluer la sensibilité graphotactique (orthographe sublexicale), puisque l'élève doit ici choisir parmi deux non mots, celui qui aurait, selon lui, l'orthographe correspondant le plus à un mot de la langue française. Cela pourrait donc expliquer la part significative de variance expliquée par l'orthographe sublexicale lors de cette tâche, alors qu'elle n'explique aucune part de variance pour la tâche de production écrite. L'orthographe lexicale, quant à elle, participe bien aux performances ultérieures en orthographe sublexicale à partir du CE1 ; ainsi, le lien sous-tendant orthographe lexicale et sublexicale semble dépendre en partie de la tâche, c'est-à-dire de la finalité demandée à l'élève, et donc des habiletés à mobiliser pour arriver à effectuer l'exercice.

Ces études ont donc mis en évidence la diversité des prédicteurs précoces de l'orthographe lexicale et sublexicale, venant renforcer le postulat selon lequel ce sont deux habiletés distinctes dans l'apprentissage orthographique. En effet, ces études ont fait apparaître des prédicteurs majoritairement différents pour l'orthographe lexicale (dictée lexicale, décodage, connaissances syntaxiques, vocabulaire, compréhension écrite) et sublexicale. De même, les prédicteurs communs à ces deux habiletés orthographiques sont peu nombreux parmi

les prédicteurs connus de la lecture et de l'écriture (dictée lexicale, et sensibilité graphotactique préalable, selon le type de tâche choisie).

Chapitre 8 : Discussion générale

8.1 Synthèse

Ce travail de thèse avait pour objectif d'étudier la composante orthographique sous deux angles, l'un plus global et l'autre plus local. L'aspect global concernait l'étude de l'évolution du lien entre lecture et écriture du CP au CM2. Il s'agissait ensuite de préciser les liens, en début d'apprentissage, sous-tendant cette composante orthographique, et plus spécifiquement l'orthographe lexicale et l'orthographe sublexicale.

Notre première étude présentait un modèle *cross-lagged* sur l'évolution conjointe de trois habiletés, deux liées à la lecture (lecture de mots isolés, fluence en contexte) et une liée à l'écriture (orthographe lexicale). Cette analyse a montré des évolutions différentes de chacune de ces habiletés au cours de l'école élémentaire : l'orthographe lexicale reste un prédicteur fort de sa propre progression, qui a tendance à augmenter légèrement avec le temps. Au contraire, la lecture de mots isolés semble perdre en importance vers la fin de l'école élémentaire, alors qu'elle était très fortement impliquée dans les compétences ultérieures en lecture de mots isolés en début d'apprentissage. Enfin, la fluence de lecture en contexte reste stable tout au long de la scolarité élémentaire, avec une augmentation plus nette en milieu d'école élémentaire. Ainsi, avec le temps l'importance de chacune des habiletés varie : la lecture de mots isolés, par exemple, ne semble plus peser autant dans l'explication des compétences ultérieures de cette même habileté, tandis que la fluence de lecture en contexte prend de plus en plus d'importance dans l'explication de son propre développement. Les habiletés de lecture ne semblent donc pas suivre la même trajectoire au cours de l'apprentissage.

Les résultats ont montré un lien bilatéral entre lecture et écriture, infirmant notre première hypothèse. Le poids de ce lien est toutefois très inégal, avec une influence plus forte de la lecture vers l'écriture. Si la plupart des études longitudinales évoquées précédemment avaient montré un lien unilatéral de la lecture vers l'écriture en début d'apprentissage (e.g. Deacon et al., 2012 ; Schaars et al., 2017), quelques-unes faisaient toutefois état d'un lien

bilatéral (Ahmed et al., 2014 pour le niveau phrase seulement ; Shanahan & Lomax, 1986, 1988). L'inégalité dans ce poids, avec la supériorité de la lecture sur l'écriture, malgré ce lien bilatéral, semble correspondre avec la théorie de l'auto-apprentissage de Share (1995) : c'est en passant par l'exposition à l'écrit, et donc par l'activité de lecture, que l'enfant va pouvoir créer une représentation orthographique correcte. Ainsi, c'est l'activité de lecture qui va en retour permettre le développement de l'activité d'écriture. De plus, cette correspondance avec l'auto-apprentissage semble d'autant plus marquée grâce à l'évolution de l'inégalité des poids entre lecture et écriture : en effet, si cette inégalité semble très marquée en début d'apprentissage, elle s'atténue avec le temps. Selon l'hypothèse d'auto-apprentissage, l'effet de l'exposition diminue avec l'acquisition de l'expertise en lecture ; en cela, notre étude semble conforter cette hypothèse.

Le postulat des niveaux de langue proposé par Abbott et collaborateurs (2010) peut être une explication concernant les différences de résultats de l'ensemble des études présentées au Chapitre 4. C'est également ce qu'avaient conclu Ahmed et collaborateurs (2014), puisqu'ils trouvaient un lien différent selon le niveau de langue étudié (unilatéral de la lecture vers l'écriture pour les niveaux mots et texte, bilatéral pour le niveau phrase). Par conséquent, il serait intéressant de comparer ces études et les résultats mis en avant en fonction du niveau de langue concerné. Ainsi, pour notre étude, en étudiant le lien entre trois habiletés, dissociant parallèlement la lecture en contexte de la lecture isolée, ceci nous permet de comparer deux niveaux de langue différents : au niveau du mot, pour la lecture isolée, et au niveau du texte, pour la lecture en contexte. L'habileté orthographique étudiée se situait quant à elle au niveau du texte.

Cette première étude avait également pour but de vérifier la distinguabilité de deux habiletés liées à la lecture : la lecture de mots isolés et la fluence de texte, cette dernière correspondant à de la lecture en contexte. Si les deux habiletés ont montré leur distinctivité dans les résultats de cette étude, il est intéressant de constater que leur lien évolue de manière opposée dans le temps. En effet, la lecture de mots isolés, qui correspond au décodage, habileté nécessaire au bon développement de la compréhension en lecture (selon le modèle simple de la lecture, Gough & Tunmer, 1986) perd en importance au fil du temps, que ce soit pour sa propre évolution (caractère autorégressif) ou pour son importance dans l'explication de la fluence en contexte. A l'inverse cependant, la fluence en contexte, qui correspond à une habileté mobilisant des compétences cognitives de plus haut niveau, croît avec le temps, tant d'un point de vue autorégressif que par son importance dans l'explication des compétences ultérieures en

lecture de mots isolés. Ainsi, en plus d'une distinction entre ces deux habiletés de lecture, nous constatons également une inversion, au fur et à mesure de la scolarité, de leur influence, tant sur leur propre développement que sur leur développement réciproque. La distinction entre décodage et fluence en contexte avait été vérifiée par Goigoux et collaborateurs (2016), dont l'étude avait mis en évidence l'apport significatif de variance expliquée dans les performances en lecture CP de la fluence en contexte, même après avoir pris en compte les deux habiletés principales du modèle simple de la lecture (décodage et compréhension orale). L'importance ultérieure dans l'apprentissage de la lecture de la fluence en contexte a également été validée par l'étude de Tilstra et collaborateurs (2009), en fin de primaire et collège, notamment dans le poids explicatif des performances en compréhension écrite. Cette distinction entre lecture de mots isolés et fluence en contexte est également une piste potentiellement explicative de l'absence d'effet de la fluence trouvée par Adlof et collaborateurs (2006), qui avaient utilisé comme mesure de fluence une épreuve de lecture de mots isolés.

L'objectif de nos deux autres études était de vérifier le lien existant entre orthographe lexicale et sublexicale en début d'apprentissage : nous souhaitons en effet étudier leur distinguabilité, ainsi que leurs particularités. L'hypothèse d'indépendance de ces deux habiletés orthographiques est bien validée, confortant ainsi le postulat de Binamé et Poncelet (2016) qui distinguent ces deux habiletés dès le début de l'apprentissage. Au-delà de leur distinction dans la compétence orthographique, c'est également au niveau des facteurs externes que ces deux habiletés se différencient : en effet, l'orthographe sublexicale, c'est-à-dire, la capacité à repérer les fréquences statistiques de co-occurrences de lettres, n'est pas dépendante d'un apprentissage formel en classe, contrairement à l'orthographe lexicale. Plusieurs études ont toutefois testé l'apprentissage explicite des règles implicites sous-tendant l'orthographe sublexicale (voir par exemple Stanké et al., 2021, pour un entraînement aux régularités graphotactiques chez des élèves faibles orthographes) ; les entraînements ont montré un effet positif sur la compétence sublexicale en post-test. Si l'orthographe sublexicale n'est donc pas classiquement impactée par l'enseignement orthographique fait en classe, elle pourrait toutefois donc bénéficier d'un enseignement explicite, d'autant plus que l'orthographe sublexicale est impliquée dans les performances en production sous la dictée.

Enfin, de manière plus globale, les trois études présentées dans ce travail nous permettent de nous interroger sur le type de tâche choisie dans la construction d'une expérimentation. En effet, si nous prenons comme exemple la fluence ou l'orthographe lexicale, nos études ont permis de mettre en avant la différence obtenue dans les résultats en fonction du type de tâche choisi pour mesurer une même habileté. Concernant l'orthographe lexicale, les résultats montrent des différences dans les variables explicatives impliquées pour les deux épreuves choisies en CE2 (choix orthographique *versus* dictée). Ainsi, si certains prédicteurs restent communs aux deux épreuves, l'utilisation d'une épreuve de reconnaissance orthographique, par exemple, mobilisera d'autres habiletés (ici, la sensibilité aux régularités graphotactiques) qu'une épreuve de production sous la dictée. Cet effet du type de tâche est cohérent avec les propositions de Chen et collaborateurs (2018), qui, en mesurant l'apprentissage orthographique via deux épreuves (reconnaissance orthographique et production orthographique) ont trouvé des effets différents. Les auteurs postulent ainsi que ces différences pourraient s'expliquer par les capacités cognitives requises pour chacune des tâches. En effet, si une tâche de reconnaissance fait appel, comme attendu, à de la reconnaissance, mécanisme cognitif de plus bas niveau, une tâche de production demanderait davantage de ressources cognitives et attentionnelles.

8.2 Limites

Plusieurs limites sont toutefois à prendre en compte concernant le travail que nous venons de présenter. Tout d'abord, concernant l'échantillon, nous avons fait le choix d'inclure dans les études tous les élèves qui avaient, avec l'accord de leurs tuteurs légaux, accepté de participer à l'étude. Les élèves ayant ensuite été officiellement diagnostiqués comme ayant des troubles des apprentissages dans la grande majorité des cas ont pu choisir d'arrêter leur participation (choix que la plupart des familles ont fait, afin d'éviter une trop grande fatigue ou un potentiel stress face aux épreuves proposées) ; d'autres ont choisi de continuer, en présence de leur AESH pour ceux qui en bénéficiaient. Ces élèves représentaient, au total, un très faible nombre (12 élèves) ; le diagnostic n'étant pas disponible au début de l'étude (la dyslexie, par exemple, devant faire état d'un retard de lecture de 24 mois), ces élèves qui présentaient potentiellement déjà des difficultés de lecture ont été pris en compte dans nos analyses. Les données présentent ainsi des résultats sur un échantillon tout-venant, ce qui peut avoir une influence sur les résultats que nous avons présentés. Toutefois, il pourrait être intéressant

d'étudier les performances avant diagnostic officiel de ce sous-groupe d'élèves, afin d'étudier l'évolution des habiletés liées à la lecture et à l'écriture pour ce sous-groupe. Enfin, l'échantillon étudié n'est pas représentatif de l'ensemble de la population française (voir Chapitre 5). Les résultats présentés ici sont donc potentiellement à nuancer : une analyse similaire sur un échantillon représentatif serait nécessaire pour venir confirmer ou infirmer nos résultats.

Ensuite, une deuxième limite concerne la construction des épreuves utilisées tout au long de cette étude. En effet, le projet ACQUISLONGIT, dans lequel ce travail de thèse s'insère, était un projet de relativement grande ampleur, pour lequel les élèves étaient suivis dès le début du CP et jusqu'à la fin du CM2. Ils n'étaient donc pas encore « lecteurs » lors du premier recueil de données. Ainsi, certaines mesures ont dû évoluer au fur et à mesure que les compétences des élèves s'affinaient ; certaines mesures ne sont donc pas strictement identiques tout au long des temps de mesure. Même si cela s'entend aisément, cela a pu compliquer quelque peu les traitements statistiques ou les choix qui ont été faits lors de la construction des passations. En effet, il était indispensable de prendre en compte les capacités attentionnelles des élèves, leurs compétences sur les habiletés étudiées, le temps global des évaluations et son impact sur l'organisation de la classe pour l'enseignant, et il a bien sûr fallu faire des choix concernant les épreuves proposées aux élèves. Nous avons également dû surveiller le possible effet test-retest, étant donné qu'un grand nombre d'épreuves était identique sur plusieurs temps de mesure ; si cela facilite la comparaison des résultats, il y a toutefois un risque de mémorisation, même partielle, des élèves sur ce contenu. Toutes ces contraintes écologiques rendent ce travail limité notamment dans la continuité des habiletés étudiées ou dans le choix des habiletés étudiées à chaque temps de mesure.

Une troisième limite qui pourrait être faite à ce travail concerne l'amplitude des passations ; en effet, les passations de fin d'année scolaire s'étendaient généralement de la reprise suite aux vacances de printemps à la fin de l'année scolaire. En cette période, l'organisation des passations était ardue : les impératifs des classes (sorties pédagogiques, temps réservé à la pratique sportive ou musicale, etc.), des disponibilités des vacataires, et du respect du protocole (passations collectives en premier, si possible en matinée) ont fait que l'ensemble des passations se déroulait sur environ huit semaines chaque année. Il est donc possible que les élèves ayant été évalués dans les premières semaines n'aient pas encore appris certaines notions, qui ont pu être travaillées en classe pour les élèves ayant été évalués dans les

dernières semaines d'expérimentation. Nous avons, dans la mesure du possible, essayé de réduire cette amplitude au maximum.

8.3 Perspectives

Ce travail vient toutefois soulever d'autres questions, qui mériteraient d'être étudiées plus avant. Tout d'abord, en se focalisant sur les niveaux de langue proposés par Abbott et collaborateurs (2010), une revue fine de la littérature sur les études longitudinales portant sur le lien lecture écriture pourrait permettre une meilleure comparaison des différents résultats trouvés par les équipes de recherche, en comparant par exemple les études qui se concentrent sur les mêmes niveaux de langue, et en analysant les résultats trouvés selon ce critère. Cela permettrait potentiellement de distinguer les effets liés au niveau de langue étudié de ceux liés à la langue en elle-même, et aux différences dans la temporalité d'acquisition des langues en fonction de l'opacité de leur système d'écriture. En effet, dans notre première étude, si un lien bilatéral a été mis en évidence, on constate également une inégalité dans le poids des habiletés de lecture et d'écriture, la lecture pesant davantage dans ce lien bilatéral que l'écriture.

Cette asymétrie dans le poids entre lecture et orthographe peut être mise en lien avec l'opacité de la langue française : en effet, le français a un nombre de graphèmes nettement supérieur au nombre de phonèmes, c'est-à-dire qu'il existe plus de façons d'orthographier un phonème que de façons de prononcer un graphème. Le poids plus fort de la lecture sur l'orthographe en début d'apprentissage, mis en avant par notre première étude, peut donc s'expliquer en partie par cette inégalité dans les correspondances entre graphèmes et phonèmes, comme étant le reflet de cette inégalité de transcription. Afin de vérifier cette possibilité, il serait intéressant de comparer ces résultats à ceux d'études similaires dans des orthographe transparentes : l'objectif serait alors de regarder si cette asymétrie en début d'apprentissage se retrouve également dans les orthographe plus transparentes ou non, notamment les orthographe contenant une quasi égalité dans les correspondances entre graphèmes et phonèmes. En finnois, Leppanen et collaborateurs (2006) ont effectué des modèles *cross-lagged* pour étudier les liens entre lecture et écriture à la fin de l'équivalent de notre CP et le début de l'équivalent de notre CE2. Différents temps de mesures étaient proposés : deux pour la première et la deuxième année de l'étude, un seul en début de dernière année. Les résultats montrent un lien bilatéral entre lecture et écriture, mais les auteurs ne font pas mention de la

différence des poids de chacune des habiletés. Ils mentionnent cependant que ces liens ne se maintiennent pas dans le temps, contrairement aux résultats de notre première étude, pour laquelle des liens étaient mis en évidence à chaque temps de mesure. De plus, les différences du système scolaire finlandais avec le système scolaire français nous imposent de la prudence quant à la comparaison de ces résultats.

Au regard de nos résultats, qui montrent un lien bilatéral mais toutefois très inégal dans le poids de la lecture vers l'écriture, nous pourrions préconiser une approche pédagogique centrée sur la lecture en début d'apprentissage, le poids de la lecture vers l'écriture étant nettement plus important. Toutefois, il serait prudent de s'interroger sur ce que viennent montrer ces résultats : sont-ils le reflet de l'apprentissage réel des élèves, ou sont-ils le reflet de l'enseignement de la lecture et de l'écriture telles qu'actuellement dispensées ? En d'autres termes, ces résultats sont-ils la cause ou la conséquence des choix pédagogiques qui ont été mis en place ? Une étude comprenant un groupe contrôle et des groupes expérimentaux pourrait alors être une piste intéressante afin de répondre à ces questions.

D'un point de vue pédagogique, une étude approfondie du temps passé sur des activités d'enseignement explicite de lecture et d'orthographe permettrait alors de mieux comprendre ce lien inégal entre lecture et écriture : dans notre étude longitudinale, nous n'avons effectivement pas pris en compte cette donnée, qui pourrait pourtant apporter des informations importantes à ce sujet. En effet, des études ont montré l'importance du temps passé à enseigner ces apprentissages : Graham et collaborateurs (2018), ont par exemple proposé une méta-analyse sur l'efficacité des programmes scolaires selon le poids donné dans l'enseignement à la lecture et à l'écriture ; leurs résultats ont montré que les programmes dans lesquels lecture et écriture étaient réparties de manière équilibrée permettaient de renforcer les apprentissages. Ils ont également montré que l'enseignement conjoint de ces habiletés permettait des progrès en lecture comme en écriture, comparativement à d'autres types de programmes scolaires.

Un autre point à explorer concernerait la compétence orthographique : une étude longitudinale avec davantage d'habiletés potentiellement explicatives de cette compétence orthographique serait intéressant à mettre en place, notamment pour avoir plus d'informations sur les différentes habiletés sous-tendant l'apprentissage orthographique tôt dans la scolarité. Il s'agirait, par exemple, d'étudier l'effet précoce de la morphologie ou de la grammaire, ou encore des orthographes approchées, afin d'évaluer leur mise en place et leurs effets sur le

développement de la compétence orthographique dès le début de la scolarité et jusqu'à la fin de l'école élémentaire. Des entraînements sur le développement de cette compétence pourraient également être utiles, tant sur des groupes tout-venant que sur des groupes d'élèves identifiés comme étant potentiellement à risque, afin de voir si cela peut permettre de renforcer les acquisitions.

En lien avec cet autre point se pose la question de l'évaluation de la sensibilité aux régularités graphotactiques : comme nous l'avons vu dans les études 2 et 3, les habiletés habituellement prédictives de la lecture et de l'écriture n'expliquent que très peu cette compétence qu'est l'orthographe sublexicale. De manière générale, la sensibilité à la co-occurrence de patterns, ou apprentissage statistique, est une compétence générale, qui n'est donc pas spécifique à l'écriture. Par exemple, Arciuli et Simpson (2011) ont testé, en anglais, le lien entre l'apprentissage statistique et le langage, chez des enfants et des adultes. En utilisant un paradigme d'apprentissage de triplets visuels (des monstres de différentes formes et couleurs), les auteurs ont montré que l'apprentissage statistique avait une influence sur les habiletés liées à la lecture, chez l'enfant comme chez l'adulte, indépendamment de l'âge et du niveau de classe pour les enfants. Les participants ayant un apprentissage statistique élevé montraient des habiletés de lecture également plus élevées. Cet effet, ici visible sur un stimulus visuel, se retrouve également dans nos études qui montrent que la sensibilité aux régularités statistiques a un effet significatif sur une tâche de reconnaissance orthographique. Cependant, cette sensibilité semble très peu expliquée par les composants classiques de la lecture et de l'écriture. En effet, si nos études permettent de confirmer ce lien entre apprentissage implicite lié aux régularités graphotactiques et habiletés de lecture et d'écriture, elles ne nous permettent pas de comprendre quels sont les facteurs impliqués dans l'évolution de cet apprentissage qui ne dépend pas d'un apprentissage formel. Pour Treiman et Kessler (2022), la méthodologie classiquement utilisée pour évaluer cet apprentissage est à repenser : le type de tâche proposée n'est peut-être pas adapté à cet apprentissage, c'est pourquoi il ne rendrait pas compte des différences inter-individuelles ; de la même manière, les auteurs questionnent le concept même d'apprentissage statistique, qui serait trop vaste pour être appréhendé expérimentalement et donc rendrait plus difficile sa théorisation. La méthodologie utilisée pour évaluer l'apprentissage statistique est également une question soulevée par Witteloostuijn et collaborateurs (2021), qui ont testé cent élèves néerlandophones sur le lien entre apprentissage statistique et compétences en lecture – écriture. Cette étude n'a pas permis de mettre en évidence un lien direct entre apprentissage statistique et habiletés de lecture et d'écriture sur le

public ciblé (des élèves avec ou sans dyslexie), suggérant alors que la méthodologie utilisée dans les différentes études sur le sujet peut être un facteur explicatif des effets ou de l'absence d'effet mis en évidence sur ce lien.

Ainsi, davantage d'études sur cet apprentissage statistique en lien avec l'acquisition de l'orthographe et des prédicteurs autres que des habiletés liées à la lecture ou à l'écriture pourraient nous permettre de mettre en évidence d'autres facteurs explicatifs de cette compétence.

Toutefois, des études ont démontré que, bien que cet apprentissage se fasse de manière non consciente, il peut cependant être travaillé de manière explicite notamment via un entraînement : Stanké et collaborateurs (2021) ont par exemple réalisé un entraînement basé sur l'enseignement explicite de ces régularités graphotactiques ainsi que des régularités morphologiques sur un échantillon d'élèves à besoins éducatifs particuliers de CE2 (troubles du langage oral, dyslexie-dysorthographe, dyslexie isolée). Leur objectif était de vérifier si l'entraînement permettait l'amélioration des performances en dictée des élèves, grâce à l'instruction explicite et le travail en classe des régularités graphotactiques et morphologiques. L'étude a ainsi mis en évidence un effet positif de cet entraînement sur les performances ultérieures en orthographe lexicale des élèves. Davantage de travaux seraient à mener sur ce type d'entraînement, sur des échantillons plus importants, afin de compléter ces pistes d'amélioration de l'orthographe lexicale via un travail métalinguistique sur les règles orthographiques. De tels travaux pourraient en outre être d'autant plus utiles pour les élèves ayant des troubles liés au langage oral ou écrit.

Si le caractère non conscient de l'orthographe sublexicale a été mis en lumière, les études précédentes ont également démontré que cette habileté pouvait se travailler selon le principe de l'enseignement explicite. Il serait alors intéressant de vérifier si, après avoir rendu les élèves conscients de cette habileté et les avoir sensibilisés à sa manipulation, les performances en orthographe sublexicale deviennent dépendantes d'une situation d'enseignement, et dont l'analyse statistique requiert l'utilisation d'un modèle multiniveau, à l'instar de l'orthographe lexicale dans nos études 2 et 3.

De tels travaux sur une pédagogie explicite pour enseigner des compétences non étudiées en classe, comme la sensibilité aux régularités graphotactiques, nous amènent à réfléchir sur les pratiques pédagogiques qui pourraient être également testées. En effet, si nos études ont montré un lien bilatéral dès le début de l'apprentissage en français, avec un poids

plus fort de la lecture vers l'écriture qui s'atténue par la suite, il pourrait être intéressant d'étudier comment cela pourrait se transcrire de manière plus concrète dans les pratiques pédagogiques. Si le poids des habiletés évolue avec le temps, nous pouvons donc supposer que le travail mené sur l'enseignement de ces habiletés en début et en fin de scolarité sera donc appréhendé différemment.

Conclusion

L'objectif de ce travail était d'investiguer les liens sous-tendant les habiletés de lecture et d'écriture ainsi que leur évolution, sur un temps long, en français. Nous avons également proposé un focus sur le lien entre deux habiletés issues de la composante orthographique en début d'apprentissage formel.

Nous avons, dans une première partie, présenté les aspects théoriques liés à ce travail. Dans un premier chapitre, nous avons explicité la conception de l'activité de lecture selon le modèle simple de la lecture (Gough & Tunmer, 1986), puis celui de l'écriture (Juel et al., 1986). Le modèle simple de la lecture postule que la lecture est le fruit de l'interaction entre le décodage et la compréhension orale ; le modèle simple de l'écriture postule quant à lui que l'écriture est le produit de l'épellation et de l'idéation. Ensuite, nous avons présenté les spécificités du système alphabétique français, à savoir le degré de transparence de la langue, la dissymétrie entre les correspondances graphèmes-phonèmes et les correspondances phonèmes-graphèmes, ainsi que les spécificités morphologiques du français (morphologie flexionnelle et dérivationnelle). Dans un deuxième chapitre, nous avons présenté quelques modèles théoriques majeurs dont l'objectif est d'expliquer les processus impliqués dans la lecture experte ou dans le développement de la lecture ; nous avons ensuite présenté les modèles théoriques concernant la compétence orthographique, sa composition, ainsi que les caractéristiques universelles de la lecture. Dans un troisième chapitre, nous avons mis en avant certains facteurs environnementaux qui impactent les apprentissages de lecture et d'écriture, comme l'effet classe, l'enseignement explicite, le travail sur la phonologie, l'écriture sous la dictée, les écritures inventées ou encore l'effet du temps. Enfin, dans un quatrième et dernier chapitre théorique, nous avons présenté les principaux prédicteurs de la lecture et de l'écriture en début de scolarité : décodage, connaissance du nom des lettres, fluence, conscience phonologique, habiletés cognitives générales, dénomination rapide ou encore mémoire de travail. Nous avons, pour finir, comparé plusieurs études longitudinales s'intéressant à l'évolution du lien entre lecture et écriture au niveau élémentaire.

La seconde partie de ce travail était consacrée à l'empirie. Nous avons présenté tout d'abord une première étude, consacrée à l'évolution des habiletés de lecture et d'écriture du CP au CM2. L'originalité de cette étude résidait principalement dans sa durée (cinq années

scolaires, soit tout le cursus de l'école élémentaire française) ainsi que dans le nombre d'élèves impliqués (561 sur l'ensemble de l'étude). Nous avons choisi d'étudier les liens entre lecture de mots isolés, fluence de lecture en contexte, et orthographe lexicale. Le modèle statistique utilisé a permis de mettre en évidence un lien bilatéral entre la lecture et l'orthographe, avec un poids plus important de la lecture vers l'orthographe, notamment en début d'apprentissage. Ceci peut s'expliquer à la fois par le mécanisme d'apprentissage si l'on se base sur l'hypothèse d'auto-apprentissage, postulant que le décodage est la fondation nécessaire à l'assimilation de représentations mentales orthographiques correctes ; ces résultats peuvent également s'expliquer, en français, par l'inconsistance de l'orthographe, à savoir l'inégalité entre les correspondances graphèmes-phonèmes et les correspondances phonèmes-graphèmes. Cette première étude aura également permis de mettre en avant l'indépendance des deux habiletés de lecture, à savoir la lecture de mots isolés et la lecture de mots en contexte (fluence de textes).

Les deux études suivantes étaient consacrées au lien entre orthographe lexicale et sublexicale, au début de la scolarité (respectivement, du CP au CE2 et du CE1 au CE2). Ces études ont permis de mettre en évidence la distinguabilité entre les deux types d'orthographe, qui ont, d'après nos résultats, des prédicteurs communs mais aussi des prédicteurs propres à chacune. Ainsi, les résultats ont mis en évidence une indépendance entre orthographe lexicale et orthographe sublexicale, via des prédicteurs distincts apportant chacun une part de variance explicative dans les performances mesurées en fin de CE2. De plus, le pouvoir explicatif des modèles différait largement selon que l'on cherche à expliquer les prédicteurs de l'orthographe lexicale ou sublexicale.

Enfin, l'ensemble des études réalisées a montré l'importance du type de tâche évalué : que ça soit en lecture ou en orthographe, le choix de la tâche s'avère critique dans les résultats mesurés par les différentes méthodes statistiques employées. En effet, des différences de facteurs explicatifs ou de poids ont été mis en évidence selon les tâches utilisées au sein d'une même habileté.

Ce travail aura donc permis de mettre en évidence un lien bilatéral entre lecture et écriture tout au long de l'école élémentaire ; toutefois, les poids de chacune des habiletés sont inégaux et changent au fil du temps. Ce travail aura également montré la distinction entre fluence de lecture de mots isolés et fluence de texte, ainsi que l'indépendance des habiletés d'orthographe lexicale et sublexicale.

Bibliographie

- Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281-298. <https://doi.org/10.1037/a0019318>
- Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading Include a Fluency Component? *Reading and Writing, 19*(9), 933–958. <https://doi.org/10.1007/s11145-006-9024-z>
- Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of educational psychology, 106*(2), 419. <https://doi.org/10.1037/a0035692>
- Andreu, S., Dalibard, E., & Étève, Y. (2016). CEDRE 2003-2009-2015. Maîtrise de la langue en fin d'école : l'écart se creuse entre filles et garçons. Note d'information de la DEPP, 20, 1-4.
- Ans, B., & Carbonnel, S., & Valdois, S. (1998). A Connectionist Multiple-Trace Memory Model for Polysyllabic Word Reading. *Psychological Review, 105*(4), 678-723. <https://doi.org/10.1037/0033-295X.105.4.678-723>
- Arciuli J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711). <http://dx.doi.org/10.1098/rstb.2016.0058>
- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental science, 14*(3), 464-473. <https://doi.org/10.1111/j.1467-7687.2009.00937.x>
- Bianco, M. (2015). *Du langage oral à la compréhension de l'écrit*. PUG.

- Bianco, M. (2019). Vers un enseignement de la compréhension des textes. *A. Bentolila & B. Germain (Éd.), L'apprentissage de la lecture*, 47-63.
- Bianco, M., Pellenq, C., Lambert, E., Bressoux, P., Lima, L., & Doyen, A. L. (2012). Impact of early code-skill and oral-comprehension training on reading achievement in first grade. *Journal of research in reading*, 35(4), 427-455. <https://doi.org/10.1111/j.1467-9817.2010.01467.x>
- Binamé, F., & Poncelet, M. (2016). The development of the abilities to acquire novel detailed orthographic representations and maintain them in long-term memory. *Journal of experimental child psychology*, 143, 14-33. <https://doi.org/10.1016/j.jecp.2015.10.010>
- Bissonnette, S., Richard, M., Gauthier, C. & Bouchard, C. (2010). Quelles sont les stratégies d'enseignement efficaces favorisant les apprentissages fondamentaux auprès des élèves en difficulté de niveau élémentaire ? résultats d'une méga-analyse. *Revue de recherche appliquée sur l'apprentissage*, 3(1),1-35.
- Bocquillon, M. (2020). *Quel dispositif pour la formation initiale des enseignants ? Pour une observation outillée des gestes professionnels en référence au modèle de l'enseignement explicite* [Thèse de doctorat]. Université de Mons.
- Bosse M-L., & Valdois S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. *Journal of Research in Reading*, 32, 230-253. <https://doi.org/10.1111/j.1467-9817.2008.01387.x>
- Bosse, M.-L., Brissaud, C., & Le Levier, H. (2021). French Pupils' Lexical and Grammatical Spelling from Sixth to Ninth Grade: A Longitudinal Study. *Language and Speech*, 64(1), 224-249. <https://doi.org/10.1177/0023830920935558>

- Bressoux, P. (2012). L'influence des pratiques enseignantes sur les acquisitions scolaires des élèves. *Regards croisés sur l'économie*, (2), 208-217.
<https://doi.org/10.3917/rce.012.0208>
- Bressoux, P. (2022, juin). *L'enseignement explicite : de quoi s'agit-il, pourquoi ça marche et dans quelles conditions ? Synthèse de la recherche et recommandations*. Conseil Scientifique de l'Education Nationale.
- Brissaud, C., Cogis, D., Jaffré, J. P., Pellat, J. C., & Fayol, M. (2011). *Comment enseigner l'orthographe aujourd'hui ?* Hatier.
- Brissaud, C. et Fayol, M., (2018). *Étude de la langue et production d'écrits*. Rapport scientifique, conférence de consensus. *Ecrire et rédiger*, CNESCO.
- Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The Foundations of Spelling Ability: Evidence from a 3-Year Longitudinal Study. *Journal of Memory and Language*, 45(4), 751-774. <https://doi.org/10.1006/jmla.2000.2785>
- Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5-51.
<https://doi.org/10.1177/1529100618772271>
- Chaves, N., Totereau, C., & Bosse, M. L. (2012). Acquérir l'orthographe lexicale : quand savoir lire ne suffit pas. *Approche Neuropsychologique des Apprentissages Chez L'enfant*, 118, 271-279.
- Chen, Y. J. I., Irey, R., & Cunningham, A. E. (2018). Word-level evidence of the role of phonological decoding during orthographic learning: A direct test of the item-based

- assumption. *Scientific Studies of Reading*, 22(6), 517-526.
<https://doi.org/10.1080/10888438.2018.1473403>
- Chevrie-Muller, C. (1997). Langage oral écrit mémoire attention (L2MA). ECPA.
- Chung, S. C., Koh, P., Chen, X., & Deacon, S. H. (2022). Orthographic knowledge: a predictor or an outcome of word reading and spelling in bilingual children?. *Reading and Writing*, 1-23. <https://doi.org/10.1007/s11145-022-10302-5>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204-256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach?. *Journal of experimental child psychology*, 95(1), 56-77.
<http://dx.doi.org/10.1016/j.jecp.2006.03.008>
- Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and writing*, 14(5), 549-568.
<https://doi.org/10.1023/A:1011100226798>
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4), 561-584. <https://doi.org/10.1037/0278-7393.9.4.561>
- Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg? Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition*, 122(1), 110-117. <https://doi.org/10.1016/j.cognition.2011.09.003>
- DEPP-MENJ, (2019). Evaluations 2019 Repères CP, CE1 : premiers résultats – Document de travail 2019-03 – Série études - Novembre 2019.

- Desrochers, A., Kirby, J.R., Thompson, G.L., & Fréchette, S. (2009). Le rôle de la conscience phonologique dans l'apprentissage de la lecture. *Revue du Nouvel-Ontario*, 34, 59-82.
<https://doi.org/10.7202/038720ar>
- Desrochers, A., Carson, R., & Daigle, D. (2012). Une analyse des facteurs de risque dans l'apprentissage de la lecture chez l'enfant. *Enfance en difficulté*, 1, 47-83.
<https://doi.org/10.7202/1012123ar>
- Dunn, L. M., Theriault-Whalen, C. M., & Dunn, L. M. (1993). Echelle de Vocabulaire en Images Peabody (EVIP), adaptation française du Peabody Picture Vocabulary Test-Revised. ATM.
- Écalle, J., Dujardin, É., Labat, H., Thierry, X., & Magnan, A. (2022). Prédicteurs de réussite en lecture dans l'Étude longitudinale française depuis l'enfance (ELFE). *Enfance*, 2(2), 195-216. <https://doi.org/10.3917/enf2.222.0195>
- Ehri, L. C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading*, 9(2), 167-188. https://doi.org/10.1207/s1532799xssr0902_4
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. & Shanahan, T. (2001a). Phonemic Awareness Instruction Helps Children Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. *Reading Research Quarterly*, 36(3), 250-287. <https://doi.org/10.1598/rrq.36.3.2>
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001b). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of educational research*, 71(3), 393-447.
<https://doi.org/10.3102/00346543071003393>
- Eme, E., & Rouet, J. F. (2002). Aspects métacognitifs dans l'apprentissage de la lecture-compréhension. *L'orientation scolaire et professionnelle*, (31/1).
<https://doi.org/10.4000/osp.4871>

- Fayol, M. (2009). Chapitre 8. L'orthographe et son apprentissage. *Réussir à Apprendre*, 127.
<https://doi.org/10.3917/puf.labad.2009.01.0127>
- Fayol, M., & Jaffré, J.P. (1999). Note de synthèse. L'acquisition/apprentissage de l'orthographe. *Revue Française de Pédagogie*, 126(1), 143-170.
<https://doi.org/10.3406/rfp.1999.1101>
- Fayol, M., & Jaffré, J. P. (2008). *Orthographier*. Presses Universitaires de France.
- Fayol, M., & Jaffré, J.-P. (2014). *L'orthographe*. Presses Universitaires de France.
- Fisher, C., & Nadeau, M. (2014). Usage du métalangage et des manipulations syntaxiques au cours de dictées innovantes dans des classes du primaire. *Repères*, (49), 169–191.
<https://doi.org/10.4000/reperes.742>
- Fitzgerald, J., & Shanahan, T. (2000) Reading and Writing Relations and Their Development. *Educational Psychologist*, 35(1), 39-50, https://doi.org/10.1207/S15326985EP3501_5
- Foulin, J.-N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, 18(2), 129–155. <https://doi.org/10.1007/s11145-004-5892-2>
- Foulin, J.-N. & Pacton, S. (2006). La connaissance du nom des lettres : précurseur de l'apprentissage du son des lettres. *Éducation et francophonie*, 34(2), 28–55.
<https://doi.org/10.7202/1079021ar>
- Frith, U. (1985). Beneath the surface of developmental dyslexia. I K. Patterson, M. Coltheart, & J. Marshall (Red.), *Surface dyslexia* (p. 301–330). Erlbaum.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of dyslexia*, 36(1), 67-81.
- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263-279. <https://doi.org/10.1017/S0140525X11001841>
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis.

Scientific Studies of Reading, 5(3), 239-256.

https://doi.org/10.1207/S1532799XSSR0503_3

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology*, 92(2), 377–390.

<https://doi.org/10.1037/0022-0663.92.2.377>

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, 25(2), 321-346. <https://doi.org/10.1007/s11145-010-9271-x>

Goigoux, R. dir. (2016). Lire et écrire. Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages. Lyon : IFÉ et ENS.

Goigoux, R., Cèbe, S., & Pironom, J. (2016). Les facteurs explicatifs des performances en lecture-compréhension à la fin du cours préparatoire. *Revue Française de Pédagogie*, (196), 67–84. <https://doi.org/10.4000/rfp.5076>

Gombert, J. E., Bryant, P., & Warrick, N. (1997). Les analogies dans l'apprentissage de la lecture et de l'orthographe. Dans L. Rieben, M. Fayol, & C.A. Perfetti, *Des orthographes et leur acquisition* (pp.319-334). Delachaux et Niestlé.

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading*, 22(3), 217-240. <https://doi.org/10.1111/1467-9817.00087>

Goswami, U., & Bryant, P. E. (1990). Phonological skills and learning to read. Hillsdale, NJ: Erlbaum. <https://doi.org/10.4324/9781315785110>

- Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 7(1), 6-10.
<https://doi.org/10.1177/074193258600700104>
- Graham, S., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2018). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*, 53(3), 279-304.
- Grigorenko, E. L., Klin, A., & Volkmar, F. (2003). Hyperlexia: disability or superability? *Journal of Child Psychology and Psychiatry*, 44(8), 1079-1091.
<https://doi.org/10.1111/1469-7610.00193>
- Hughes, C.A., Morris, J.R., Therrien, W.J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities, Research and Practice*, 32(3), 140-148. <https://doi.org/10.1111/ldrp.12142>
- INSEE (2011, septembre). Synthèse de territoire : le pays de Grésivaudan. *La Lettre Analyses*, 6.
- INSEE (2019). Recensement de la population, exploitation principale.
- Jaffré, J. P., & Morin, M. F. (2008). Les activités pré-orthographiques : nature, validité et conceptions. *Pratiques. Linguistique, littérature, didactique*, (139-140), 189-208.
<https://doi.org/10.4000/pratiques.1238>
- Joyce, T., & Borgwaldt, S. R. (2011). Typology of writing systems: Special issue introduction. *Written Language & Literacy*, 14(1), 1-11.
<https://doi.org/10.1075/wll.14.1.04joy>
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of Literacy: A Longitudinal Study of Children in First and Second Grade. *Journal of Educational Psychology*, 78(4), 243-255. <https://doi.org/10.1037/0022-0663.78.4.243>

- Kandel, S. (in press). The APOMI model of word writing: Anticipatory Processing of Orthographic and Motor Information. In *Cognitive Processes of Language Production*, K. Strijkers & R. Hartsuiker (Eds.), Current issues in the Psychology of Language. Routledge, Taylor & Francis Group.
- Le Brun, I., Bosse, M.-L. & Valdois, S. (2016). Améliorer l'orthographe chez les élèves de 10 à 13 ans : entraînement par dictées guidées. *Nouveaux cahiers de la recherche en éducation*, 19(3), 55–71. <https://doi.org/10.7202/1045177ar>
- Lecocq, P. (2011). Epreuve de Compréhension Syntaxico-Sémantique (E.CO.S.SE). Deuxième édition. Presse Universitaires du Septentrion.
- Leppanen, U., Nieme, P., Aunola, K., & Nurmi ; J.-E. (2006). Development of Reading and Spelling Finnish From Preschool to Grade 1 and Grade 2, *Scientific Studies of Reading*, 10(1), 3-30, https://doi.org/10.1207/s1532799xssr1001_2
- Li , X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, 1-12. <https://doi.org/10.1038/ s44159-022-00022-6>
- Manesse, D., & Cogis, D. (2007). *Orthographe : à qui la faute ?*. ESF éditeur.
- McBride-Chang, C. (1998) The Development of Invented Spelling, *Early Education and Development*, 9(2), 147-160, https://doi.org/10.1207/s15566935eed0902_3
- Ministère de l'Éducation Nationale (2019). Pour enseigner la lecture et l'écriture au CP. Récupéré le 30/11/2021 sur https://cache.media.eduscol.education.fr/file/Actualites/23/2/Lecture_écriture_version WEB_939232.pdf
- MEN-DEPP (2017). PIRLS 2016 : évaluation internationale des élèves de CM1 en compréhension de l'écrit. *Note d'information*, 17(24).

- MENJ-DEPP (2019). Evaluations repères 2018 de début de CP : premiers résultats. *Note d'information*, 19(13).
- Morin, M.-F. & Montésinos-Gelet, I. (2007). Effet d'un programme d'orthographe approchées en maternelle sur les performances ultérieures en lecture et en écriture d'élèves à risque. *Revue des sciences de l'éducation*, 33(3), 663–683.
<https://doi.org/10.7202/018963ar>
- Mussar, R., Sénéchal, M., & Rey, V. (2020). The Development of Morphological Knowledge and Spelling in French. *Frontiers in Psychology*, 11, 146.
<https://doi.org/10.3389/fpsyg.2020.00146>
- Núñez-Regueiro, F., Juhel, J., Bressoux, P., & Nurra, C. (2022). Identifying Reciprocities in School Motivation Research: A Review of Issues and Solutions Associated With Cross-Lagged Effects Models. *Journal of Educational Psychology*, 114(5), 945.
<https://doi.org/10.1037/edu0000700>. hal-03437199v2
- Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child development*, 76(2), 324-339.
https://doi.org/10.1111/j.1467-8624.2005.00848_a.x
- Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings? *Frontiers in Psychology*, 4, 701.
<https://doi.org/10.3389/fpsyg.2013.00701>
- Peereman, R., & Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, 31(2), 376-379. <https://doi.org/10.3758/BF03207735>

- Peereman, R., & Sprenger-Charolles, L. (2018). Manulex-MorphO, une base de données sur l'orthographe du français intégrant les morpho-phonogrammes. *Langue française*, (3), 99-109. <https://doi.org/10.3917/lf.199.0099>
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Lawrence Erlbaum Associates.
- Perfetti, C. A. (2007). Reading Ability: Lexical Quality to Comprehension, *Scientific Studies of Reading*, 11(4), 357-383, <https://doi.org/10.1080/10888430701530730>
- Perfetti, C. A., Rieben, L., & Fayol, M. (Eds.). (1997). *Learning to spell: Research, theory, and practice across languages*. Routledge.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86. <https://doi.org/10.1075/swll.11.14per>
- Perfetti, C. A., & Harris, L. N. (2013). Universal Reading Processes Are Modulated by Language and Writing System. *Language Learning and Development*, 9(4), 296-316. <https://doi.org/10.1080/15475441.2013.813828>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. *Journal of Memory and Language*, 72, 98-115. <https://doi.org/10.1016/j.jml.2014.01.003>
- Pollo, T. C., Kessler, B., & Treiman, R. (2009). Statistical patterns in children's early writing. *Journal of Experimental Child Psychology*, 104(4), 410–426. <https://doi.org/10.1016/j.jecp.2009.07.003>
- Pulido, L., & Morin, M.-F. (2016). L'accompagnement des premières écritures: effets et pratiques – une synthèse. In M.-F. Morin, D. Alamargot & C. Gonçalves (Ed), *Perspectives actuelles sur l'apprentissage de la lecture et de l'écriture/Contributions*

about learning to read and write – Editions de l'Université de Sherbrooke,

<https://doi.org/10.17118/11143/10229>

- Schaars, M. M. H., Segers, E., & Verhoeven, L. (2017). Predicting the integrated development of word reading and spelling in the early primary grades. *Learning and Individual Differences, 59*, 127-140. <https://doi.org/10.1016/j.lindif.2017.09.006>
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review, 96*(4), 523.
- Sénéchal, M., Gingras, M., & L'Heureux, L. (2016). Modeling Spelling Acquisition: The Effect of Orthographic Regularities on Silent-Letter Representations, *Scientific Studies of Reading, 20*(2), 155-162, <https://doi.org/10.1080/10888438.2015.1098650>
- Seymour, P.H.K., (1997). Les fondations du développement orthographique et morphographique. Dans L. Rieben, M. Fayol, & C.A. Perfetti, *Des orthographes et leur acquisition* (pp.385–403). Delachaux et Niestlé.
- Seymour, P. H., Aro, M., Erskine, J. M., & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology, 94*(2), 143-174. <https://doi.org/10.1348/000712603321661859>
- Shanahan, T. (1984). Nature of the reading–writing relation: An exploratory multivariate analysis. *Journal of Educational Psychology, 76*(3), 466.
<https://doi.org/10.1037/0022-0663.76.3.466>
- Shanahan, T., & Lomax, R. G. (1986). An analysis and comparison of theoretical models of the reading–writing relationship. *Journal of Educational Psychology, 78*(2), 116.
<http://dx.doi.org/10.1037/0022-0663.78.2.116>
- Shanahan, T., & Lomax, R. G. (1988). A developmental comparison of three theoretical models of the reading-writing relationship. *Research in the Teaching of English, 22*(2), 196-212.

- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Share, D. L. (1999). Phonological recoding and orthographic learning: a direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95-129. <https://doi.org/10.1006/jecp.1998.2481>
- Snowling, M. (2014). Dyslexia: A language learning impairment. *Journal of the British Academy*, 2. <https://doi.org/10.5871/jba/002.043>
- Sprenger-Charolles, L. (2017). Une progression pédagogique construite à partir de statistiques sur l'orthographe du français (d'après Manulex-Morpho) : pour les lecteurs débutants et atypiques. *Approche Neuropsychologique des Apprentissages chez l'Enfant (ANAE)*, 148(5), 247-256.
- Sprenger-Charolles, L., Colé, P., Piquard-Kipffer, A. & Leloup, G. (2010). EVALEC : Une batterie d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture. Editions ORTHO.
- Sprenger-Charolles, L., & Colé, P. (2013). Quelles sont les méthodes les plus efficaces pour apprendre à lire ? Études avec des enfants ayant appris à lire en langue maternelle seconde. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, 123, 127-134.
- Sprenger-Charolles, L., Desrochers, A., & Gentaz, E. (2018). Apprendre à lire-écrire en français. *Langue française*, (3), 51-67. <https://doi.org/10.3917/lf.199.0051>
- Stanké, B., Rezzonico, S., Moreau, A., Robidoux, J., & Royle, P. (2021). Impact de l'enseignement explicite des régularités graphotactiques et morphologiques sur l'apprentissage de l'orthographe lexicale auprès de différents profils d'élèves faibles

- orthographieurs de CE2. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, 33(173).
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. <https://doi.org/10.1037/0012-1649.38.6.934>
- Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32(4), 383–401. <https://doi.org/10.1111/j.1467-9817.2009.01401.x>
- Treiman, R. (2017). Learning to Spell Words: Findings, Theories, and Issues, *Scientific Studies of Reading*, 21(4), 265-276, <https://doi.org/10.1080/10888438.2017.1296449>
- Treiman, R. (2018). Teaching and learning spelling. *Child Development Perspectives*, 12(4), 235-239. <https://doi.org/10.1111/cdep.12292>
- Treiman, R., & Kessler, B. (2022). Statistical Learning in Word Reading and Spelling across Languages and Writing Systems, *Scientific Studies of Reading*, 26(2), 139-149, <https://doi.org/10.1080/10888438.2021.1920951>
- Verhoeven, L., & Perfetti, C. (2022). Universals in Learning to Read Across Languages and Writing Systems. *Scientific Studies of Reading*, 26(2), 150-164. <https://doi.org/10.1080/10888438.2021.1938575>
- Wechsler, D. (2004). WPPSI-III. Echelle d'Intelligence de Wechsler pour la période Pré-scolaire et Primaire. Troisième édition. ECPA.
- Wechsler, D. (2005). WISC IV. Echelle d'Intelligence de Wechsler pour Enfants. Quatrième édition. ECPA.
- Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2021). The contribution of individual differences in statistical learning to reading and spelling performance in

- children with and without dyslexia. *Dyslexia*, 27(2), 168-186.
<https://doi.org/10.1002/dys.1678>
- Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, Retrieval Processes, and Reading: A Longitudinal Study in Average and Impaired Readers. *Child Development*, 57(4), 988.
<https://doi.org/10.2307/1130373>
- Ziegler, J. C. (2018). Différences inter-linguistiques dans l'apprentissage de la lecture. *Langue française*, 199(3), 35. <https://doi.org/10.3917/lf.199.0035>
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments & Computers*, 28(4), 504-515. <https://doi.org/10.3758/BF03200539>
- Ziegler J. C., Castel C., Pech-Georgel C., & George F. (2008). Lien entre dénomination rapide et lecture chez les enfants dyslexiques. *L'année psychologique*, 108(3), 395-42.
<http://dx.doi.org/10.4074/S0003503308003011>
- Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faisca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, 21(4), 551-559. <https://doi.org/10.1177/0956797610363406>
- Ziegler, J.C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369.
<http://dx.doi.org/10.1098/rstb.2012.0397>

Liste des annexes

Annexe A : RAN chiffres

Annexe B : RAN lettres

Annexe C : Vocabulaire en étendue (EVIP)

Annexe D : Compréhension orale

Annexe E : ECOSSE

Annexe F : Lecture de mots

Annexe G : Lecture de pseudomots

Annexe H : Fluence de lecture (textes)

Annexe I : Sensibilité graphotactique

Annexe J : Reconnaissance orthographique

Annexe K : Compréhension écrite

Annexe L : modèle CLPM complet

Annexes

- **Annexe A** : RAN chiffres

4	1	5	3	2	5	1	4
5	4	3	1	2	1	3	5
1	3	4	2	5	3	4	2
3	5	2	4	1	5	2	1
4	3	1	2	3	2	4	5

- **Annexe B** : RAN lettres

O A U I E U A O
U O I A E A I U
A I O E U I O E
I U E O A U E A
O I A E I E O U

- **Annexe C** : Vocabulaire en étendue (EVIP)

Liste des 32 mots utilisés fin CP (numéro correspondant dans la liste globale de l'EVIP suivi du mot choisi) :

45. Robinet	57. Pot	69. Vase
46. Voile	58. Ecorce	70. Tir à l'arc
47. Narine	60. Balcon	72. Casserole
49. Surpris	61. Hameçon	73. Pédale
50. Groupe	62. Récompenser	74. Colère
51. Remplir	63. Fatigué	75. Tranquillité
52. Peler	64. Cérémonie	89. Classer
53. Dispute	65. Mécanicien	130. Mendiant
54. Plonger	66. Fragile	94. Médaille
55. Livrer	67. Tronc	159. Ligoter
56. Démolir	68. Anneau	

Liste des 45 mots utilisés en CE1 (numéro correspondant dans la liste globale de l'EVIP suivi du mot choisi) :

52. Peler	127. Lucarne	158. Agrume
56. Démolir	94. Médaille	123. Exténuée
58. Écorce	108. Cascade	117. Ornithorynque
61. Hameçon	149. Spectre	135. Spatule
64. Cérémonie	91. Pyramide	136. Filtre
73. Pédale	159. Ligoter	115. Sphérique
76. Cylindre	133. Arrogant	99. Reptile
129. Seringue	146. Bourrasque	141. Affûter
89. Classer	122. Boulon	156. Obélisque
125. Confiance	124. Félin	155. Incandescent
130. Mendiant	110. Quatuor	93. Délabré

132. Assaillir

151. Entravé

152. Lustrer

128. Constellation

134. Péninsule

160. Diverger

113. Parallélogramme

- **Annexe D** : Compréhension orale : textes utilisés pour l'épreuve de compréhension orale en fin de CP et fin de CE1, et questions de compréhension posées à la suite de la lecture orale :

Texte 4 : Le dromadaire (fin CP, CE1)

Le dromadaire est l'ami des gens du désert. Ils boivent son lait, fabriquent des vêtements avec sa laine et surtout, ils montent sur son dos pour voyager. Dans sa bosse, il fait des provisions de nourriture. C'est son frigo à lui ! Quand il part dans le désert, sa bosse est grosse et bien remplie. Mais à la fin du voyage, il l'a toute vidée. Comme l'eau est rare, quand il en trouve, il peut en boire plus de 100 litres à la fois. C'est comme s'il buvait toute l'eau de ta baignoire !

- 1) Pourquoi le dromadaire est-il utile aux gens du désert ? (*Inciter l'enfant à donner 3 raisons, le relancer : te rappelles-tu d'une autre raison ? Et d'une autre encore ?*)
- 2) A quoi sert sa bosse ?
- 3) Comment est la bosse du dromadaire à la fin du voyage dans le désert ?
- 4) Pourquoi ?
- 5- Pourquoi le dromadaire boit-il beaucoup en une seule fois ?

Texte 5 : Le paquet surprise (fin CP)

Samedi, un paquet surprise est arrivé pour Pierre et Julie. Pierre a regardé les tampons sur le paquet. Il n'avait jamais vu des lettres aussi bizarres. Julie a enlevé la ficelle. Puis ils ont poussé des cris de joie. Leur oncle avait envoyé des rollers pour Julie et un train électrique pour Pierre. C'était ce que les enfants désiraient depuis très longtemps.

- 1) Quel jour le paquet est-il arrivé ?
- 2) Comment sait-on que les enfants n'attendaient pas le paquet ?
- 3) Qui a enlevé la ficelle ?
- 4) Comment sait-on que le paquet vient d'un autre pays ?
- 5) Qui a envoyé le paquet ?
- 6) Qu'y avait-il dans le paquet pour Julie ?
- 7) Qu'y avait-il dans le paquet pour Pierre ?
- 8) Pourquoi les enfants étaient-ils heureux de ces cadeaux ?

Texte 6 : Au cirque (fin CP)

Les lions avaient presque terminé leur numéro. Jacques attendait pour nettoyer la piste. Le tonnerre au-dessus de la tente du cirque avait rendu les lions nerveux. Soudain, Tina, la dompteuse des lions, trébucha. Son fouet tomba. Le plus jeune des lions sauta vers elle. Jacques se glissa doucement à l'intérieur de la cage, ramassant le fouet avec une grande habileté. Son action rapide permit à Tina de reprendre très vite le contrôle de la situation. Après cette petite aventure, Jacques savait le métier qu'il voulait faire plus tard.

- 1) Où se passe l'histoire ?
- 2) Les lions en étaient-ils au début, au milieu ou à la fin de leur numéro ?
- 3) Pourquoi Jacques attendait-il ?
- 4) Pourquoi les lions étaient-ils nerveux ?
- 5) Qu'est-il arrivé à Tina ?
- 6) Qu'a fait Jacques quand le jeune lion a sauté ?
- 7) Qui a terminé le numéro ?
- 8) Vers qui le jeune lion a-t-il sauté ?
- 9) A ton avis, quel métier Jacques veut-il faire ?

Texte 7 : Les pingouins (CE1)

La nuit tombait. Beaucoup de gens s'étaient rassemblés pour assister à un étrange spectacle. Tout d'abord, il n'y eut aucun signe de vie, ni sur les dunes, ni sur la plage.

Puis, un léger murmure s'éleva dans la foule. Des projecteurs balayaient la plage pendant qu'une rangée de pingouins chargés de poissons pour leurs petits sortit des vagues en se dandinant. Ils grimpèrent sur les dunes en vacillant jusqu'à leurs nids, le long des sentiers bien tracés. Les petits sortaient leur tête avec impatience et soudain, la magnifique parade prit fin.

- 1) Qu'est-ce que les gens venaient voir ?
- 2) A quel moment de la journée se passe l'histoire ?
- 3) Où les gens se sont-ils rassemblés ?
- 4) Grâce à quoi les spectateurs peuvent-ils voir les pingouins ?
- 5) Pourquoi les pingouins sortent-ils de l'eau ?
- 6) Où les pingouins ont-ils fait leurs nids ?
- 7) Comment sait-on que les pingouins font souvent le même trajet ?

- 8) Comment sait-on que le trajet de la mer jusqu'au nid est fatigant pour les pingouins ?
- 9) Comment sait-on que les petits avaient faim ?

• **Annexe E** : ECOSSE :

Liste des phrases énoncées aux élèves en fin de CP :

- 1) Le garçon pousse la dame.
- 2) Le crayon est moins grand que le stylo.
- 3) Le crayon qui est sur le livre est noir.
- 4) Le crayon noir est sur le stylo.
- 5) La reine qui regarde le roi est debout.
- 6) La dame qui montre la fille est assise.
- 7) La fille est portée par le garçon.
- 8) Le bol gris est sous la tasse.
- 9) La pomme est la moins grande.
- 10) La dame que montre la fille est brune.
- 11) La vache poursuivant le chat est noire.
- 12) Le chien a fait tomber la deuxième quille.
- 13) Le tigre qui regarde le lion est couché.
- 14) Quelqu'un montre le troisième crayon noir.
- 15) La fée qui pousse la reine est blonde.
- 16) Le monsieur regarde la vache que poursuit le chien.
- 17) La fourchette grise est derrière le couteau.
- 18) La moto est suivie par le vélo.
- 19) Le cercle dans lequel il y a une étoile est noir.
- 20) Le crayon n'est ni long ni noir.
- 21) Le carré qui est dans l'étoile est gris.
- 22) La dame est assise sur la dernière chaise blanche.

Liste des phrases énoncées aux élèves en CE1 :

- 1) Le crayon qui est sur le livre est noir.
- 2) Le crayon noir est sur le stylo.
- 3) La dame qui montre la fille est assise.
- 4) La fille est portée par le garçon.
- 5) Le bol gris est sous la tasse.
- 6) La pomme est la moins grande.

- 7) La vache poursuivant le chat est noire.
- 8) Le chien a fait tomber la deuxième quille.
- 9) Le tigre qui regarde le lion est couché.
- 10) Quelqu'un montre le troisième crayon noir.
- 11) La fée qui pousse la reine est blonde.
- 12) Le monsieur regarde la vache que poursuit le chien.
- 13) La fourchette grise est derrière le couteau.
- 14) La moto est suivie par le vélo.
- 15) Le cercle dans lequel il y a une étoile est noir.
- 16) Le crayon n'est ni long ni noir.
- 17) Le carré qui est dans l'étoile est gris.
- 18) La dame est assise sur la dernière chaise blanche.

• **Annexe F** : lecture de mots (1min) : version expérimentateur

(essai : «ta » et «bol »)			Total		
a	ou	la	au	tu	/5
un	il	été	on	mur	/5
ni	sur	qui	vélo	par	/5
feu	ce	peur	ami	moto	/5
peau	lune	car	lire	bon	/5
mardi	col	avril	roi	faire	/5
facile	cheval	vrai	ligne	porte	/5
autre	loup	soir	page	raisin	/5
gare	sucré	chat	matin	trésor	/5
grand	lundi	rose	visage	six	/5
ciseau	août	pays	balai	fille	/5
sept	lourd	femme	garage	hibou	/5
loupe	pétrole	stade	industrie	examen	/5
saladier	orchestre	magie	rhum	quai	/5
bocal	écharpe	domicile	seconde	square	/5
condamné	farce	ennui	hache	bouchon	/5
album	balai	agréable	emmener	fable	/5
vigne	vingtaine	poêle	gentiment	clown	/5
comptoir	piège	balance	club	capitale	/5
poire	écho	décor	désespoir	maximum	/5
gendarme	aquarium	foot	cauchemar	bague	/5
aiguille	coque	crochet	marmite	août	/5
million	fuite	galop	rigueur	tronc	/5
sauvage	respect	approche	dix	caravelle	/5
orchidée	avanie	équateur	vague	aquarelle	/5
courroie	ville	faute	dolmen	asile	/5
fusil	animé	compteur	bottine	cacahuète	/5
montagne	agenda	dorade	tabac	pardon	/5
escroc	astronome	moyen	talon	cake	/5
baril	parfum	mission	chorale	acrobate	/5
Nombre total de mots correctement lus en une minute			/150		

• **Annexe G** : lecture de pseudomots (1min) : version expérimentateur

(essai : « ti » et « buk »)			Total		
o	i	bi	ul	titi	/5
ja	ol	ata	dik	nar	/5
vaf	zou	dul	lévo	tur	/5
veur	co	neul	opa	timo	/5
neau	lupe	cal	tibe	chir	/5
pirda	gu	ablir	sar	vaïpe	/5
cifale	champir	vro	rigne	rosan	/5
aubre	louré	gur	talpe	ragife	/5
noge	torpe	chil	tésin	trosir	/5
bran	ranli	ousir	vigase	bapre	/5
gésin	cibe	krolir	mapé	vige	/5
cipre	gofi	tasin	cipé	puse	/5
poute	bluque	tropule	aximon	prachet	/5
ammoper	blafe	moupe	minarte	taïdinte	/5
uspindrie	toirponte	vagne	cloi	vique	/5
panxiton	flai	rhupé	chacaumar	clon	/5
ommui	quiliatre	dorgente	canséte	rolser	/5
quéreil	dinolice	ambe	naque	dosela	/5
cédor	chur	miège	lapadier	fissato	/5
boïpe	ogréïple	irchustré	rigun	gadelot	/5
tageloi	rac	rigende	caldon	plour	/5
sande	chon	givor	bondeuse	girettos	/5
vatriche	pisal	bertale	aivron	pacirande	/5
taubage	mic	mardiron	fudin	robisse	/5
anchovée	agante	courlone	stipe	torac	/5
trane	splindon	tagin	modan	tandir	/5
casine	bate	coginte	abranise	glon	/5
taparelle	abindeur	gental	ontage	clautaille	/5
guébe	diste	dargoné	esan	mejasse	/5
ochirpe	sule	rapsir	lucelis	gavin	/5
Nombre total de mots correctement lus en une minute			/150		

- **Annexe H** : fluence de lecture de textes (1min)

Monsieur Petit (fin CP, CE1)

Monsieur Petit

C'est l'histoire de Monsieur Petit qui vit dans une vieille maison située	12
au cœur d'un vieux village. La maison est entourée d'un jardin avec	24
une barrière ; il y a des concombres, des choux frisés, toutes sortes	36
de légumes. Au fond du jardin, le portillon reste toujours fermé	47
pour que Chien à Puces ne s'échappe pas. Chien à Puces aime se	60
coucher près de la poubelle, à l'ombre d'un oranger couvert de fruits	72
délicieux. Chien à Puces est gourmand, il croque tout ce qui lui passe	85
sous la dent : des oranges pourries qui tombent sur le sol, des fleurs	98
fanées, un morceau de buvard... Un jour, Monsieur Petit décide de	109
mettre Chien à Puces dans une niche. Chien à Puces n'aime pas	121
être enfermé, il préfère s'endormir en regardant les étoiles dans le	132
ciel. Toutes les nuits, il aboie quand Monsieur Petit va se coucher.	144
Monsieur Petit décide de dormir dans le grenier de sa jolie maison	156
pour prendre un peu de repos. Il ne trouve plus le sommeil ! Une nuit	170
d'insomnie, hop ! Il saute du lit et ouvre la grande malle qui se trouve	184
devant lui, dans un coin sombre du grenier. Et là, surprise, toute	196
sa vie, qu'il pensait sans histoire, lui revient en mémoire : il sort les	209
mouchoirs brodés par sa grand-mère, ses petites dents de lait, son	221
pot de chambre ébréché, une tête de poisson séché, un sac plein	233
de billes, une montre qui fait tic, tac, tic, tac, son carnet de notes, un	248
bout de lacet, son vieux transistor à pile.	256

Le voyage de Georges (CE1, CE2)

Dans la rue, derrière le mur sur lequel a grimpé Georges, le petit singe, se tient	16
--	----

un marchand de ballons. Une petite fille achète justement un ballon pour son	29
petit frère. Georges regarde. Lui aussi a envie d'un ballon. Il se penche et	43
essaie de se servir. Mais au lieu d'un ballon, c'est toute la grappe qui se	58
détache. Georges ne veut pas lâcher les fils. Il se sent brutalement soulevé de	72
terre.	73
En un clin d'œil, le vent emporte les ballons et Georges avec eux. Georges se	88
cramponne des deux mains. Et voilà Georges qui s'envole de plus en plus haut.	102
Les immeubles ressemblent de plus en plus à des jouets et les gens à des	117
poupées, puis à des fourmis. Georges a peur. Il ne faut surtout pas lâcher !	131
Alors il s'agrippe aux ballons.	136
Le vent qui soufflait par rafales se calme et tombe. Georges est très fatigué à	151
force de se crisper. Il descend, descend et boum ! Il atterrit sur le sommet d'un	166
feu tricolore ! Vous savez, un de ces feux qui règlent la circulation.	178
Tout le monde est étonné. Les piétons lèvent la tête et bing, dans le réverbère !	193
Les automobilistes lèvent la tête et bang ! dans le pare-chocs ! Tout cela crée	206
un embouteillage et le gendarme qui intervient ne sait plus où donner de la	220
tête.	221

La maitresse a peur du noir (CE2, CM1, CM2)

Cette semaine, les enfants partent en classe verte. Ils vont faire des randonnées, des mathématiques et des dictées en profitant de l'air pur des montagnes. Dans le car, Éline ne quitte pas son doudou. C'est une vieille poupée toute moche, mais qu'elle aime beaucoup. Elle explique que, sans elle, elle ne pourrait jamais s'endormir dans le noir. Les autres enfants commencent à rire, mais la maîtresse se fâche tout rouge. - « Eh bien moi, je parie que la maîtresse aussi a peur du noir ! », chuchote Marine. La première journée passe très vite : on s'installe, on défait les bagages, on dîne ... Et le soir, les enfants se retrouvent tout seuls dans les dortoirs. Dehors, il y a des bruits bizarres. C'est peut-être des loups, des monstres, ou un fantôme ! Et dans le couloir, on entend des pas légers. Le lendemain matin, au petit déjeuner, les enfants s'interrogent. La maison serait-elle hantée ? Lucas est le seul à se taire. Il écoute, un petit sourire aux lèvres. Au bout d'un moment, Lucas lâche : - « Moi, je l'ai vu, le fantôme du

couloir ! C'est une femme. » - « Alors, c'est une dame blanche ! s'écrie Éline. Elle a fait jadis quelque chose de terrible ici, et elle revient hanter les lieux de son crime ! » - « Elle n'était pas du tout blanche, rectifie Lucas. Elle portait une robe rose avec des pois. » À cet instant, la maîtresse s'approche de la table pour dire aux enfants d'aller se brosser les dents avant la promenade en forêt. - « Regardez, elle a de tout petits yeux, on dirait qu'elle n'a pas bien dormi », murmure Éline. - « La pauvre, elle a dû entendre le fantôme, elle aussi. Déjà qu'elle a peur du noir ! », répond Lucas.

Quand le soir arrive, les enfants montent sagement à l'étage, pour rejoindre les chambres. Mais ils sont bien décidés à essayer de surprendre la dame blanche.

Comme la maîtresse a peur du noir, aucun risque qu'elle vienne les gronder parce qu'ils ne dorment pas. Ils attendent longtemps, longtemps. Enfin, à minuit, on entend un grincement sinistre dans le couloir. Les enfants entrouvrent tout doucement la porte des dortoirs. Ils aperçoivent une ombre rose s'engager dans l'escalier. - « C'est elle, suivons-la ! », souffle Lucas. Et quatre petites silhouettes se fauillent pour descendre les marches. Une faible lueur brille dans la grande cuisine. Le coeur battant, les enfants s'approchent à pas de loup et découvrent leur maîtresse. Vêtue d'une drôle de chemise de nuit rose à pois, elle farfouille dans le réfrigérateur en chantonnant. Soudain, elle se retourne, aperçoit ses élèves et s'exclame : - « Ah, vous aussi, vous avez toujours une petite faim, la nuit ? » Et elle poursuit : - « J'ai une terrible envie de glace à la vanille ! » Lucas éclate de rire : - « Vous avez bien raison : rien de tel qu'un peu de glace pour chasser la peur du noir et faire fuir les fantômes ! »

La naissance de bébé phoque (CM1, CM2)

C'est le mois de février. Depuis la fin de l'été, les phoques du Groenland ont fait un long voyage pour aller se reproduire. Les femelles sont près de trois mille groupées sur les glaces flottantes. Bientôt, elles vont mettre bas. Le vent souffle et la mer gèle petit à petit. Ce matin, un bébé phoque est né sur la glace. C'est une grosse boule de poils mouillés. La maman phoque renifle son bébé. Elle reconnaîtra son odeur toute la vie. Le bébé phoque a de grands yeux sombres. Son pelage, d'abord jaune, devient vite tout blanc. Comme cela, il passe inaperçu sur la glace. On appelle le petit un « blanchon ». Saisi par le froid, il grelotte. Maman le caresse du bout du nez avant de le nourrir de son bon lait chaud. Quand la mère part pêcher, le blanchon attend seul dans son berceau de glace. Bien nourri, grâce à un lait dix fois plus riche que celui de la vache, il grossit de plus de deux kilos et demi par jour. Son manteau de graisse mesure près de

dix centimètres d'épaisseur et le protège du froid. Mais il s'impatiente et pleure car si sa mère tarde, il peut mourir gelé. Quand la mère sort enfin la tête de l'eau, le petit s'approche du bord. Pour l'instant le bébé ne sait pas encore nager. D'ailleurs il n'aime pas l'eau et son poil n'est pas très imperméable. Il préfère l'attirer sur la glace pour une dernière tétée car elle va bientôt le quitter. À deux semaines, le petit phoque doit déjà se débrouiller tout seul. Il est abandonné sur la banquise par sa mère. Au début, il l'appelle car il ne peut pas se nourrir seul. Puis, il reste avec ses copains. Son pelage manteau argenté. Un jour, poussé par la faim, il finit par rejoindre l'eau froide. Bientôt, il suivra les adultes vers le nord.

• **Annexe I** : sensibilité graphotactique (CE1, CE2)

1.	attoge	akkoge
2.	irramé	ijjamé
3.	boxxit	bottit
4.	billot	bihhot
5.	ommile	oxxile
6.	tihhas	tinnas
7.	exxiva	effiva
8.	ttamir	tammir
9.	rammin	rramin
10.	ffurois	furrois
11.	niffor	nnifor
12.	tinnot	ttinot
13.	nummar	nnumar
14.	rrafout	raffout
15.	acclomir	acclomir
16.	aplover	applover
17.	ripllave	ripllave
18.	gabiffler	gabiffler
19.	accriver	accriver
20.	mouffrive	mouffrive
21.	afrrunir	afrrunir

• **Annexe J** : reconnaissance orthographique (CE2)

1. □ des lunettes □ des lunêtes □ des lunètes
2. □ une camérat □ une caméras □ une caméra
3. □ ascide □ acide □ asside
4. □ une quète □ une quette □ une quête
5. □ un dépard □ un départ □ un dépar
6. □ un pissenlid □ un pissenlie □ un pissenlit
7. □ un bassin □ un bascin □ un bacin
8. □ un lavabos □ un lavabo □ un lavabot
9. □ contemporein □ contemporain □ contemporain
10. □ un cadran □ un quadran □ un kadran
11. □ il est roux □ il est rout □ il est rou
12. □ un plancher □ un planché □ un planchée
13. □ un nénuphar □ un nénuphard □ un nénuphart
14. □ un artichaut □ un artichau □ un articheau
15. □ debou □ deboux □ debout
16. □ un tuyeau □ un tuyaut □ un tuyau
17. □ une écurie □ une écurit □ une écurid
18. □ la beautet □ la beauter □ la beauté
19. □ un sénario □ un scénario □ un cénario

20. ancien hancien encien
21. un écla un éclat un éclas
22. un épinart un épinar un épinard
23. une acension une assension une ascension
24. un quangourou un kangourou un cangourou
25. le repo le repot le repos
26. un frain un frein un frin
27. un repas un repa un repat
28. de l'acier de l'acié de l'aciet
29. une planête une planète une planette
30. un tablau un tableau un tablaut
31. le duvé le duvet le duver
32. un nit un nid un nie
33. un engar un angar un hangar
34. un lycer un lycée un lycé
35. une cigale une sigale une scigale
36. un cachalot un cachalos un cachalo
37. un dauphin un dauphain un dauphein
38. un hanclos un enclos un anclos
39. un bambout un bambou un bamboux
40. un canapée un canaper un canapé

41. une scerrure une cerrure une serrure

42. Le kartier le cartier le quartier

- **Annexe K** : compréhension écrite de textes

Texte 1 : La boîte (fin CP)

👣 Ma mère m'a donné une grande boîte. Je l'ai posée sur la table. Je cherchais un jouet dans la boîte. C'est un lapin blanc qui est apparu.

Qui donne une boîte à l'enfant ?

Où l'enfant pose-t-il la boîte ?

Que pense-t-il trouver dans la boîte ?

Quelle est la surprise ?

Texte 2 : Le chaton (fin CP)

🐾 Une chatte noire est venue près de ma maison. Elle a apporté un chaton devant la porte. Puis, elle est partie. Maintenant, j'ai un petit chaton à moi.

Qui est venu près de la maison ?

Où la chatte noire a-t-elle déposé le chaton ?

Qui est partie ensuite ?

Qui a un chaton maintenant ?

Texte 3 : L'oiseau (fin CP)

🐦 Un oiseau s'est posé sur ma fenêtre. Je lui ai donné du pain. Il a fait un nid dans mon jardin. Maintenant, je peux veiller ses petits.

Où l'oiseau s'est-il posé ?

Qui a mangé du pain ?

Qu'a fait l'oiseau dans le jardin ?

Qui est né dans le jardin?

Texte 4 : Le pêcheur (fin CP, CE1)

🕒 Anne et Jean cherchaient des têtards. Soudain, ils ont entendu un grand plouf. Un pêcheur était tombé dans le lac. Il ne pouvait pas nager car il s'était blessé. Les enfants ont essayé de le tirer vers le bord mais il était trop lourd. Alors, Anne a tenu la tête de l'homme hors de l'eau pendant que Jean a couru chercher de l'aide.

Que faisaient Jean et Anne près du lac ?

Quel bruit ont-ils entendu ?

Pourquoi Jean et Anne ont-ils entendu un grand plouf ?

Pourquoi le pêcheur ne peut-il pas nager ?

Qu'essayent de faire les enfants ?

Pourquoi sont-ils incapable de tirer le pêcheur vers le bord ?

Comment Anne aide-t-elle le pêcheur ?

Comment Jean l'aide-t-il ?

Texte 5 : Bouddha (CE1) :

Au parc, Simon et Tania ont vu un homme qui courait. Un morceau de papier est tombé de sa poche. Les enfants l'ont ramassé et ont lu un mystérieux message : *Un trésor qui a du nez. Les pieds du bouddha.*

- « C'est sûrement un message codé, dit Simon. Il doit indiquer où a été caché un trésor. » Les enfants, qui ne savaient pas ce qu'était un bouddha, allèrent demander à Marion, la grande soeur de Simon.

- « Bouddha était un homme très bon qui savait beaucoup de choses, dit-elle. Dans la carrière de sable de Fontainebleau, il y a une grande statue de Bouddha. Si vous voulez, je vais vous la montrer. »

Ils prirent tous les trois leur vélo. Arrivés à la carrière, ils se retrouvèrent à l'entrée d'un étrange couloir à ciel ouvert, entre deux falaises. Le couloir était très sombre. Simon et Tania se rapprochèrent l'un de l'autre pour se donner du courage et entrèrent. Soudain, un hurlement jaillit du fond du couloir.

- « On dirait un cri d'animal, dit Marion. Peut-être un renard prisonnier d'un piège ? »

- « Il faut aller voir », affirma Simon, d'une voix qui tremblait un peu.

Ils avancèrent encore un peu jusqu'à se retrouver face à la statue de Bouddha. Le hurlement continuait. Tania se baissa et dans l'obscurité, elle distingua, sous la statue, une cage, puis deux yeux qui la suppliaient et ... le museau d'un petit chien !

-« Oh, le pauvre, dit-elle, il faut le sortir de là ! »

Simon, Tania et Marion attrapèrent les barreaux et tirèrent de toutes leurs forces pour sortir la cage. Puis Tania ouvrit la porte. Le chien bondit avec une force imprévue et courut à toute vitesse.

Les enfants le retrouvèrent près de leurs vélos en train de croquer les biscuits de leur goûter.

- « Il devait garder le trésor », dit Tania.

- « Non, dit Simon. Dans la cachette, il n'y avait qu'une cage. Le trésor, c'est forcément lui ! »

- « Il n'a pas l'air d'avoir une grande valeur, il est tout petit », dit Tania, déçue.

Marion montra les restes du goûter.

- « Je ne sais pas si c'est un chien de pure race mais il est allé tout droit sur mes gâteaux préférés. Quel flair ! Quel nez ! »

Simon dansa d'excitation :

- « Marion a raison, le message disait : *Un trésor qui a du nez* »

1) Quand Simon dit « il doit indiquer où a été caché un trésor », ça veut dire que

...

- Simon doit indiquer où est caché un trésor ^a

- Le message doit indiquer où est caché un trésor ^b

- L'homme doit indiquer où est caché un trésor ^c

- Le bouddha doit indiquer où est caché un trésor ^d

2) Pourquoi Simon et Tania vont-ils voir Marion ?

- Pour lui demander ce que veut dire le message ^a

- Pour lui demander quel est le trésor b

- Pour lui demander ce qu'est un bouddha c

- Pour lui demander de garder le papier d

3) *Pourquoi les enfants vont-ils à la carrière de Fontainebleau ?*

4) *Comment les enfants se rendent-ils à la carrière ?*

5) *Simon a la voix qui tremble un peu parce que ...*

- Il a froid a

- Il a peur b

- Il est fatigué c

- Il est essoufflé d

6) *Dans la carrière, qui hurlait ?*

- Le renard a

- Simon b

- Le chien c

- Bouddha d

7) *Que trouvent les enfants sous la statue ?*

8) *Quel trésor les enfants ont-ils trouvé ?*

9) *Finally, why did the message say « A treasure that has a nose. The feet of the Buddha » ?*

- Le trésor est un chien avec un grand museau ; il est caché sous la statue du Bouddha ^a

- Un chien qui a du flair vous conduira au trésor du Bouddha ^b

- Le nez et les pieds de la statue du Bouddha cachent un trésor ^c

- Le trésor est un chien qui a du flair ; il est caché sous la statue du Bouddha ^d

- **Annexe L** : modèle *cross-lagged* complet

Figure 2. Modèle *cross-lagged* complet sur l'évolution des liens entre orthographe d'usage (USAGE), fluence en contexte (FLUENCE) et lecture de mots isolés (LEC_MOTS) du CE1 au CM2. Les variables prédictrices début CP sont la dénomination rapide alphanumérique (RAN) et les matrices (MATRICES). En vert sont présentés les liens autorégressifs ; en rouge les liens croisés entre les variables au fil du temps ; en violet les covariances.

