



**UNIVERSITÉ
DE GENÈVE**

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Master

2019

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Le lien entre vocabulaire dans la petite enfance et apprentissage de la lecture

Burnand, Laure-Anne Solenne

How to cite

BURNAND, Laure-Anne Solenne. Le lien entre vocabulaire dans la petite enfance et apprentissage de la lecture. Master, 2019.

This publication URL: <https://archive-ouverte.unige.ch/unige:121663>

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.



**UNIVERSITÉ
DE GENÈVE**

**FACULTÉ DE PSYCHOLOGIE
ET DES SCIENCES DE L'ÉDUCATION**

Section de psychologie

MÉMOIRE DE MAITRISE UNIVERSITAIRE EN LOGOPÉDIE

**Le lien entre vocabulaire dans la
petite enfance et apprentissage de
la lecture**

Laure-Anne BURNAND

Sous la direction de

Prof. Pascal Zesiger

Laura Alaria

Jury

Prof. P. Zesiger | Laura Alaria | Céline Lauffs

Juin 2019

19'752 mots

Remerciements

Je souhaite tout d'abord remercier celle qui m'a accompagnée et guidée au cours de ces deux années, Madame Laura Alaria. Elle a été présente pour moi tout au long de ce travail en se rendant très disponible et en sachant m'apporter une aide et un soutien précieux. Je remercie également le Professeur Pascal Zesiger d'avoir été le directeur de cette recherche et pour ses relectures éclairantes. Finalement, merci à Céline Lauffs d'avoir accepté d'être membre du jury.

J'aimerais aussi remercier tous les enfants ainsi que leurs parents d'avoir participé à cette étude, et ce, depuis plusieurs années et sans qui ce travail de master n'aurait pas vu le jour.

Merci bien entendu à Célia, ma partenaire de mémoire, dont le soutien et la collaboration ont été précieux durant la totalité de ce travail.

Un grand merci à tous mes proches et surtout à mes parents pour leur soutien, leurs encouragements et leur présence tout au long de mon cursus. Ils ont su m'encourager dans les bons comme dans les moins bons moments. Merci aussi à Florian d'avoir été à mes côtés durant ces derniers mois de master et dont le soutien et la présence m'ont permis de trouver la motivation nécessaire dans les moments les plus difficiles.

Finalement, j'aimerais aussi remercier toutes mes camarades de logopédie sans qui ces deux années de master n'auraient pas eu la même saveur ! Je vous remercie pour tout ce que nous avons pu partager ensemble, et je me réjouis de vous avoir comme collègues dans ma future pratique !

TABLE DES MATIÈRES

RÉSUMÉ.....	4
INTRODUCTION : pourquoi s'intéresser aux prémices de la lecture ?.....	5
1. CADRE THÉORIQUE.....	6
1.1. LA LECTURE	6
1.2. L'APPROCHE COGNITIVE DE LA LECTURE.....	6
1.2.1. <i>L'identification des mots écrits</i>	7
1.2.2. <i>La compréhension en lecture</i>	8
1.3. LE SYSTÈME D'ÉCRITURE DU FRANÇAIS	9
1.4. LIEN ENTRE LANGAGE ORAL ET LECTURE	9
1.5. LES PRÉDICTEURS DE LA LECTURE AU COURS DU DÉVELOPPEMENT	10
1.5.1. <i>Prédicteurs de l'identification des mots/décodage</i>	11
1.5.2. <i>Prédicteurs de la compréhension en lecture</i>	13
1.6. COMMENT EXPLIQUER LE LIEN ENTRE CES COMPÉTENCES ET LA LECTURE	14
1.6.1. <i>Vocabulaire</i>	14
1.6.2. <i>Dénomination rapide automatisée (DRA)</i>	15
1.6.3. <i>Conscience phonologique</i>	16
1.6.4. <i>Connaissance des lettres</i>	17
1.6.5. <i>Mémoire à court terme phonologique</i>	18
1.7. LIMITES DE LA LITTÉRATURE CONCERNANT CES PRÉDICTEURS DE LA LECTURE.....	18
1.8. PROBLÉMATIQUE ET HYPOTHÈSES THÉORIQUES.....	20
2. PARTIE EXPÉRIMENTALE	22
2.1. LA PRÉSENTE ÉTUDE	22
2.2. MÉTHODE.....	22
2.2.1. <i>Participants</i>	22
2.2.2. <i>Procédure</i>	23
2.2.3. <i>Matériel</i>	23
2.2.3.1. <i>Épreuves et mesures à 22 mois (t1)</i>	24
2.2.3.2. <i>Épreuves et mesures à 60 mois (t2)</i>	24
2.2.3.3. <i>Épreuves et mesures après 7-8 mois d'instruction formelle de la lecture (t3)</i>	26
2.2.4. <i>Plan d'expérience</i>	27
2.2.5. <i>Hypothèses opérationnelles</i>	28

3. RÉSULTATS.....	30
3.1. STATISTIQUES DESCRIPTIVES ET NETTOYAGE DES DONNÉES	30
3.2. ANALYSES DE CORRÉLATIONS	32
3.3. ANALYSES DE RÉGRESSIONS LINÉAIRES MULTIPLES.....	33
3.3.1. <i>Analyse du lien entre les prédicteurs et la lecture</i>	34
3.3.1.1. Les prédicteurs de la lecture et le décodage	34
3.3.1.2. Les prédicteurs de la lecture et la compréhension	35
3.3.2. <i>Analyse du lien entre le vocabulaire avant 2 ans et la lecture</i>	36
3.3.2.1. Lien entre vocabulaire à 22 mois et décodage.....	36
3.3.2.2. Lien entre vocabulaire à 22 mois et compréhension en lecture.....	37
3.3.3. <i>Analyse du lien entre vocabulaire à 22 mois et lecture, médiatisé par les</i> <i>prédicteurs classiques de la lecture</i>	37
3.3.3.1. Lien entre vocabulaire à 22 mois et décodage, médiatisé par les prédicteurs	38
3.3.3.2. Lien entre vocabulaire à 22 mois et compréhension, médiatisé par les prédicteurs	39
4. DISCUSSION	42
4.1. RAPPEL DES HYPOTHÈSES ET DES PRINCIPAUX RÉSULTATS OBTENUS.....	42
4.2. INTERPRÉTATION ET DISCUSSION DES RÉSULTATS	43
4.2.1. <i>Lien entre prédicteurs classiques et la lecture</i>	43
4.2.1.1. Les prédicteurs du décodage.....	43
4.2.1.2. Les prédicteurs de la compréhension.....	45
4.2.2. <i>Liens entre vocabulaire dans la petite enfance et lecture</i>	47
4.3. LIMITES DE LA PRÉSENTE ÉTUDE.....	49
4.4. PERSPECTIVES FUTURES ET OUVERTURE CLINIQUE	50
5. CONCLUSION	52
6. BIBLIOGRAPHIE.....	53
7. ANNEXES	59

Résumé

De nombreux chercheurs se sont déjà intéressés aux liens qui existent entre certaines compétences acquises avant la scolarité et la lecture. Si certains auteurs sont d'accord sur le rôle prédictif de certaines de ces mesures, toutes ne font pas l'unanimité. De plus, la plupart de ces études ont été conduites auprès de populations anglophones et les mesures se font relativement tard dans le développement de l'enfant. C'est pourquoi notre étude a pour premier but d'étudier le lien entre ces prédicteurs classiques et la lecture chez une population francophone mais aussi d'explorer le lien entre la taille du vocabulaire dans la petite enfance et le niveau de lecture au début de son apprentissage formel.

Pour ce faire, nous avons mené une étude longitudinale auprès de 44 enfants domiciliés à Genève et environs. Au cours des différentes années, diverses mesures ont été récoltées. À 22 mois, les participants ont effectué une tâche de désignation d'images novatrice par son design simple et tactile, permettant d'obtenir une mesure de vocabulaire réceptif dans la petite enfance. À 60 mois, ils ont passé diverses épreuves parmi lesquelles une tâche de vocabulaire réceptif, de conscience phonologique, de connaissance des lettres, de mémoire à court terme phonologique ainsi qu'une tâche de DRA automatisée. Finalement, après 8 mois d'apprentissage formel de la lecture, des tâches d'identification de mots et une tâche de compréhension de texte leur ont été administrées.

Les analyses de régression ont permis de montrer que les mesures de connaissance des lettres et de DRA à 60 mois prédisent les futures compétences de décodage alors que la mesure de conscience phonologique au même âge permet de prédire les futures compétences de compréhension en lecture. Ensuite, les résultats ont mis en évidence que le vocabulaire réceptif dans la petite enfance permet de prédire les futures capacités de décodage et de compréhension. Finalement, un modèle global prenant en compte à la fois le vocabulaire à 22 mois ainsi que les trois mesures citées ci-dessus permet d'expliquer plus de variance des compétences de décodage et de compréhension qu'un modèle ne prenant en compte que les prédicteurs classiques à 60 mois. Ces différentes mesures permettent ainsi d'expliquer près de 40% de la compréhension en lecture et près de 50% du décodage. Ces résultats intéressants apportent de nouvelles informations dans le domaine francophone mais restent à interpréter avec prudence compte tenu de la taille de l'échantillon.

Introduction : pourquoi s'intéresser aux prémices de la lecture ?

Un des principaux objectifs de la scolarité primaire est l'apprentissage du langage écrit. Celui-ci va venir supporter la plupart des nouveaux apprentissages scolaires et devient alors un outil important pour communiquer et pour acquérir de nouvelles informations. Sa mise en place est donc cruciale pour tout enfant d'âge scolaire. Après les deux années dévolues à son enseignement, la lecture est considérée comme acquise et comme étant une base sur laquelle l'enfant peut s'appuyer pour acquérir de nouvelles connaissances. Ainsi, les élèves éprouvant des difficultés à ce niveau-là sont fragilisés et pénalisés dans la majorité des disciplines, ce qui peut entraîner des échecs scolaires. Si ces échecs sont vécus à répétition, ils peuvent notamment impacter la construction de l'identité. En effet, Maltais et Henry (1997) ont montré que les élèves avec des difficultés d'apprentissage ont un concept de soi plus faible que les élèves sans difficulté. Il peut alors se mettre en place un cercle vicieux, car selon Bandura (2007), il y a un lien important entre ce qu'un individu pense de lui-même dans un domaine et sa performance dans celle-ci. De ce fait, les enjeux autour de la lecture sont conséquents, pour la scolarité d'une part mais aussi par son impact sur le développement personnel d'un individu (He, Xue, Chen, Chen, Lu & Dong, 2013).

Ainsi, il nous semble important de comprendre quels sont les mécanismes sous-jacents à la mise en place de la lecture afin de pouvoir intervenir de manière préventive avant que cette acquisition soit censée être totalement maîtrisée. Comme nous le verrons dans ce travail, il existe de nombreux liens entre certaines capacités mises en place au tout début de la scolarité et la lecture. Cette étude s'intéresse plus précisément à mettre en évidence quelles sont les différentes capacités qui permettent de prédire les performances ultérieures de lecture au début de son apprentissage alors que cette dernière n'est pas encore mise en place. Cela permettrait, dans le futur, de repérer plus rapidement les enfants à risque de développer des difficultés de lecture et ainsi d'intervenir le plus tôt possible.

Nous commencerons donc par une brève revue de la littérature, focalisée sur l'apprentissage de la lecture ainsi que ses prédicteurs au cours du développement. Nous décrirons ensuite la méthodologie adoptée pour mener cette recherche puis nous présenterons les résultats obtenus à la suite des différentes analyses et nous finirons par discuter leur interprétation. Finalement, certaines limites de l'étude et perspectives cliniques seront exposées.

1. Cadre théorique

1.1. La lecture

Tout d'abord, il nous semble important de définir la lecture avant d'expliquer les mécanismes sous-tendant son acquisition. De manière générale, la lecture est souvent décrite comme étant une tâche complexe pour l'enfant car elle implique diverses composantes qui sont à la fois perceptives, cognitives et langagières (de Weck & Marro, 2010).

Concernant les composantes perceptives, il s'agit essentiellement des processus visuels qui permettent l'extraction d'informations visuelles. Ces processus sont principalement liés aux mouvements de nos yeux lors d'une activité de lecture. En effet, dans son article, Zagar (1992) dégage les différentes caractéristiques oculaires qui sont en jeu. Lorsque nous lisons, les mouvements de nos yeux peuvent être catégorisés en deux moments distincts. Premièrement, il y a ce que nous appelons les fixations, qui correspondent à un moment de pause dans le mouvement des yeux. Deuxièmement, il y a les saccades qui correspondent au moment où l'œil se déplace. Ainsi, notre œil capte l'information visuelle durant les temps de fixation oculaire.

Quant aux processus cognitifs et langagiers, ceux-ci font l'objet de divers modèles théoriques, mais nous ne présentons ici que celui qui semble faire l'unanimité des chercheurs actuellement (de Weck & Marro, 2010). Ce modèle se situe dans une perspective interactionniste où le lecteur doit, pour lire, mettre en lien des informations prises dans le texte avec d'autres connaissances plus globales telles que ses connaissances sur le monde. De ce fait, le lecteur doit apprendre « à organiser l'information écrite en fonction des connaissances précédemment acquises » (de Weck & Marro, 2010, p. 116). Une des manières d'illustrer cette perspective interactionniste est l'approche cognitive de la lecture que nous allons développer ci-dessous.

1.2. L'approche cognitive de la lecture

Comme nous l'avons vu précédemment, la lecture est basée sur des processus perceptifs et cognitifs. De ce fait, il est possible de définir la lecture comme le produit du processus de décodage (ou de reconnaissance des mots) et de compréhension du langage (Gough & Tunmer, 1986). Ces auteurs ont proposé une équation qui permet d'illustrer l'interaction de ces deux processus : $L = R \times C$. Dans cette équation, L représente la performance globale de lecture, R la reconnaissance des mots et C réfère à la capacité linguistique de comprendre et donc de

donner du sens. De ce fait, la lecture implique l'acquisition de connaissances qui sont spécifiques à la lecture (pour l'identification des mots écrits) et d'autres qui lui sont non spécifiques (comme la compréhension). Il s'agit notamment de l'accès à la signification de phrases et de textes ou certaines connaissances générales sur le monde. Ainsi, lorsqu'il est question de la lecture, il s'agit bien à la fois de prendre en compte les compétences de reconnaissance des mots mais également les processus de compréhension car ces deux composantes sont toutes deux nécessaires, mais pas suffisantes individuellement. En effet, le décodage sans la compréhension ne permet pas la lecture, tout comme la compréhension orale avec des processus de décodage insuffisamment élaborés ne mène pas non plus à la lecture.

1.2.1. L'identification des mots écrits

D'après Coltheart (1978), l'identification des mots écrits chez le lecteur expert peut se faire via deux processus parallèles, mais hautement indépendants. Ce principe est décrit par son modèle, dit « à double voie ». La première voie, appelée *voie d'adressage* (ou voie lexicale), permet un accès direct à la représentation orthographique du mot stockée dans le lexique mental. Ce lexique orthographique se construit tout au long de l'apprentissage de la lecture et bien au-delà et il permet l'accès à la connaissance globale du mot. Cette voie permet donc de lire des mots réguliers et irréguliers, pour autant que ceux-ci soient connus et donc stockés par le lecteur dans son lexique. La deuxième voie est appelée *voie d'assemblage* (ou voie non lexicale) et permet, contrairement à l'autre voie, de lire des mots rares, inconnus (c.-à-d. non stockés) ou des pseudo-mots. Cette voie traite séquentiellement chaque élément du mot et convertit les graphèmes en phonèmes à l'aide de règles de conversion (on parle de conversion grapho-phonémique). Finalement, le lecteur expert utilise plus rarement cette voie d'assemblage car elle est plus lente et coûteuse sur le plan cognitif. Ainsi, la voie d'adressage est souvent privilégiée car elle est plus rapide et efficace.

Cependant, les apprentis lecteurs ne peuvent pas utiliser la voie d'adressage dès le départ puisqu'ils n'ont pas encore construit leur lexique orthographique. Certains auteurs proposent donc une acquisition de la lecture qui a lieu en suivant des stades de développement. Frith (1986) propose ainsi trois stades d'acquisition de la lecture, en commençant par le stade logographique. À ce moment-là, l'apprenti lecteur ne sait pas encore lire, mais peut reconnaître certains mots grâce à certains indices visuels. Il reconnaît globalement un mot (voire une forme) mais on ne parle pas encore de lecture. Puis, vient le stade alphabétique où l'enfant peut, à l'aide de ses connaissances de l'alphabet, commencer à décoder des mots en s'aidant de la phonologie.

Il met en place des règles de conversion grapho-phonémique qui lui permettent de déchiffrer la plupart des mots (réguliers et pseudo-mots), mais cela ne lui permet pas de lire les mots irréguliers qui ne respectent pas ces règles. Cela nous amène au dernier stade de son modèle, le stade orthographique. À ce moment, l'enfant n'a plus recours à la phonologie (conversion grapho-phonémique) pour lire les mots puisqu'il reconnaît les mots comme une entité à part entière qu'il récupère dans son lexique. Ainsi, nous pouvons dire que le stade alphabétique correspond à la voie d'assemblage et que le stade orthographique correspond à la voie d'adressage du modèle de Coltheart (1978).

Plus tard, Share (1995) propose un autre modèle d'acquisition du lexique orthographique chez l'enfant. Cette acquisition ne se ferait non pas en stades, mais grâce à l'expérience de lecture de l'enfant. Pour tenter d'expliquer comment les enfants développent ce lexique et acquièrent ainsi l'orthographe spécifique des mots, Share (1995) propose l'hypothèse d'auto-apprentissage. Selon cette hypothèse, c'est le décodage des mots (par la voie d'assemblage) qui va permettre à l'enfant de petit à petit construire son lexique. En effet, lorsqu'il va décoder un mot correctement, l'enfant va progressivement construire la représentation orthographique de celui-ci et le mémoriser.

Pour conclure, nous pouvons voir que, quelle que soit l'approche adoptée, l'enfant semble commencer à lire et à développer son lexique orthographique grâce à la voie d'assemblage. La voie d'adressage prendrait ainsi petit à petit le pas sur la voie phonologique. Une étude menée par Sprenger-Charolles, Siegel, Béchennec & Serniclaes (2003) a d'ailleurs permis de mettre en évidence ce développement avec une utilisation d'abord privilégiée de la voie d'assemblage (phonologique), suivie d'un passage progressif à la voie d'adressage. Ainsi, la voie d'assemblage ou phonologique semble être une étape cruciale dans l'acquisition de la lecture.

1.2.2. La compréhension en lecture

Gough et Tunmer (1986) spécifient bien que lorsqu'ils parlent de compréhension, ils ne font pas référence à la compréhension en lecture, mais bien à ce qu'ils appellent la compréhension linguistique. Il s'agit selon eux « du processus par lequel des informations lexicales (c.-à-d., des mots), des phrases et des discours sont interprétés » [traduction libre] (1986, p. 7). De ce fait, cette composante de la lecture fait appel à des connaissances qui sont non spécifiques au langage écrit, telles que des connaissances sémantiques, syntaxiques, discursives et pragmatiques qui sont déjà mises en place antérieurement dans le langage oral.

1.3. Le système d'écriture du français

Un des facteurs pouvant influencer l'apprentissage de la lecture est la nature du système d'écriture impliqué. Dans les langues alphabétiques, certaines sont considérées comme opaques (l'anglais par exemple), et d'autres comme transparentes (l'italien). Une langue transparente est une langue dont les correspondances graphèmes-phonèmes sont très régulières. Selon Jaffré (1991), ce degré de transparence ou d'opacité aurait plus d'influence sur les apprentis lecteurs que sur les lecteurs experts. D'ailleurs, le français est considéré comme une des langues alphabétiques les plus complexes (après l'anglais). En effet, il existe une trentaine de phonèmes qui peuvent être représentés par une centaine de graphèmes, et ce à l'aide des 26 lettres de l'alphabet et de leurs combinaisons (Jaffré, 1991). De plus, Seymour, Aro et Erskine (2003), ont trouvé qu'à un niveau de connaissance des lettres similaires, les connaissances des enfants en lecture de mots et de pseudo-mots sont meilleures dans les langues qui ont un code orthographique plus transparent. Ainsi, le système de correspondances entre graphèmes et phonèmes du français est complexe car il y a différentes manières de prononcer une lettre ou plusieurs façons d'écrire un son. Néanmoins, le système français est moins irrégulier en lecture qu'en écriture.

1.4. Lien entre langage oral et lecture

Il existe de nombreuses différences entre le langage oral et le langage écrit, que ce soit au niveau de leur acquisition ou de leur enseignement. En effet, le langage oral est une acquisition qui se fait sans apprentissage explicite et qui est le produit d'une interaction entre l'enfant et son entourage. A contrario, le langage écrit est le fruit d'un apprentissage explicite (Bernicot & Bert-Erboul, 2009). Giasson (1997) met en évidence trois principales différences entre langage oral et langage écrit. Ces deux langages diffèrent d'abord par ce que cette auteure appelle leur mode de représentation. Effectivement, le langage oral passe par le canal auditif alors que le langage écrit passe par le canal visuel. Ensuite, elle situe la deuxième différence sur le plan de la structure syntaxique. À l'oral, les phrases sont moins complexes et le langage moins formel. À l'inverse, le langage écrit permet souvent de produire des phrases plus complexes. De plus, certaines tournures de phrases sont plus utilisées à l'écrit qu'à l'oral. Finalement, la dernière différence mentionnée par Giasson est la variation de contexte. En effet, comme le langage oral se situe dans une situation concrète qui se déroule dans l'ici et maintenant, on dit qu'il est concret et contextualisé alors que le langage écrit est lui, décontextualisé.

Malgré ces différences, ces deux modalités s'inscrivent dans une importante continuité (Zesiger, Brun & Patrucco-Nanchen, 2004). Cette continuité s'illustre tout d'abord par le fait que la lecture est « un processus de langage au même titre que le langage oral » (Giasson, 1997, p. 24). En effet, le langage oral constitue une première base nécessaire à la compréhension du monde écrit, les mots oraux étant traduits par les mêmes mots écrits. De plus, il y a des liens étroits entre les aspects de maîtrise de la phonologie et la capacité à identifier les mots écrits. Nous pensons notamment aux capacités que développent les enfants pour associer un phonème à un graphème (conversion grapho-phonémique) ou encore aux capacités de conscience phonologique (cette notion sera définie plus loin). En outre, certains auteurs (Gathercole & Baddeley, 1993) ont mis en évidence un autre lien entre les capacités de mémoire à court terme phonologique (mesurées à l'aide de la répétition de non-mots) et les capacités d'identification de mots. Aussi, comme nous l'avons vu précédemment, la lecture est définie comme le produit entre l'identification des mots et la compréhension (Gough & Tunmer, 1986). Or, il se trouve que la compréhension est un processus global issu du langage oral et qui, d'une certaine manière, en dépend. Une étude de McArthur, Hogben, Edwards, Heath et Mengler (2000) permet d'illustrer ce lien en s'intéressant aux troubles développementaux. Des enfants dyslexiques ont été comparés à des enfants avec troubles du langage oral et il apparaît que 51% des enfants présentant un trouble développemental du langage oral ont aussi des difficultés au niveau du langage écrit. Inversement, 55% des enfants dyslexiques ont aussi des troubles au niveau du langage oral. Ainsi, même si l'un n'est pas forcément la conséquence de l'autre, il semble qu'il y a un lien non négligeable entre les deux.

Ainsi, nous allons explorer plus précisément ce lien en nous intéressant aux différents prédicteurs de la lecture. Ces différents prédicteurs sont identifiables dans le langage oral au cours du développement de l'enfant dès 2 ans et jusqu'au début de l'apprentissage formel de la lecture.

1.5. Les prédicteurs de la lecture au cours du développement

Il a été clairement établi au cours des dernières années qu'il existe des liens très spécifiques entre certaines capacités acquises par l'enfant au cours de son développement langagier et ses futures capacités en lecture. Au moyen d'une revue de littérature riche et variée, nous allons tenter d'expliquer en quoi ces compétences forment ce qu'on appelle des prédicteurs et leurs liens directs avec les futures compétences de lecture.

Comme mentionné précédemment, la lecture est le produit d'un processus de décodage et d'un processus de compréhension. De ce fait, nous présentons séparément les prédictors liés au décodage de ceux liés à la compréhension en lecture.

La plupart des auteurs sont plus ou moins d'accord sur le fait qu'il existe cinq principaux prédictors de l'identification des mots écrits chez les enfants d'âge scolaire. Il s'agit des compétences lexicales, métaphonologiques, de connaissance des lettres, de mémoire à court terme ainsi que des performances à une tâche de dénomination rapide automatisée. Néanmoins, tous ne sont pas forcément d'accord sur ces liens et leur nature.

Il existe moins d'études qui s'intéressent au lien entre ces prédictors et la compréhension en lecture. Jusqu'ici, nous n'avons évoqué qu'un possible lien entre vocabulaire et compréhension en lecture dans le chapitre « La compréhension en lecture ». En effet, la compréhension en lecture demande d'avoir certaines connaissances sémantiques issues du langage oral. Mais comme nous le verrons plus loin, il existe d'autres prédictors qu'il est possible de mettre en lien avec celle-ci.

1.5.1. Prédictors de l'identification des mots/décodage

Une étude longitudinale s'intéressant à la lecture et ses prédictors a été réalisée par Schatschneider, Fletcher, Francis, Carlson et Foorman en 2004. Ces auteurs s'intéressent aux différents prédictors de la lecture. Différentes mesures ont été réalisées sur 945 enfants âgés de 4 ans qui ont été suivis jusqu'à leurs 6-8 ans. Au cours de leur première année de jardin d'enfants, différentes tâches leur ont été administrées : une fois au début de l'année et une fois à la fin de celle-ci. Parmi celles-ci, diverses tâches de conscience phonologique telles que la fusion de syllabes pour former un mot, la fusion de phonèmes pour former un mot, la fusion de phonèmes pour former un non-mot, la comparaison des sons initiaux, l'élision de phonème, la segmentation en phonèmes et l'identification d'un intrus. Les autres tâches sont des mesures de connaissance du nom et du son des lettres, une tâche de dénomination rapide automatisée (DRA) sur objets et une sur les lettres ainsi qu'une tâche de vocabulaire réceptif. Puis, afin de mesurer leurs capacités en lecture, les enfants ont passé une tâche de compréhension de phrases et une tâche de décodage (mesurant la précision et la fluence) à 7-8 ans. Les auteurs ont trouvé que les mesures de conscience phonologique, de DRA (objets et lettres) et de connaissance du son des lettres permettent de mieux prédire les performances de décodage que de compréhension. Un résultat notable concernant la tâche de DRA est que la DRA de lettres permet de prédire une part plus importante de la variance du décodage (précision et fluence)

que la DRA d'objets. Néanmoins, les auteurs n'ont pas trouvé que le vocabulaire permette de prédire les futures compétences de lecture en décodage. D'autres auteurs tels que Hulme et Snowling (2013) sont arrivés aux mêmes résultats. Ils mettent en lien les capacités de conscience phonologique, de connaissance des lettres et la tâche de DRA avec le décodage mais n'établissent pas de liens entre vocabulaire et compétences de décodage.

Plus récemment, une étude longitudinale de Catts, Nielsen, Bridges, Liu & Bontempo (2015) a elle aussi porté sur le rôle prédictif de certaines mesures issues de batteries de screening sur les futures compétences en lecture. Les auteurs sont arrivés aux mêmes résultats que les précédents : les mesures de connaissance des lettres, de conscience phonologique (identification du phonème initial ou reconnaissance de sons dans le mot) et de DRA d'images à 4 ans permettent de prédire les futures compétences de décodage d'enfants âgés de 7 ans. Leur étude apporte un résultat supplémentaire en identifiant la répétition de non-mots comme étant un autre prédicteur. Puis, d'autres auteurs ont trouvé, à la suite d'une méta-analyse, que la conscience phonologique est le meilleur prédicteur des futures capacités de décodage (Melby-Lervag, Halaas Lyster & Hulme, 2012).

Comme nous l'avons vu dans l'étude de Schatschneider et collaborateurs (2004), le vocabulaire ne semble pas être un prédicteur du décodage (voir aussi Muter, Hulme, Snowling & Stevenson, 2004). Cependant, d'autres auteurs ont mis en évidence un rôle prédictif du vocabulaire en âge scolaire (à 5 ans) sur les futures compétences de décodage (voir notamment Scarborough, 2001 ; Catts, Fey, Zhang, & Tomblin, 1999). Pour d'autres, le vocabulaire est important pour certaines compétences de décodage, mais pas toutes. En effet, Ricketts, Nation et Bishop (2007) ont trouvé que le niveau de vocabulaire ne permettait de prédire que la lecture de mots irréguliers, et non pas la lecture de non-mots ni de mots réguliers. De ce fait, le rôle du vocabulaire dans le décodage semble controversé.

Ainsi, juste avant l'apprentissage de la lecture, quatre mesures semblent être de bons prédicteurs des futures compétences de décodage en lecture. Ces mesures sont les suivantes : connaissance des lettres (Hulme, Browyer-Crane, Carroll, Duff & Snowling, 2012 ; Rvachew & Savage, 2006), conscience phonologique (Hulme et al., 2012 ; Rvachew & Savage, 2006 ; Roth, Speece & Cooper, 2002), DRA (Furnes et Samuelsson, 2011) et mémoire à court terme phonologique (Catts et al., 2015). En revanche, le rôle prédictif du vocabulaire sur le décodage semble débattu.

1.5.2. Prédicteurs de la compréhension en lecture

Plus rares sont les auteurs à s'être intéressés spécifiquement à la compréhension en lecture. Or, comme nous l'avons vu précédemment, s'il est effectivement important de pouvoir décoder les mots, il est tout aussi important de pouvoir comprendre les mots et les phrases qui sont lus. Catts, Nielsen, Bridges & Liu (2016) se sont directement intéressés à cette composante de la lecture. Ils ont trouvé que le vocabulaire (autant réceptif qu'expressif), la connaissance des lettres, la conscience phonologique et les performances à une tâche de DRA sont de bons prédicteurs des futures compétences de compréhension en lecture.

Nous réutilisons l'étude de Schatschneider et collaborateurs (2004) décrite précédemment afin d'illustrer quels sont les prédicteurs de la compréhension en lecture. Pour rappel, ces enfants ont été testés une première fois à leurs 4 ans, et une autre fois plus tard après une année scolaire d'apprentissage formel de la lecture. Les auteurs rapportent que la compréhension en lecture peut être prédite par les trois mêmes éléments qui permettent de prédire le décodage, soit la conscience phonologique, la DRA objet et image, ainsi que la connaissance du son des lettres. La part de variance expliquée du niveau de compréhension à 7-8 ans par ces trois mesures est relativement égale et elles permettent d'expliquer entre 20 et 30% de la variance totale. De même que pour le décodage, les auteurs n'ont pas trouvé que le vocabulaire permette de prédire les futures compétences de compréhension en lecture.

Le lien entre vocabulaire et compréhension en lecture est pourtant démontré depuis longtemps par d'autres auteurs (Catts, 1993). Dans cette étude, Catts (1993) a comparé des enfants avec un trouble spécifique du langage oral et des enfants tout-venant âgés en moyenne de 6;2 ans. Son étude a permis de mettre en évidence que ce sont les mesures de langage en production et en compréhension telles que les connaissances sémantiques et syntaxiques qui sont les meilleurs prédicteurs de la compréhension en lecture. D'autres auteurs sont parvenus aux mêmes conclusions, tout en précisant la nature de cette relation. Dans leur étude plus récente, Catts et al. (2016) ont testé 266 enfants au début du jardin d'enfants (à 5 ans) et plus tard au niveau de la lecture, notamment au niveau de la compréhension écrite à 9 ans. Un premier test de screening leur a été administré lorsqu'ils avaient 5 ans, comprenant une tâche de connaissance des lettres, de conscience phonologique, de DRA, de répétition de non-mots ainsi que de vocabulaire (expressif et réceptif). Les chercheurs ont mesuré leurs compétences de décodage vers 7-8 ans, ainsi que leur compétence de compréhension en lecture à l'aide de deux tâches vers leurs 9 ans. Leurs résultats ont notamment permis de mettre en évidence que

le vocabulaire (expressif et réceptif) vers l'âge de 5 ans est le meilleur prédicteur de la compréhension en lecture plus tard (8-9 ans), alors même que l'influence du décodage est contrôlée. D'autres auteurs sont arrivés aux mêmes conclusions (voir notamment Muter, Hulme, Snowling et Stevenson, 2004 ; Ricketts et al., 2007).

Nous pouvons donc relever qu'à ce jour la littérature permet de mettre en évidence quatre prédicteurs de la compréhension en lecture, soit le vocabulaire (expressif et réceptif), la conscience phonologique, la DRA (voir aussi Norton et Wolf, 2012) et la connaissance des lettres. La mémoire à court terme phonologique ne semble ainsi pas prédire la compréhension en lecture car nous n'avons pas trouvé d'études les mettant en lien. Néanmoins, les auteurs ne sont pas unanimes quant aux prédicteurs de la compréhension.

1.6. Comment expliquer le lien entre ces compétences et la lecture

1.6.1. Vocabulaire

Tout d'abord, et avant d'illustrer le lien entre vocabulaire et lecture, il est important de préciser qu'il faut distinguer deux types de vocabulaire. En effet, nous n'avons pas un seul vocabulaire ou stock lexical, mais nous en distinguons deux lorsqu'il est question du langage oral. De ce fait, nous dissocions le vocabulaire réceptif (en compréhension) du vocabulaire expressif (en production). Cette distinction est particulièrement pertinente lorsque nous nous intéressons aux enfants car nous observons la plupart du temps une compréhension meilleure que la production au cours de leur développement. Nous allons donc maintenant tenter d'expliquer en quoi le vocabulaire est relié au décodage, puis à la compréhension en lecture.

Selon Duff, Reen, Plunkett et Nation (2015) le lien entre le vocabulaire et le décodage pourrait s'expliquer soit d'une manière directe, soit d'une manière indirecte. En effet, ce lien semble encore débattu actuellement. Pour certains auteurs (Ouellette, 2006 ; Ricketts et al., 2007), il s'agirait d'un rôle indirect via la phonologie. En effet, l'habileté à encoder des représentations phonologiques détaillées et organisées dans le lexique pourrait expliquer l'association entre le vocabulaire oral et la lecture. Les mots étant encodés phonologiquement dans leur lexique, les enfants peuvent ainsi faire des liens et des associations entre une nouvelle forme orthographique rencontrée et une forme phonologique déjà encodée en mémoire. Ainsi, ces représentations seraient impliquées dans le succès du décodage car comme l'enfant ne possède pas encore de représentations orthographiques du mot cible, il peut se baser sur le stock

phonologique de mots qu'il a déjà encodés. Ceci explique l'importance de la taille du vocabulaire et des représentations phonologiques dans le succès du décodage.

D'autres auteurs soutiennent plutôt l'hypothèse d'un rôle direct du vocabulaire sur le décodage (en lien avec l'hypothèse d'auto-apprentissage de Share, 1995). Leur hypothèse est que lorsque l'enfant tente de décoder un mot et que cela est fait de manière incorrecte, cette tentative pourrait être corrigée s'il possède le mot cible dans son vocabulaire oral (influence top-down du vocabulaire). Évidemment, cette hypothèse fait sens si l'on considère que l'enfant lit à haute voix. Ricketts et collaborateurs (2007) expliquent qu'un enfant avec un bon vocabulaire sera doublement avantageé lorsqu'il lit un mot de manière erronée. Premièrement, l'enfant peut se rendre compte qu'il y a un mot qui est proche phonologiquement de ce qu'il vient de lire. Deuxièmement, il peut aussi réaliser que le mot qu'il vient de lire n'existe pas et donc corriger sa lecture en cherchant un mot phonologiquement proche qui convient.

Ainsi, que nous parlions d'une influence directe ou indirecte, il paraît vraisemblable que ce sont les représentations phonologiques liées au vocabulaire oral qui viennent soutenir le décodage des mots.

Finalement, le lien prédictif entre vocabulaire et compréhension en lecture semble assez évident. En effet, le sens d'un texte ne peut pas être compris si le sens des mots qui le constituent (connaissance du vocabulaire) n'est pas connu (Duff et al., 2015). Ces propos sont d'ailleurs repris par Kamhi et Catts (2012) car selon eux, la taille du vocabulaire a une influence dans le sens où, pour être efficace, l'enfant a besoin de connaître toutes les classes de mots grammaticaux comme les noms, les verbes et les adjectifs. De plus, Ouellette (2006) en est arrivé aux mêmes conclusions en trouvant que c'est surtout la taille du vocabulaire ainsi que sa précision (connaissances sémantiques) qui sont reliées à la compréhension en lecture.

Pour conclure, il semble donc que le vocabulaire soit relié de manière différente aux deux processus de lecture.

1.6.2. Dénomination rapide automatisée (DRA)

Les tâches de DRA sont des tâches qui consistent à dénommer le plus rapidement et le plus justement possible des séries de lettres, de couleurs, d'objets ou de symboles qui se répètent. Ces tâches se déroulent en deux étapes. Premièrement, on demande à la personne d'identifier les différents éléments de l'image afin de s'assurer qu'elle puisse les dénommer

correctement. En effet, on ne cherche pas à savoir si elle connaît ou non ces éléments, mais plutôt à voir à quelle vitesse elle parvient à y accéder. Ainsi, après avoir vérifié que les éléments sont connus et correctement dénommés, on demande à la personne de dénommer le plus rapidement possible ces éléments qui se répètent dans un ordre aléatoire. La lecture et la DRA seraient liées car elles partagent plusieurs processus communs. Pour certains auteurs (notamment Wolf & Bowers, 1999, cités dans Araújo, Reis, Peterson & Faísca, 2015), la lecture ainsi que la DRA impliquent cinq processus cognitifs similaires. Selon eux, toutes deux demandent une attention aux stimuli, un processus visuel qui est responsable de la détection, de la discrimination visuelle et de l'identification du stimulus, une intégration des informations visuelles à des représentations en mémoire, des processus lexicaux incluant l'accès et la récupération de codes phonologiques et finalement, l'organisation d'un output d'articulation. Pour d'autres auteurs (Wimmer, Mayringer & Landerl, 2000, cité par Lervag & Hulme, 2009), la tâche de DRA permet de prédire la lecture car tout comme cette dernière, de bonnes performances sont associées à la vitesse à laquelle les représentations phonologiques peuvent être récupérées en mémoire à long terme.

Enfin, Ziegler, Castel, Pech-Georgel et George (2008) s'intéressent au lien entre DRA et lecture chez une population d'enfants dyslexiques. Grâce à cette étude, ils proposent le point de vue selon lequel DRA et lecture partagent essentiellement un mécanisme qui est « la mise en relation et l'automatisation du couplage entre les représentations visuelles et phonologiques » (p. 415). Ils font ainsi l'hypothèse que des déficits dans des tâches de DRA reflèteraient des difficultés de l'enfant à accéder à son lexique phonologique (voir aussi Treillet, 2004). Cette hypothèse est étayée par certaines études d'imageries cérébrales par résonance magnétique fonctionnelle (IRMf) qui montrent que les mêmes aires cérébrales et circuits neuronaux sont activés chez des lecteurs experts lorsqu'ils réalisent une tâche de DRA ou une tâche de lecture de mots (Misra, Katzir, Wolf et Poldrack, 2004, cités dans Ziegler et al., 2008).

1.6.3. Conscience phonologique

Tout d'abord, il nous semble primordial de définir cette notion. La conscience phonologique est la capacité à percevoir, à découper et à manipuler les unités sonores du langage, telles que le phonème et la syllabe. Il faut donc faire la différence entre ce qu'on pourrait appeler la conscience syllabique et la conscience phonémique car ce n'est pas le même « matériel » oral qui est manipulé. En d'autres termes, c'est la capacité qu'acquiert l'enfant à « réfléchir » sur du matériel phonologique et à le manipuler.

Dans ses études princeps, Liberman (1973, 1974) a mis en évidence que la conscience phonologique est acquise avant la conscience phonémique. Cette dernière ne serait acquise qu'au moment où l'enfant commence l'apprentissage de la lecture car elle serait en lien avec la connaissance des lettres et avec la compréhension du principe alphabétique. La conscience phonologique serait mise en place plus tôt autour de 3-4 ans.

Historiquement, la conscience phonologique est mentionnée dans la littérature comme un prédicteur seulement depuis les années 1970 (Schatschneider et al., 2004). Le lien entre conscience phonologique et identification des mots écrits s'explique notamment au travers du modèle à double voies de Coltheart (1978). Comme nous l'avons vu, l'une des deux voies de lecture est la voie d'assemblage et elle demande à l'enfant de pouvoir identifier les unités de parole (les phonèmes), de les relier aux unités graphémiques et ensuite de les réassembler pour former le mot. En d'autres termes, l'enfant doit « mettre en correspondance les unités sublexicales de l'écrit, les graphèmes, avec les unités correspondantes de l'oral, les phonèmes. Il faut ensuite assembler les unités résultant du transcodage pour accéder aux mots » (Sprenger-Charolles et Colé, 2006, p. 170). Certaines étapes de ce processus demandent donc une manipulation du matériel phonologique afin d'accéder aux sens des mots. De plus, la compréhension du système alphabétique demande de comprendre qu'à une lettre ou à un groupe de lettres (graphème) correspond un phonème et pour le réaliser, il faut être capable d'isoler et de manipuler l'unité phonémique. Néanmoins, certains auteurs (Giasson et Vandecasteele, 2012) nous rendent attentifs au fait que le développement d'une bonne conscience phonologique ne doit pas être l'objectif visé, mais bien un moyen de découvrir le principe alphabétique. Pour d'autres, il s'agit même d'une des conditions de réussite de l'apprentissage de la lecture (Ouzoulias, 2004).

1.6.4. Connaissance des lettres

Certains auteurs se sont intéressés au rôle que pouvait avoir la connaissance des lettres dans l'acquisition de la lecture. Notamment, Foulin (2005), pour le français, considère que la connaissance du nom des lettres est un boost à l'acquisition de la littéracie, et ce de trois manières. Premièrement, cela permet de faire un pont vers l'alphabétisation. Ensuite, cet auteur considère le nom des lettres comme un précurseur du son des lettres. En effet, l'enfant peut s'appuyer sur le nom de la lettre pour en déduire le son car elles sont nombreuses à avoir le son de la lettre dans leur nom (notamment toutes les voyelles du français sauf « y »). Par exemple, la lettre « a » se dit /a/ et son son est aussi /a/ ; la lettre « b » se dit /be/, mais contient le son /b/.

Finalement, connaître le nom des lettres permettrait de donner un premier cadre supportant la conscience phonémique en utilisant par exemple le nom des lettres pour les premières activités d'épellation. De plus, connaître les lettres et leur nom aiderait l'enfant à comprendre et réaliser que les mots utilisés à l'oral sont composés de différents segments, amenant ainsi à les identifier et les manipuler. D'autres auteurs amènent des éléments allant dans le même sens, en spécifiant que le fait d'apprendre le nom des lettres ne demande pas de compétences de conscience phonologique, mais que cela semble néanmoins être un facteur qui facilite le développement de cette compétence (Suortti & Lipponen, 2016). L'apprentissage de la connaissance du nom des différentes lettres commence déjà au cours de la maternelle (dès 3-4 ans) et se poursuit jusqu'au début de l'apprentissage formel de la lecture (Briquet-Duhazé, 2015).

Finalement, même si cette connaissance est un bon prédicteur de la lecture, un entraînement spécifique sur le nom des lettres ne permet pas directement d'augmenter les capacités en lecture (Rvachew & Savage, 2006). Néanmoins, cette compétence serait, selon certains auteurs, le meilleur prédicteur des capacités ultérieures de lecture (identification des mots écrits) (Catts et al., 2015).

1.6.5. Mémoire à court terme phonologique

La manière dont la mémoire à court terme phonologique et le décodage sont reliés semble plutôt indirecte. En effet, comme nous l'avons vu dans le chapitre sur la conscience phonologique, l'enfant doit être capable de maintenir en mémoire certaines informations phonologiques avant de les remettre ensemble pour parvenir à décoder un mot. Il nous semble donc que cette compétence est liée au décodage via la conscience phonologique, et c'est d'ailleurs ce que suggèrent Melby-Lervag et collaborateurs (2012) dans leur méta-analyse. De plus, les deux voies de lecture exposées dans le modèle à double voie demandent au lecteur de maintenir l'information en mémoire durant l'articulation.

1.7. Limites de la littérature concernant ces prédicteurs de la lecture

Comme nous avons pu le voir, les liens entre certaines compétences acquises au début de la scolarité et le développement des futures compétences de lecture sont bien documentés et relativement bien établis dans la littérature.

Cependant, aucune de ces études ne prend en compte des capacités qui se situent avant l'âge de 4 ans. Or, comme nous l'avons vu dans l'introduction, un dépistage ainsi qu'une intervention plus précoce permettraient de limiter l'apparition de certains troubles. De plus, il

existe peu d'études dans ce domaine qui se basent sur une population francophone. En effet, la majeure partie des études mentionnées précédemment concernent des populations anglophones. Pourtant, comme nous l'avons vu plus haut, ces deux langues n'ont pas la même opacité (le français étant plus transparent), c'est pourquoi les résultats ne sont pas forcément transposables d'une langue à l'autre.

Il existe néanmoins quelques études qui s'intéressent à des capacités plus précoces de langage oral chez les enfants et qui les mettent en lien avec les capacités émergentes de lecture. Il y a notamment deux études qui se sont intéressées au pouvoir prédictif du vocabulaire dans la petite enfance (autour de 2 ans) sur les capacités futures de lecture. Cependant, ces études portent elles aussi sur des populations anglophones.

Une des premières grandes études à s'être intéressée au lien entre vocabulaire dans la petite enfance et développement ultérieur de la lecture a été faite par Lee (2011). Dans son étude, cet auteur s'est principalement intéressé à la capacité prédictive du vocabulaire expressif à 2 ans sur le langage oral ainsi que sur les capacités de littéracie de 3 à 11 ans. À l'aide d'un questionnaire parental, des données de vocabulaire expressif de plus de 1'000 enfants au développement typique de 2 ans ont été récoltées. Deux groupes ont été formés, l'un composé des enfants avec un faible niveau de vocabulaire et l'autre des enfants avec un niveau de vocabulaire élevé. Le décodage et la compréhension en lecture ont été testés plus tard, notamment lorsque les enfants avaient 8-9 ans. Les résultats indiquent que le vocabulaire expressif à 2 ans permet de prédire l'identification de lettres, la conscience phonologique, le vocabulaire ultérieur et la compréhension en lecture sur une durée de 9 ans. Plus spécifiquement, c'est la taille totale du vocabulaire (en comparaison à la taille totale des verbes) qui est un meilleur prédicteur des capacités ultérieures de lecture.

Plus récemment, Duff et collaborateurs (2015) se sont intéressés au lien entre le vocabulaire du jeune enfant entre 16 et 24 mois et certaines compétences d'âge scolaire, telles que la conscience phonologique, la précision en lecture, le vocabulaire et la compréhension en lecture chez des enfants âgés de 4 à 9 ans. Dans leur étude, les auteurs ont donc récolté des mesures à deux temps différents auprès de 300 enfants anglais. Premièrement, lorsque les enfants avaient entre 14 et 16 mois, les parents ont rempli un questionnaire afin de savoir lesquels des 416 mots listés leur enfant était capable de comprendre ou de produire, ce qui constituait la mesure de vocabulaire. Dans un deuxième temps, lorsqu'ils avaient entre 4 et 9 ans, les mêmes enfants ont passé des tâches permettant de récolter des informations concernant

leur vocabulaire (réceptif et expressif), leur conscience phonologique (élision de syllabe), leurs capacités de décodage (précision), de compréhension en lecture et de raisonnement non verbal. Les principaux résultats de cette étude montrent que le vocabulaire du jeune enfant entre 16 et 24 mois permet de prédire 4% de la variance de la conscience phonologique, 11% de la précision en lecture, 16% du vocabulaire futur et 18% de la compréhension en lecture lorsque ces enfants sont en âge scolaire. Même si ces résultats sont significatifs, nous pouvons néanmoins voir qu'il reste une grande part de variance non expliquée dans ce modèle. Cependant, la variance expliquée augmente si on ajoute au modèle le risque familial de dyslexie (avec 21% expliquée pour précision lecture et 30% pour la compréhension). Les auteurs arrivent donc à la conclusion que les mesures de vocabulaire vers 2 ans ont une influence causale sur les futures capacités de lecture, mais à un niveau théorique seulement. À un niveau pratique, il ne s'agit pas d'un prédicteur fiable des futures compétences de langage écrit.

Ainsi, même si ces études prennent en compte des mesures plus précoces, elles ne permettent d'expliquer que peu de variances et elles sont basées sur des mesures de vocabulaire qui sont subjectives (questionnaires parentaux). De plus, comme nous l'avons mentionné précédemment, ces études ont été menées sur des populations anglophones et cela pose problème au niveau de la transposition des résultats que l'on peut faire à une population francophone.

Notre étude a donc pour but d'étudier la relation entre le vocabulaire dans la petite enfance et l'acquisition future de la lecture chez une population d'une cinquantaine d'enfants francophones. Ce sujet est encore largement débattu et pourtant peu étudié, et ce d'autant moins sur la base de données objectives. Finalement, nous interpréterons nos résultats afin de voir leurs éventuelles implications cliniques. En effet, l'intérêt clinique recherché est de pouvoir déterminer si des mesures chez le jeune enfant telles que le vocabulaire permettent de prédire ou non de futures compétences de lecture, notamment dans une visée de prévention.

1.8. Problématique et hypothèses théoriques

Pour rappel, nous nous intéressons au développement de la lecture via les relations qu'elle entretient avec ses différents prédicteurs au cours du développement de l'enfant. Sur la base de ce qui a été exposé précédemment dans la littérature, ainsi que sur les diverses limites de celle-ci, nous proposons deux principales hypothèses théoriques.

La première est que le niveau de lecture d'un apprenti lecteur au début de son apprentissage formel devrait être expliqué par les prédicteurs classiques de la lecture mesurés à 5 ans, tels que le vocabulaire, la conscience phonologique, la connaissance des lettres, la mémoire à court terme phonologique et les performances à une tâche de dénomination rapide automatisée.

Puis, dans la continuité des études de Lee (2011) et celle de Duff et collaborateurs (2015), nous supposons que ces mêmes compétences de lecture sont expliquées par des compétences plus précoces de langage oral, et plus précisément par le niveau de vocabulaire avant 2 ans.

2. Partie expérimentale

2.1. La présente étude

Cette étude s'inscrit dans une étude longitudinale plus large initiée par les Professeurs Zesiger, Friend et Poulin-Dubois (Genève, San Diego et Montréal) qui explore les liens entre langage oral précoce et développement des compétences de littéracie. La présente étude s'intéresse plus particulièrement aux liens entre la lecture et ses prédicteurs et plus précisément au lien entre le vocabulaire dans la petite enfance et le développement ultérieur de la lecture. Notre étude s'inscrit donc en quelque sorte dans la continuité des travaux de Lee (2011) et de Duff et al. (2015).

Aux vues des différents éléments et limites mis en lumière dans la littérature, nous nous posons deux questions principales qui guident ce travail. Premièrement, nous cherchons à reproduire ce qui a été déjà avéré dans la littérature scientifique au niveau des prédicteurs de la lecture, mais au sein d'une population francophone. Deuxièmement, pour aller plus loin que les études citées, nous nous baserons exclusivement sur des données objectives d'enfants plus jeunes dans le but d'évaluer si le vocabulaire avant 2 ans permet ou non de prédire les compétences émergentes de lecture, telles que le décodage et la compréhension en lecture au début de son apprentissage formel.

2.2. Méthode

Notre méthodologie a été développée dans le cadre d'une étude longitudinale qui est menée conjointement avec San Diego (USA) et Montréal (Canada). À Genève, 65 enfants ont été vus une première fois au début de l'année 2012 puis ont été suivis à sept temps d'observation différents depuis leurs 16 mois jusqu'à leurs 7 ans. Ils ont été évalués au moyen d'un ensemble de tâches expérimentales incluant des tests de langage oral et de langage écrit pour le dernier temps d'observation au cours du deuxième trimestre de l'année 2018.

2.2.1. Participants

Les enfants ont été recrutés initialement via l'office cantonal de la population de Genève et sur la base du volontariat des parents, selon un critère d'âge (lors de la première passation, ils avaient 16 mois +/- 15 jours) et d'exposition linguistique (exposition au français > 80%). Pour chaque vague sauf la dernière, les enfants ont été vus au même âge à +/- 15 jours près.

Pour cette dernière phase, 56 enfants sur les 65 de la phase initiale ont été testés, ce qui donne un taux d'attrition de 1.6%, ce qui est remarquable pour une étude menée sur 6 années. Pour cette session, ce n'est pas le critère d'âge qui a été privilégié, mais le temps d'exposition à un apprentissage formel de la lecture. Ainsi, tous les enfants ont été vus sur une période allant du 14 mars au 18 avril 2018 après 7 à 8 mois d'exposition à l'apprentissage formel de la lecture.

Pour la présente recherche et compte tenu de nos hypothèses, seules les observations faites à 22 mois (âge moyen : 21.96 mois) et à 60 mois (âge moyen : 59.13 mois) nous intéressent, ainsi que celles réalisées au printemps 2018 par deux membres de l'équipe de recherche, lorsque les enfants avaient 7 ans (âge moyen : 86.86 mois), soit après 7-8 mois d'apprentissage formel de la lecture. De manière générale, les dates auxquelles les enfants ont été vus sont espacées au minimum dans le but de réduire le plus possible la variabilité interindividuelle liée à l'enseignement reçu à l'école.

2.2.2. Procédure

Avant de débiter cette étude, le projet a fait l'objet d'une demande d'autorisation au comité d'éthique. Avant chaque nouvelle passation, les parents ont été contactés à nouveau afin de savoir s'ils étaient d'accord de participer une nouvelle fois. Les enfants accompagnés de leur(s) parent(s) ont chaque fois été accueillis à l'Université de Genève, au laboratoire de psycholinguistique (babylab). Ils y ont été testés individuellement dans une pièce prévue à cet effet et ont été filmés ainsi qu'enregistrés afin de pouvoir analyser les données récoltées. Chaque session a duré environ une heure et ils ont tous passé les épreuves décrites plus loin qui évaluent divers domaines tels que le langage oral, les compétences reliées à la lecture et diverses compétences scolaires. Les participants étaient libres de demander une pause à l'expérimentateur si besoin. Lorsque les enfants étaient plus petits, les parents étaient présents dans la salle. Lors de leur dernière venue, il est arrivé que le parent de l'enfant soit présent si l'enfant le désirait, mais il lui était demandé de ne pas intervenir durant la passation. Finalement, les parents ont dû remplir, à chacune de leur venue, un formulaire de consentement qui résumait l'étude et qui nous autorisait à utiliser les données de leur enfant de manière anonyme.

2.2.3. Matériel

Comme nous l'avons mentionné ci-dessus, cette étude s'inscrit dans une étude longitudinale plus large. Ainsi, nous décrivons ici uniquement les différentes épreuves passées aux différents âges (aux temps t1, t2 et t3) correspondants à nos hypothèses.

2.2.3.1. Épreuves et mesures à 22 mois (t1)

Computerized Comprehension Task (CCT) (Friend&Keplinger, 2003)

Il s'agit d'une tâche de désignation d'images dans laquelle l'enfant doit toucher sur un écran tactile l'image qui correspond au mot cible énoncé oralement. Comme il s'agit d'enfants jeunes, un son renforçateur est produit lorsque l'enfant touche l'image cible (son normalement relié de manière sémantique à l'image), dans le but de conserver leur intérêt. Les items sont présentés par paires, composés de l'item cible et d'un distracteur. Dans l'adaptation française de cette tâche (Schoenhals, Mounir, Friend & Zesiger, 2009), l'enfant dispose de quatre paires d'items d'entraînement, puis viennent 41 paires d'items pour le test. Il existe une version A et une version B de ce test où chaque item joue le rôle de la cible dans une version (A) et le rôle du distracteur dans l'autre version (B). Les différentes paires sont composés de noms (« *Où est le/la ___ ? Touche ____.* »), de verbes (« *Qui est en train de ___ ? Touche ____.* ») et d'adjectifs (« *Lequel est ___ ? Touche ____.* »). De plus, cette tâche est à la fois composée de mots faciles, modérément difficiles et difficiles en proportions égales (critères basés sur les données de production issues de questionnaires parentaux, IFDC). Finalement, une fois les 41 items tests passés, l'enfant repasse directement 13 essais de retest afin de vérifier la fiabilité de ses réponses. La note maximale pouvant être obtenue au test par l'enfant est donc de 41.

2.2.3.2. Épreuves et mesures à 60 mois (t2)

Mesure de conscience phonologique – Adaptée de la WJ Battery (2014)

Afin de mesurer les capacités de conscience phonologique, deux subtests d'une tâche de la batterie de tests de Schrank, McGrew & Mather (WJ Battery, 2014) ont été adaptés en français. Le premier subtest est celui de la sensibilité à la rime, il est demandé aux enfants de nommer des mots rimant avec un son donné. Le score de cette épreuve est le score cumulé de 3 épreuves sur un total de 24 items. Pour le second subtest, il s'agit de la manipulation de phonèmes avec une tâche de déletion de phonèmes composée de 20 items. Ainsi, la note maximale conjointe pour ces deux épreuves est de 44.

The Lollipop Test – mesure de connaissance des lettres (Chew, 1981)

La connaissance des lettres a été mesurée à l'aide d'une tâche issue de la batterie « Lollipop-test » de Chew (1981) qui mesure différentes aptitudes préscolaires. Les enfants ont passé la tâche d'identification de lettres et écriture. On a demandé aux enfants de désigner 5 lettres (B, L, C, P et F) et d'en dénommer 5 (M, E, S, D, H). Puis, il lui était demandé d'écrire 3 lettres (A, B et C), ainsi que d'écrire son prénom (avec un maximum de 5 points). La note

maximale pouvant être obtenue à cette tâche est de 18, correspondant au nombre de bonnes réponses pouvant être obtenues.

Dénomination rapide automatisée (DRA)

La DRA a été testée à l'aide d'une tâche de DRA d'images. Huit noms (4 monosyllabiques et 4 bisyllabiques) et leur dessin en noir et blanc correspondants ont été sélectionnés dans des bases de données lexicales françaises (Alario & Ferrand, 1999 ; Bonin, Peerman, Malardier, Méot, & Chalard, 2003). Les items ont été contrôlés pour la fréquence (pour chaque groupe de 4 mots, deux sont fréquents et deux sont rares), pour l'âge d'acquisition (l'intervalle des stimuli était de 1,12 à 2,16 sur une échelle de cinq points dans laquelle 1 correspond aux mots qui sont appris entre 0 et 3 ans, et 4 aux mots qui sont appris entre 9 et 12 ans), et pour une grande concordance (moyenne = 96,4%). Pour cette dernière mesure, des chiffres obtenus à partir des réponses des enfants lors de la constitution de la base de données nous permettent de faire un dernier contrôle qui assure qu'ils donnent tous un nom équivalent pour chaque image donnée. Ils ont été présentés sur deux feuilles A4 (au format paysage) où chaque image est répétée 4 fois dans un ordre aléatoire (16 stimuli au total/feuille). On demande d'abord à l'enfant de dénommer les 4 images correspondant à la feuille pour s'assurer qu'il les connaisse et donne le bon mot correspondant à chaque image. Puis, il doit dénommer les 16 stimuli présentés sur une feuille et il est chronométré pour chaque feuille. Le score final correspond à un ratio moyen avec la moyenne des scores (nombre de réponses correctes/temps(sec)) obtenus pour la série des mots monosyllabiques et la série des mots bisyllabiques.

Répétition de pseudo-mots – mesure de MCT phono

Pour évaluer la mémoire à court terme phonologique des enfants, ces derniers ont dû répéter des pseudo-mots de 2 à 5 syllabes. Pour chaque « bloc » de pseudo-mots, il y a 4 items de complexité croissante en longueur et en structure pour les pseudo-mots de 2 et 3 syllabes, 7 items pour ceux de 4 syllabes et 5 items pour ceux de 5 syllabes. Ces pseudo-mots ont été créés à partir des structures syllabiques des mots largement produits à cet âge (basé sur les mots de l'IFDC qui s'arrête à 30 mois). Le dispositif est le suivant : on présente à l'enfant une petite figurine de jeu dont le nom correspond au pseudo-mot qu'il doit répéter afin d'induire une motivation supplémentaire à la tâche. Dans la présente version, le nombre d'items correctement répétés constitue le score de l'enfant et le score maximal pouvant être obtenu à cette tâche est de 20.

Échelle de Vocabulaire en Image Peabody (EVIP) (Dunn, Thériault-Whalen & Dunn, 1993)

L'EVIP est une épreuve standardisée qui évalue les habiletés de vocabulaire réceptif chez des enfants âgés de 30 mois à 18;0 ans. Il s'agit d'une épreuve qui se passe sous un format « papier-crayon » et qui a une durée de passation d'environ 8 à 15 minutes. Cette épreuve est une adaptation française du *Peabody Picture Vocabulary Test (PPVT)* (Dunn & Dunn, 2007) qui est un outil d'évaluation du langage réceptif.

Il s'agit d'une tâche de désignation d'image durant laquelle l'examineur formule un mot et l'enfant doit pointer l'image qui correspond selon lui le mieux au mot énoncé oralement. La consigne est la suivante : « *Je vais te dire des mots et tu devras me montrer l'image qui correspond le mieux à chaque mot que je dirai. Tu as bien compris ? On commence.* ».

L'épreuve commence dans tous les cas par quatre items d'essai pour s'assurer de la bonne compréhension de la consigne par l'enfant. Puis, les items étant classés par âge, on commence par les items correspondant à celui de l'enfant afin de ne pas tous les faire passer. La base de départ doit être de huit items consécutifs réussis (on considère les items précédents comme réussis) et le critère d'arrêt correspond à 6 échecs dans une série de huit items consécutifs. Le score brut de l'épreuve se mesure de la manière suivante : on soustrait au nombre total d'items réussis le nombre d'items échoués par l'enfant. L'épreuve se composant de 170 items, le score maximal pouvant être obtenu est de 170.

2.2.3.3. Épreuves et mesures après 7-8 mois d'instruction formelle de la lecture (t3)

BELO – lecture de mots irréguliers

Cette tâche, tirée de la Batterie d'Évaluation de Lecture et d'Orthographe (Pech-Georgel & George, 2006), permet d'évaluer la voie d'adressage de la lecture (voir chap. 2.1. pour une définition). En effet, douze mots irréguliers courts sont proposés à l'enfant et ce dernier doit les lire à haute voix. Le score maximal pouvant être obtenu à cette épreuve est de 12.

Lecture en Une Minute (LUM) de mots familiers et de non-mots (Sprenger-Charolles, Colé, Kipffer-Piquard & Leloup, 2010)

Cette tâche de fluence en lecture se décompose en deux sous-tâches : une lecture de mots familiers et une lecture de non-mots.

Mots familiers : lors de cette tâche, l'enfant dispose d'une minute pour lire un maximum de mots de la manière la plus exacte possible. Il y a au total 60 mots que l'enfant peut lire dans ce

délai d'une minute. Il s'agit de mots familiers, tant réguliers qu'irréguliers (50 réguliers pour 10 irréguliers), dont douze sont composés de graphèmes contextuels (le phonème correspondant dépend du contexte, c'est-à-dire des lettres avoisinantes). L'enfant dispose de deux items d'essai, puis la consigne suivante lui est présentée : « *Maintenant, tu vas travailler seul et tu vas lire le mieux que tu peux tous les mots* ». L'enfant est arrêté au bout d'une minute. Le score maximal à obtenir est donc de 60.

Pseudo-mots : concernant les pseudo-mots, la démarche est exactement la même. L'enfant commence avec deux items d'exemple, puis il y a 60 items à lire, dont 20 avec des graphèmes contextuels et 20 qui sont composés de digraphes (ou, eu, oi, eau, ai, gn, an, au, ou, ch, on, in, un).

Compréhension du texte « les deux poules » (Sprenger-Charolles et al., 2010)

Durant cette tâche on demande à l'enfant de lire le texte à haute voix, puis on lui pose des questions en rapport avec ce texte. Il ne dispose pas du texte sous les yeux pour répondre aux questions et c'est pourquoi il est prévenu au début. Ainsi, la consigne de cette épreuve est la suivante : « *Tu vas lire tout seul, à voix haute, une petite histoire et après je te poserai des questions sur ce que tu as lu. Tu es prêt ? On commence.* »

Pour tester la compréhension en lecture, on pose à l'enfant une série de huit questions. Ces dernières sont composées de quatre questions de compréhension littérale, de deux questions d'inférences et de deux de vocabulaire. La bonne réponse est indiquée dans le questionnaire, ce qui ne laisse ainsi pas de place à la subjectivité. Ainsi une bonne réponse vaut un point et une mauvaise réponse en vaut zéro, le score maximum est de 8 points.

2.2.4. Plan d'expérience

Variables contrôlées

Comme mentionné dans la partie « participants », certains éléments ont été contrôlés, tels que l'âge, le niveau SES (de 9 à 22 années d'étude) et l'exposition linguistique.

Variables indépendantes

Comme nous venons de le voir dans la description des tâches, les variables indépendantes sont relativement nombreuses (pour un résumé, voir Tableau I). À 22 mois, seul le score à l'épreuve de vocabulaire réceptif (CCT) est une variable indépendante. À 60 mois, les variables indépendantes sont le score total obtenu aux tâches de sensibilité à la rime et manipulation de phonèmes, le score obtenu à la tâche de connaissance des lettres (*Lollipop test*),

le score obtenu à la tâche de dénomination rapide automatisée, le score obtenu à la tâche de vocabulaire réceptif (EVIP) et le score obtenu à la tâche de répétition de pseudo-mots.

Variables dépendantes

Les variables dépendantes sont au nombre de deux et correspondent aux deux mesures de lecture effectuées après 7 ou 8 mois d'apprentissage formel (voir Tableau I). La première correspond à un score composite obtenu grâce aux différentes épreuves de décodage. Ces épreuves sont la lecture de mots irréguliers (BELO), ainsi que la lecture de mots familiers et de pseudo-mots (Sprenger-Charolles et al., 2010) et permettent ainsi d'avoir un score global de décodage. La deuxième variable dépendante est obtenue en calculant le score total obtenu à la tâche de compréhension de lecture du texte « les deux poules » (Sprenger-Charolles et al., 2010).

Tableau I

Résumé des différentes variables aux différents âges de mesures

Âge de passation	Variables indépendantes (scores aux différentes tâches de...)						Variables dépendantes (scores aux différentes tâches de...)	
	CCT	EVIP	Lollipop test Mesure de connaissance des lettres	Sensibilité à la rime + Manipulation de phonèmes	DRA image	Répétition de pseudo-mots	Décodage (score composite BELO + LUM)	Compréhension texte « les deux poules »
22 mois	X							
60 mois		X	X	X	X	X		
84 mois							X	X

2.2.5. Hypothèses opérationnelles

Pour rappel, notre première hypothèse théorique est la suivante : après 7 ou 8 mois d'apprentissage formel, le niveau de lecture devrait être expliqué par les prédicteurs classiques de la lecture tels que le vocabulaire, les performances à la DRA, la conscience phonologique, la connaissance du nom des lettres et la mémoire à court terme phonologique. Ainsi, nous faisons quatre hypothèses opérationnelles en lien avec cette première hypothèse théorique :

- a) Pour le score de décodage, nous nous attendons à ce que les performances soient prédites par les différents scores aux tâches mesurant la conscience phonologique, le vocabulaire (EVIP) et la connaissance des lettres (Lollipop test) ainsi que par les scores obtenus aux tâches de répétition de pseudo-mots et de DRA à 60 mois ;
- b) De plus, nous nous attendons à ce que les scores obtenus aux tâches mesurant la conscience phonologique ainsi que la connaissance des lettres expliquent une part de variance plus importante du score de décodage que les autres prédicteurs mentionnés dans l'hypothèse a) ;
- c) Pour le score de compréhension en lecture, nous nous attendons à ce que les performances soient prédites directement par les scores aux tâches mesurant la conscience phonologique, le vocabulaire (EVIP), la connaissance des lettres ainsi que par le score à la tâche de DRA, mais pas par le score à la tâche de répétition de pseudo-mots ;
- d) De surcroît, nous nous attendons à ce que le score à la tâche mesurant le vocabulaire (EVIP) explique une plus grande part de variance du score de compréhension en lecture que les autres prédicteurs mentionnés dans l'hypothèse c).

Ensuite, notre deuxième hypothèse théorique est que les deux composantes de lecture seraient expliquées par des compétences plus précoces de langage oral. De cette hypothèse découlent ainsi quatre autres hypothèses plus précises :

- a) Nous nous attendons à ce que le score de vocabulaire réceptif à 22 mois (CCT) prédise les performances aux épreuves mesurant le décodage ;
- b) De même, nous nous attendons à ce que le score de vocabulaire réceptif à 22 mois (CCT) prédise les performances à la tâche de compréhension de texte ;
- c) De manière plus spécifique, nous nous attendons à ce que le pourcentage de variance expliquée par le score de vocabulaire réceptif à 22 mois (CCT) soit plus important pour les performances à la tâche de compréhension que pour les tâches mesurant le décodage ;
- d) Finalement, par rapport au modèle de départ ne prenant en compte que les prédicteurs classiques de la lecture à 60 mois, nous nous attendons à ce que la prise en compte du score de vocabulaire à 22 mois (CCT) permette d'expliquer un pourcentage de variance total plus important des performances de décodage et de compréhension en lecture.

3. Résultats

Les analyses descriptives ainsi que les analyses des résultats à proprement dit ont été réalisées avec le logiciel statistique SPSS (version 25). De plus, les résultats sont considérés comme significatifs à un seuil de $p < 0.05$.

3.1. Statistiques descriptives et nettoyage des données

Avant de commencer nos analyses, nous avons réalisé des analyses descriptives de nos données afin de voir comment elles se comportaient. Notre but premier était de s'assurer de la normalité de nos données et ainsi de voir si certaines valeurs extrêmes pouvaient être mises en évidence. Nous avons donc examiné une à une nos variables. Pour ce faire, nous avons principalement regardé l'asymétrie et le kurtosis de nos données. Nous avons vérifié que la valeur de l'asymétrie soit le plus proche possible de zéro et que la valeur du kurtosis soit comprise entre -2 et 2, auquel cas nous avons identifié les valeurs extrêmes pouvant influencer ces valeurs.

Pour rappel, à 22 mois les enfants ont passé une tâche de désignation d'image qui permettait de nous fournir une mesure de vocabulaire réceptif. Dans ces données, un individu (le 663) se comportait très différemment des autres (valeur extrême). Lorsqu'on le retire de l'échantillon, l'asymétrie et le kurtosis redeviennent dans la norme et les données suivent une distribution normale. Nous avons donc exclu ce participant de nos analyses.

À 60 mois les enfants ont passé différentes tâches. Parmi celles-ci, seules les données de la tâche de mémoire à court terme phonologique ne se distribuent pas normalement. En effet, un individu (le 616) se comporte très différemment. Là aussi, en excluant cette donnée, l'asymétrie et le kurtosis redeviennent dans la norme et les données suivent une distribution normale. Les données des tâches de connaissance des lettres, de conscience phonologique, de DRA et de vocabulaire réceptif suivent une distribution normale.

Pour la dernière session de mesures, après 7-8 mois d'apprentissage formel de la lecture, les performances des participants suivent une distribution normale pour le décodage, mais pas pour la compréhension en lecture. Effectivement, le score d'un individu (le 665) apparaît comme étant extrême. De plus, en retournant dans nos données récoltées lors de la passation, il nous paraît évident d'exclure cette donnée. En effet, le participant en question a eu beaucoup de mal lors de la lecture de texte et a dû être aidé par l'expérimentateur et nous n'avons donc

pas pu lui faire passer la tâche de compréhension. En excluant cette donnée, l'asymétrie et le kurtosis sont à nouveau dans les normes.

Finalement, lors d'analyses préliminaires de régression, deux autres individus ressortaient comme ayant des résultats aberrants, et ce, dans toutes les analyses. Ainsi, nous avons aussi exclu ces deux derniers individus des données (le 621 et le 623). En effet, si l'on regarde les résidus de ces deux individus, on remarque que l'écart entre ce qui est prédit et ce qui est réellement mesuré chez ces enfants est extrême. Du fait de cet écart, cela influence grandement les paramètres de notre modèle et c'est pourquoi ils ont été retirés des analyses.

Sur les 56 enfants contactés aux dépôts, seuls 49 se sont présentés à leur rendez-vous malgré plusieurs relances de notre part. Comme nous venons de le mentionner, quatre individus ont dû être exclus de nos données car ils présentaient des valeurs extrêmes (individu 663, 616, 621 et 623) et un autre a dû être exclu pour cause de donnée manquante (à la tâche de compréhension). De ce fait, nos analyses portent sur un total de 44 participants.

Dans le Tableau II, les moyennes, écart-types ainsi que les valeurs minimales et maximales de chaque tâche sont décrites pour la population totale des analyses.

Tableau II

Moyennes, écart-types, valeurs minimales et maximales des différentes tâches

22 mois	Vocabulaire (CCT)	
	Moy. σ	Min. Max.
N = 44	29.45 5	19 40

60 mois	Conscience phonologique		DRA		Répétition pseudo-mots		Conn. nom lettres		Vocabulaire (EVIP)	
	Moy. σ	Min. Max.	Moy. σ	Min. Max.	Moy. σ	Min. Max.	Moy. σ	Min. Max.	Moy. σ	Min. Max.
N = 44	4.86 5.03	0 15	1.52 0.51	0.97 3.10	16.52 2.357	11 20	13.66 4.86	2 18	125.50 14.75	84 152

7 ans	Décodage		Compréhension	
	Moy. σ	Min. Max.	Moy. σ	Min. Max.
N = 44	0.088 0.890	-1.93 1.66	6.7 1.17	4 8

3.2. Analyses de corrélations

Après avoir fait un premier tri dans nos données, nous nous devons de vérifier les corrélations entre les différentes variables. En effet, pour pouvoir faire une analyse de régression multiple, nous nous devons de vérifier certains postulats. Tout d'abord, nous devons regarder les relations entre les variables indépendantes car nous ne voulons pas observer de (multi)colinéarité entre elles, ce qui voudrait dire que nous aurions introduit des données qui mesurent possiblement la même chose. Afin d'éviter cela, il faut donc vérifier qu'une corrélation entre deux VIs ne se situe pas à plus de 0.9 (ou -0.9), ce qui pourrait indiquer un risque important de multicolinéarité. De plus, nous voulons vérifier que nos VIs corrélaient avec notre VD de manière significative car s'il n'y a pas de corrélation et donc pas de relation linéaire entre les deux, il est impossible de les introduire dans notre modèle de régression par la suite.

Tout d'abord, les corrélations entre les différentes variables indépendantes sont présentées dans la Figure 1. L'épreuve de vocabulaire réceptif à 22 mois (cf. CCT) corréla avec la tâche de DRA (cf. RAN) à 60 mois et avec la tâche de vocabulaire réceptif (cf. EVIP) à 60 mois. L'épreuve de conscience phonologique (cf. CP) corréla avec la tâche de connaissance des lettres au même âge. Finalement, la tâche de DRA corréla avec la tâche de vocabulaire réceptif (cf. EVIP) à 60 mois. Ainsi, nous pouvons donc voir qu'aucune des corrélations ne dépasse le seuil de 0.9 (ou -0.9), ce qui signifie que nous ne devrions pas avoir de problème de multicolinéarité pour la suite de nos analyses.

Ensuite, nous examinons les corrélations entre nos variables indépendantes et nos variables dépendantes (notamment afin de pouvoir par la suite sélectionner quelles sont les variables que nous allons introduire dans notre modèle de régression). Comme nous pouvons l'observer (voir Figure 1), les performances aux tâches mesurant le décodage (cf. Décodage en lecture) corrélaient significativement avec la tâche de vocabulaire réceptif (cf. CCT) à 22 mois, avec la tâche de conscience phonologique (cf. CP), avec la tâche de connaissance des lettres, avec la tâche de DRA (cf. RAN), avec la tâche de vocabulaire réceptif (cf. EVIP) à 60 mois et finalement avec la tâche de mémoire à court terme phonologique. La compréhension en lecture ne corréla significativement qu'avec la tâche de vocabulaire réceptif (cf. CCT) à 22 mois et avec la tâche de conscience phonologique (cf. CP).

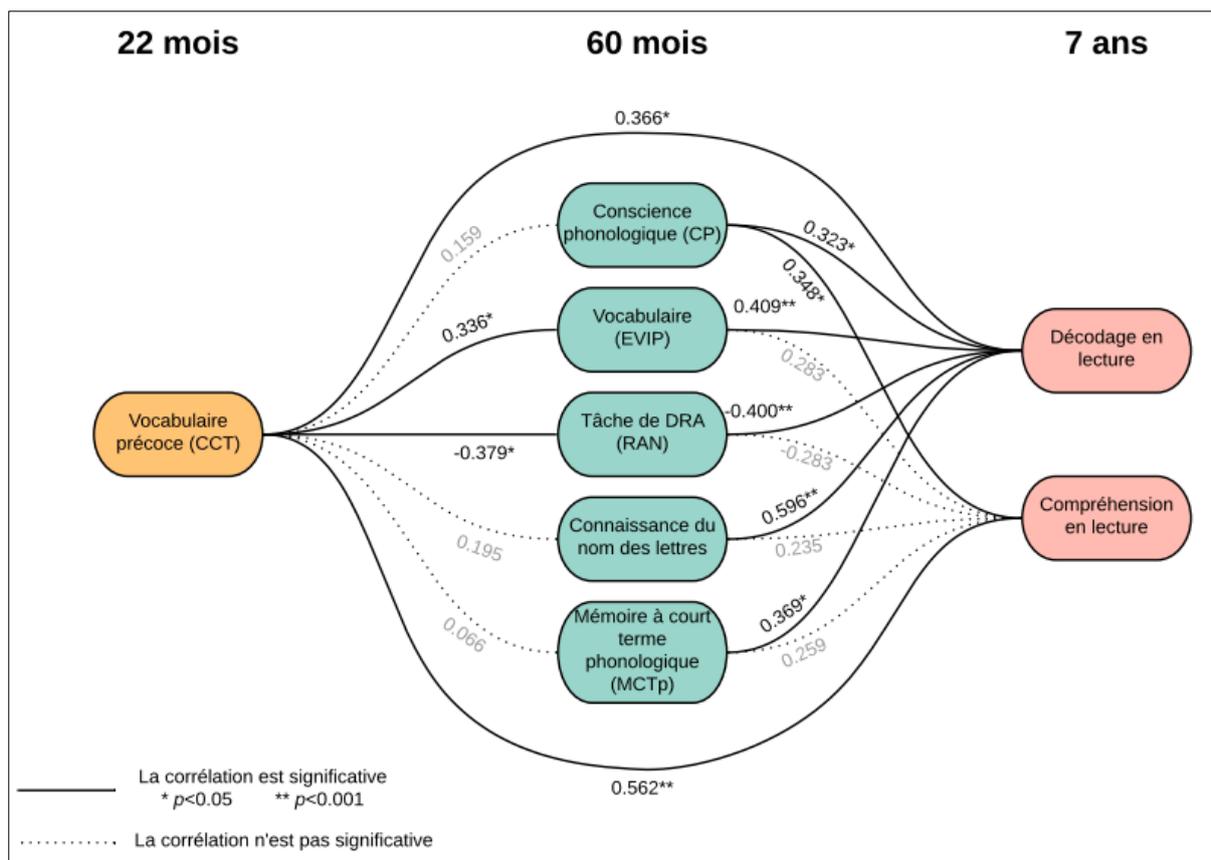


Figure 1

Résumé des analyses de corrélations

3.3. Analyses de régressions linéaires multiples

Afin de tester nos hypothèses, nous avons décidé de faire des régressions linéaires multiples et avons utilisé le logiciel statistique SPSS.

Ainsi, avant d'interpréter nos analyses, nous avons vérifié ces différentes prémisses. Premièrement, les variables indépendantes doivent être continues ou catégorielles et les variables dépendantes doivent être continues. Ensuite, nous ne devons pas avoir de variance égale à zéro, nous ne devons pas non plus observer de multicollinéarité entre les variables dépendantes (observé à l'aide du *Variance Inflation Factor (VIF)* < 10). De plus, il ne doit pas y avoir de corrélation entre les variables indépendantes et les variables externes (âge, sexe et raisonnement non-verbal). Puis, il doit y avoir une certaine homoscedasticité (homogénéité des variances résiduelles), une indépendance des erreurs (observé à l'aide du test *Durbin-Watson*) et une distribution normale des résidus. Finalement, il doit y avoir une relation linéaire entre les variables indépendantes et la variable dépendante.

En suivant l'ordre de nos hypothèses, nous nous intéressons donc d'abord aux différents prédicteurs de la lecture et leur lien avec les deux composantes de la lecture. Dans un deuxième temps, nous nous intéressons au lien entre le vocabulaire avant 2 ans et les compétences de lecture et nous observons si le vocabulaire dans la petite enfance continue de prédire une part de variance significative de la lecture lorsque les prédicteurs classiques sont pris en compte.

3.3.1. Analyse du lien entre les prédicteurs et la lecture

3.3.1.1. Les prédicteurs de la lecture et le décodage

Afin d'examiner les liens entre les prédicteurs de la lecture et le décodage, nous avons réalisé une régression linéaire multiple avec comme variable dépendante le score composite de décodage et comme variables indépendantes les scores aux mesures de conscience phonologique, de connaissance des lettres, de vocabulaire réceptif (EVIP) ainsi que les scores à une tâche de DRA et de répétition de pseudo-mots (ces mesures étant toutes corrélées avec la VD). Les variables indépendantes ont été intégrées de manière progressive au modèle avec la méthode pas-à-pas de manière à voir quelles sont les variables qui contribuent le plus et de manière significative au modèle. Ainsi, seules deux de ces variables ont été intégrées au modèle, les autres ont été écartées car leur apport n'est pas significatif ($p > 0.05$). Ces deux variables sont le score à la tâche de connaissance des lettres et le score à la tâche de DRA et rendent le modèle global significativement robuste, $F(2,41)=18.351$, $p < 0.001$. Les résultats montrent que le score de connaissance des lettres permet de prédire significativement les performances de décodage, $t(1,42)=4.924$, $p < 0.001$, et l'ajout de la tâche de DRA à ce premier modèle permet lui aussi de prédire significativement le décodage, $t(2,41)=-3.010$, $p=0.004$ (voir Tableau III).

Tableau III

Résumé des analyses de régression des variables permettant de prédire le décodage (N=44).

Variables	B	SE	Bêta	t	p
<i>VD : décodage</i>					
Connaissance des lettres	0.103	0.021	0.562	4.924	<0.001
DRA	-0.595	0.198	-0.343	-3.010	0.004

Cette analyse nous renseigne aussi sur l'apport de chaque variable. Les résultats montrent que la mesure de connaissance des lettres permet d'expliquer de manière significative 35.6% du score composite de décodage et que la performance à la tâche de DRA permet d'expliquer significativement 11.7% de variance supplémentaire du modèle. La combinaison de ces deux variables indépendantes contribue donc significativement à l'amélioration de l'explication de la variabilité du score de décodage et permet d'expliquer 47.2 % de la variance totale du modèle (voir Tableau IV).

Tableau IV

Résumé de l'ajustement du modèle de régression pour les variables prédisant le décodage (N=44).

Modèle	R ²	Variation de R ²	Variation de F	p
1	0.356	0.356	23.190	<0.001
2	0.472	0.117	9.061	0.004

VD : décodage

1 : connaissance des lettres

2 : connaissance des lettres + DRA

3.3.1.2. Les prédicteurs de la lecture et la compréhension

Dans le but d'explorer les liens entre les prédicteurs de la lecture et la compréhension en lecture, nous avons là aussi réalisé une régression linéaire multiple avec comme variable dépendante le score de compréhension en lecture et comme variable indépendante le score aux tâches mesurant la conscience phonologique (seule VI corrélée avec la compréhension). Les résultats montrent que le score de conscience phonologique permet de prédire significativement le score de compréhension en lecture, $t(1,42)=2.405$, $p=0.021$ (voir Tableau V). De plus, les résultats montrent que le score de la mesure de conscience phonologique permet d'expliquer de manière significative 12.1% de la variance du score de compréhension (voir Tableau VI).

Tableau V

Résumé des analyses de régression des variables permettant de prédire la compréhension (N=44).

Variables	B	SE	Bêta	t	p
<i>VD : compréhension</i>					
Conscience phonologique	0.081	0.034	0.348	2.405	0.021

Tableau VI

Résumé de l'ajustement du modèle de régression pour les variables prédisant la compréhension ($N=44$).

Modèle	R ²	Variation de R ²	Variation de F	<i>p</i>
1	0.121	0.121	5.783	0.021

VD : compréhension

1 : conscience phonologique

3.3.2. Analyse du lien entre le vocabulaire avant 2 ans et la lecture

3.3.2.1. Lien entre vocabulaire à 22 mois et décodage

Nous nous intéressons maintenant aux liens entre le vocabulaire réceptif à 22 mois et le décodage. Dans ce but, nous avons à nouveau réalisé une régression linéaire multiple avec comme variable dépendante le score de décodage et comme variable indépendante le score de vocabulaire réceptif (CCT) à 22 mois. Les résultats montrent que le score de vocabulaire permet de prédire significativement le score de décodage, $t(1,42)=2.547$, $p=0.015$ (voir Tableau VII). De plus, les résultats montrent que cette mesure de vocabulaire permet d'expliquer de manière significative 13.4% de la variance du score de décodage (voir Tableau VIII).

Tableau VII

Résumé des analyses de régression des variables permettant de prédire le décodage ($N=44$).

Variabes	B	SE	Bêta	<i>t</i>	<i>p</i>
<i>VD : décodage</i>					
Vocabulaire à 22 mois	0.065	0.026	0.366	2.547	0.015

Tableau VIII

Résumé de l'ajustement du modèle de régression pour les variables prédisant le décodage ($N=44$).

Modèle	R ²	Variation de R ²	Variation de F	<i>p</i>
1	0.134	0.134	6.488	0.015

VD : décodage

1 : Vocabulaire à 22 mois (CCT)

3.3.2.2. Lien entre vocabulaire à 22 mois et compréhension en lecture

Afin de déterminer le lien entre le vocabulaire à 22 mois et la compréhension en lecture à 7 ans, nous avons procédé là aussi à une analyse de régression multiple. Les analyses ont permis de mettre en évidence que le score vocabulaire à 22 mois permet de prédire de manière significative le score de compréhension en lecture plus tard, $t(1,42)=4.400$, $p<0.001$ (voir Tableau IX). De surcroît, ces analyses révèlent que cette mesure de vocabulaire à 22 mois permet d'expliquer 31.6% de la variance du score de compréhension à 7 ans (voir Tableau X).

Tableau IX

Résumé des analyses de régression des variables permettant de prédire la compréhension (N=44).

Variables	B	SE	Bêta	t	p
<i>VD : compréhension</i>					
Vocabulaire à 22 mois	0.132	0.030	0.562	4.400	<0.001

Tableau X

Résumé de l'ajustement du modèle de régression pour les variables prédisant la compréhension (N=44).

Modèle	R ²	Variation de R ²	Variation de F	p
1	0.316	0.316	19.360	<0.001

VD : compréhension

1 : Vocabulaire à 22 mois (CCT)

3.3.3. Analyse du lien entre vocabulaire à 22 mois et lecture, médiatisé par les prédictors classiques de la lecture

Dans cette partie, nous nous intéressons à un modèle qui prend en compte à la fois la mesure de vocabulaire à 22 mois (CTT) et les mesures des prédictors classiques à 60 mois afin de voir si un modèle plus global comme celui-ci est plus performant (voir Figure 2).

3.3.3.1. Lien entre vocabulaire à 22 mois et décodage, médiatisé par les prédicteurs

Pour analyser cela, nous avons repris le même modèle qu'au point 3.3.2.1., mais nous y avons ajouté les différentes mesures des prédicteurs de la lecture afin de voir si, lorsqu'il est pris en compte, le vocabulaire à 22 mois continue à expliquer une part de variance propre de façon significative. De plus, cela nous permet de voir si le fait de prendre en compte ces prédicteurs classiques permet d'ajouter de la valeur prédictive au modèle et ainsi le rendre plus fort.

L'analyse de régression multiple comprend comme variable dépendante le score de décodage et comme variables indépendantes la mesure de vocabulaire à 22 mois (CCT) et dans un autre bloc, les différentes mesures des prédicteurs du décodage. Après analyse, nous observons que seuls deux prédicteurs ont été inclus au modèle. Il s'agit de la connaissance des lettres et de la tâche de DRA. Les mesures des autres prédicteurs n'ont pas été incluses au modèle car leur ajout ne permet pas d'expliquer significativement plus de variance. Les résultats montrent que le modèle global prenant en compte le score de vocabulaire à 22 mois et le score aux tâches de connaissance des lettres et de DRA est significatif, $F(3,40)=12.899$, $p<0.001$. Ensuite, si on regarde l'apport de chaque variable au modèle, la mesure de vocabulaire à 22 mois permet d'expliquer de manière significative 13.4% du score de décodage, la mesure de connaissance des lettres permet d'expliquer 28.7% supplémentaires, et le score à la tâche de DRA permet d'expliquer 7.1% de la mesure du décodage. Ainsi, lorsqu'on tient compte de ces 3 variables, notre modèle global permet d'expliquer de manière significative un total de 49.2% du score de décodage (voir Tableau XII). Cependant, si nous regardons en détail les paramètres du modèle, et que nous regardons le coefficient de chaque variable, nous remarquons que la mesure de vocabulaire à 22 mois ne contribue pas significativement au modèle, $t(3,40)=1.228$, $p=0.226$. C'est au moment de l'ajout de la mesure de DRA au modèle que cette mesure devient non significative. Les mesures de connaissance des lettres et de DRA contribuent de leur côté significativement au modèle (voir Tableau XI).

Tableau XI

Résumé des analyses de régression des variables permettant de prédire le décodage (N=44).

Variabes	B	SE	Bêta	t	p
<i>VD : décodage</i>					
Vocabulaire à 22 mois	0.027	0.022	0.152	1.228	0.226
Connaissance des lettres	0.099	0.211	0.538	4.674	<0.001
DRA	-0.499	0.211	-0.288	-2.364	0.023

Tableau XII

Résumé de l'ajustement du modèle de régression pour les variables prédisant le décodage (N=44).

Modèle	R²	Variation de R²	Variation de F	p
1	0.134	0.134	6.488	0.015
2	0.420	0.287	20.281	<0.001
3	0.492	0.071	5.589	0.023

VD : décodage

1 : vocabulaire à 22 mois

2 : vocabulaire à 22 mois + connaissance des lettres

3 : vocabulaire à 22 mois + connaissance des lettres + DRA

3.3.3.2. Lien entre vocabulaire à 22 mois et compréhension, médiatisé par les prédictors

Nos dernières analyses portent sur le même modèle que celui exposé au point 3.3.2.2., en ajoutant le seul prédictor trouvé pour la compréhension, soit le score aux tâches mesurant la conscience phonologique. L'analyse de régression multiple intègre donc comme variables indépendantes le score de vocabulaire à 22 mois ainsi que la mesure de conscience phonologique dans deux blocs distincts et comme variable dépendante le score à la tâche de compréhension en lecture. Les résultats montrent que le score de vocabulaire (CTT) prédit significativement le score de compréhension en lecture et que cette variable contribue significativement au modèle, $t(1,42)=4.183$, $p<0.001$. Puis, lorsque la mesure de conscience

phonologique est prise en compte dans le modèle, cette dernière contribue elle aussi significativement au modèle, $t(2,41)=2.135$, $p=0.039$ (voir Tableau XIII).

Tableau XIII

Résumé des analyses de régression des variables permettant de prédire la compréhension (N=44).

Variables	B	SE	Bêta	t	p
VD : compréhension					
Vocabulaire à 22 mois	0.122	0.029	0.519	4.183	<0.001
Conscience phonologique	0.062	0.029	0.265	2.135	0.039

Finalement, lorsque nous nous intéressons à l'apport de chaque élément au modèle, nous remarquons que la mesure de vocabulaire à 22 mois permet d'expliquer à elle seule 31.6% du modèle alors que la mesure de conscience phonologique ne permet d'expliquer que 6.8% supplémentaires du score de compréhension en lecture. Ainsi, ces deux mesures permettent de prédire ensemble 38.4% de la compréhension en lecture à 7 ans (voir Tableau XIV).

Tableau XIV

Résumé de l'ajustement du modèle de régression pour les variables prédisant la compréhension (N=44).

Modèle	R²	Variation de R²	Variation de F	p
1	0.316	0.316	19.360	<0.001
2	0.384	0.068	4.557	0.039

VD : compréhension

1 : vocabulaire à 22 mois

2 : vocabulaire à 22 mois + conscience phonologique

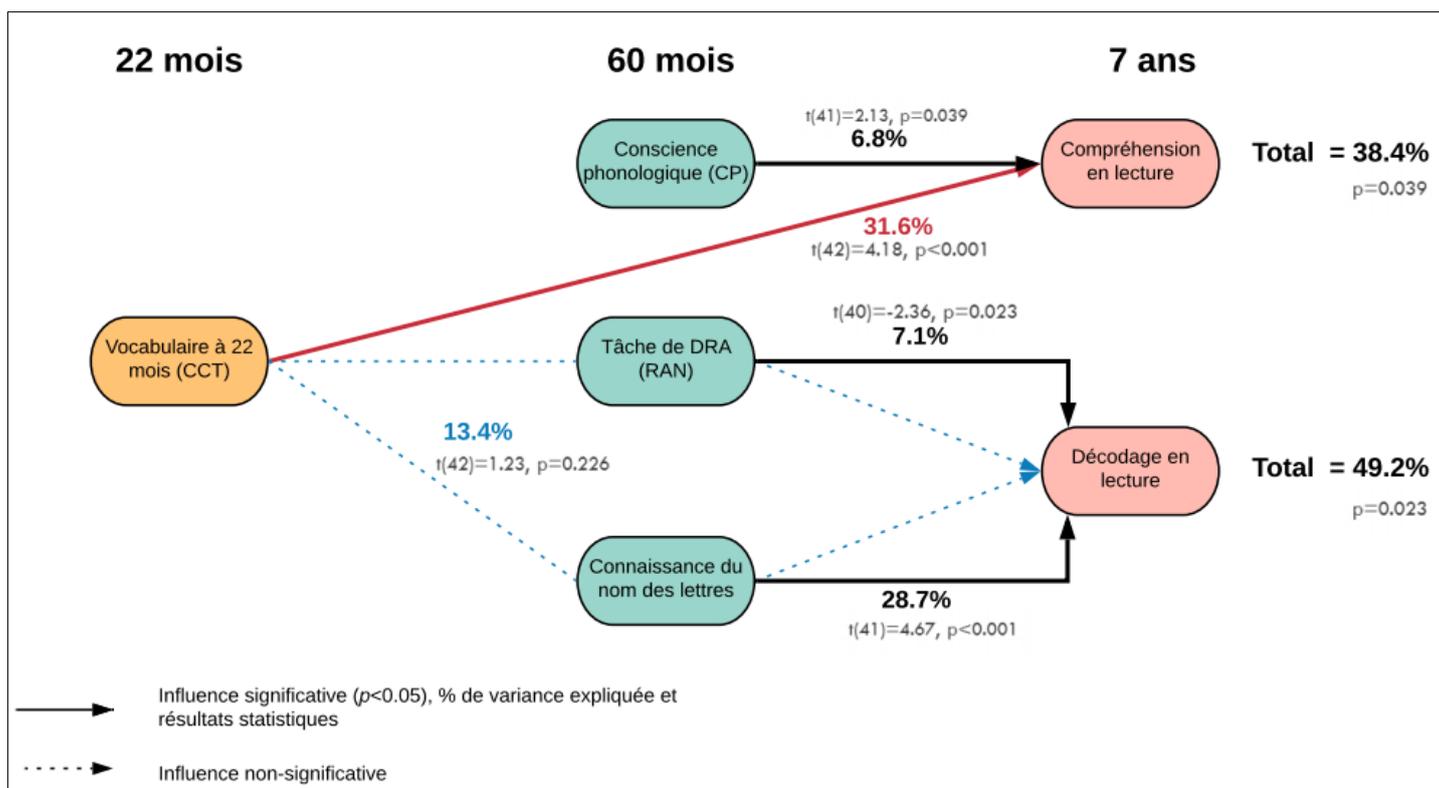


Figure 2

Résumé des analyses de régression entre vocabulaire à 22 mois et lecture, médiatisés par les prédicteurs à 60 mois

4. Discussion

4.1. Rappel des hypothèses et des principaux résultats obtenus

Pour rappel, notre recherche a comme premier but de voir si les principaux prédicteurs de la lecture précédemment décrits dans des langues telles que l'anglais s'appliquent également au français. D'autre part, elle a une visée plus exploratoire car nous cherchons à savoir si des mesures plus précoces de langage oral permettent de prédire les futures compétences de lecture. À notre connaissance, ce type de recherche n'a pas encore été menée sur une population francophone ainsi que sur la base de données objectives.

Dans cette perspective, nous avons d'abord tenté de répliquer certains résultats obtenus ces dernières années dans le domaine de la lecture. Nous avons en effet voulu savoir si certaines compétences développées et acquises assez tôt dans le parcours scolaire d'un enfant ont une capacité prédictive sur leurs compétences futures de lecture. Nos résultats ont mis en évidence que des mesures de connaissance des lettres et de DRA permettent de prédire de futures compétences de décodage, alors que des mesures de conscience phonologique, de vocabulaire et de mémoire à court terme phonologique ne semblent pas le permettre. Concernant la compréhension, seule la conscience phonologique permet de la prédire, alors que les mesures de connaissance des lettres, de DRA et de vocabulaire ne semblent pas prédictives.

Puis, en nous basant sur deux études menées en langue anglaise (Lee, 2011 ; Duff et al., 2015), nous nous sommes questionnés sur l'existence d'un lien prédictif entre une mesure de vocabulaire avant 2 ans (à 22 mois) et les deux composantes de lecture après plusieurs mois d'apprentissage de celle-ci. Nos résultats ont confirmé le lien prédictif entre le vocabulaire dans la petite enfance avec le décodage et la compréhension. De plus, le vocabulaire à 22 mois permet de prédire une part de variance plus importante de la compréhension que du décodage. Finalement, nous avons trouvé qu'un modèle global incluant une mesure de vocabulaire à 22 mois ainsi que certaines mesures à 60 mois permet d'expliquer une part plus importante de variance des composantes de lecture qu'un modèle ne prenant pas en compte une mesure de vocabulaire dans la petite enfance.

De ce fait, nous allons maintenant interpréter et discuter les différents résultats présentés ci-dessus afin de répondre plus précisément à nos différentes hypothèses exposées dans la partie méthodologique.

4.2. Interprétation et discussion des résultats

4.2.1. Lien entre prédicteurs classiques et la lecture

Nous regardons tout d'abord les différents prédicteurs du décodage et cela reprend notre première hypothèse théorique. Nous nous sommes donc intéressés à leur rôle prédictif sur les deux composantes de lecture, et ce de manière séparée.

4.2.1.1. Les prédicteurs du décodage

Suite à l'élaboration de notre cadre théorique, nous avons comme hypothèse que les différentes mesures de conscience phonologique, DRA, connaissance des lettres, vocabulaire et de mémoire à court terme phonologique soient toutes les cinq prédictrices du décodage. En nous basant sur les différents résultats obtenus aux premières analyses de corrélation, nous pouvons voir que ces 5 mesures sont corrélées de manière significative avec le décodage. Ainsi, il existe bien une relation entre les variables indépendantes mentionnées et le décodage. Néanmoins, si deux mesures corrélaient, cela ne veut pas dire que l'une permet de prédire l'autre. Les analyses de régression nous révèlent que toutes ces mesures ne peuvent pas être considérées comme des prédicteurs du décodage. En effet, seules la connaissance des lettres et la DRA permettent de prédire significativement le décodage plus tard, ce qui va dans le sens de nos hypothèses et de différents auteurs (Schatschneider et al., 2004 ; Catts et al., 2015). Qu'en est-il des trois autres mesures et comment pourrions-nous expliquer cette absence de lien prédictif dans nos résultats ?

Premièrement, nous pensons que la conscience phonologique serait un bon prédicteur du décodage, mais cela ne ressort pas dans nos analyses. D'ailleurs, il s'agit de la mesure qui corréle le moins avec le décodage. À notre sens, ces résultats peuvent s'expliquer notamment par une divergence de méthodologie entre les recherches sur lesquelles nous nous sommes basées et la nôtre. En effet, dans leur étude, Schatschneider et collaborateurs (2004) ont utilisé un total de 7 épreuves pour mesurer la conscience phonologique, dont une concernant la rime et 6 mesurant la conscience phonémique. Ainsi, leurs mesures permettent de donner une représentation assez précise de ce que peut faire l'enfant au niveau de la manipulation de phonèmes. Or, dans notre étude, les enfants n'ont passé qu'une épreuve sur la rime et qu'une seule sur le phonème. Ainsi, si la consigne de l'épreuve concernant le phonème est mal comprise ou que l'épreuve est mal réussie, le score global de conscience phonologique est tout de suite affecté et pourrait ne pas être représentatif des compétences réelles de l'enfant. De même, Catts et collaborateurs (2015) utilisent eux aussi plusieurs tâches pour établir leur

mesure de conscience phonologique. Ainsi, ces différences dans la méthodologie au niveau du choix des tâches pourraient expliquer pourquoi nous n'observons pas un lien plus fort entre la conscience phonologique vers 60 mois et le décodage au début de son apprentissage formel. De plus, pour certains auteurs (Suortti & Lipponen, 2016), le développement de cette capacité serait facilité par la connaissance des lettres. Donc comme nous observons un fort effet de cette dernière, l'effet de la conscience phonologique pourrait ainsi être masqué.

Ensuite, le vocabulaire est la deuxième mesure qui corrèle le plus avec le décodage. Cependant, cette mesure n'a pas été incluse dans le modèle car elle ne permet pas de prédire significativement une part de variance du décodage. Même si nous avions prédit le contraire, ce résultat ne nous étonne que moyennement. En effet, l'idée que le vocabulaire soit un prédicteur du décodage est controversée dans la littérature. Nos résultats vont dans le sens de ceux obtenus par Schatschneider et collaborateurs (2004) qui ont eux aussi utilisé une mesure de vocabulaire réceptif. Comme nous l'avons mentionné dans le cadre théorique, certains auteurs (Scarborough, 2001 ; Catts et al., 1999 ; Ricketts et al., 2007) ont trouvé que le vocabulaire est un prédicteur du décodage. Néanmoins, ces auteurs se sont basés soit sur des mesures de vocabulaire réceptif et expressif, soit sur une mesure de vocabulaire expressif uniquement. Ainsi, ces différences de mesures peuvent expliquer pourquoi nous n'obtenons pas des résultats en faveur d'un lien prédictif du vocabulaire sur le décodage. Finalement, cela nous questionne sur l'importance et la nature du lien que pourrait avoir le vocabulaire expressif en tant que prédicteur du décodage car certains auteurs ont mis en évidence un lien direct entre vocabulaire expressif et décodage (Wise, Sevcik, Morris, Lovett & Wolf, 2007).

Finalement, nous n'observons pas de lien prédictif de la mémoire à court terme phonologique sur le décodage. Ces résultats vont donc à l'encontre de notre hypothèse, mais supportent l'idée que cette composante soit liée au décodage via la conscience phonologique. C'est notamment une hypothèse avancée par certains auteurs (Engen & Høien, 2002). En effet, comme nous n'observons pas de lien prédictif de la conscience phonologique sur le décodage et que la mémoire à court terme phonologique pourrait y être reliée, il nous semble plausible de ne pas observer un lien prédictif de la mémoire à court terme phonologique sur le décodage.

Une deuxième hypothèse que nous avons, en relation avec le décodage, était que parmi les prédicteurs mentionnés, ce sont la conscience phonologique et la connaissance des lettres qui permettent de mieux prédire le décodage (et donc qui expliquent une plus grande part de sa variance). Comme nous l'avons vu ci-dessus, la conscience phonologique n'est pas un

prédicteur du décodage dans notre étude, et cette mesure ne peut donc pas mieux expliquer le décodage que d'autres mesures. En revanche, la connaissance des lettres permet d'expliquer 35.6% du décodage et est donc bien le meilleur prédicteur de celui-ci. Ce résultat va dans le sens de ce qui a été trouvé par d'autres auteurs (Catts et al., 2015). Cela renforce aussi l'idée émise par Foulin (2005) selon laquelle ces connaissances sont un boost à l'acquisition de la littéracie. En effet, la connaissance des lettres joue un rôle important car elle fournit un premier support pour l'acquisition de la connaissance du son des lettres ainsi que pour la correspondance entre graphèmes et phonèmes. Contrairement à ce que nous avons prédit, ce sont les performances à une tâche de dénomination rapide automatisée qui permettent ensuite de mieux prédire le décodage à hauteur de 11.7%. Comme nous l'avons vu précédemment, cette tâche demande de mettre en œuvre certains mêmes processus cognitifs que ceux impliqués durant l'activité de décodage (Wolf & Bowers, 1999, cités dans Araújo & al., 2015). C'est certainement pourquoi cette tâche permet de prédire une part de celle-ci dans notre étude.

Nous en arrivons à la conclusion que seules les mesures de connaissance des lettres et de DRA à 60 mois permettent de prédire les futures compétences de décodage après quelques mois d'apprentissage formel de la lecture. Leur capacité prédictive est d'ailleurs assez importante car à elles deux, elles permettent d'expliquer 47.2% de la variance du décodage.

4.2.1.2. Les prédicteurs de la compréhension

Pour rappel, nous avons comme hypothèse que les différentes mesures de conscience phonologique, de DRA, de connaissance des lettres et de vocabulaire à 60 mois seraient des prédicteurs de la compréhension en lecture au début de son apprentissage formel. Contrairement à ce que nous avons observé avec le décodage, seule une de ces mesures corrèle avec la compréhension. Ainsi, comme les mesures de vocabulaire, de connaissance des lettres et de DRA ne corrèlent pas avec la compréhension dans les analyses préliminaires, elles ne peuvent pas en être des prédicteurs et cela va donc à l'encontre de nos hypothèses. Néanmoins, la mesure de conscience phonologique corrèle avec la compréhension, et c'est pourquoi c'est la seule mesure qui puisse être considérée comme un prédicteur de la compréhension. Ce résultat va dans le sens d'autres études déjà mentionnées (Schatschneider et al., 2004 ; Catts et al., 2016). Comment expliquer ces différences de résultats par rapport à nos hypothèses ?

Une explication possible concernant l'absence de lien entre connaissance des lettres et la compréhension est que les tâches utilisées par certains auteurs ne mesurent pas exactement les mêmes choses. En effet, Schatschneider et collaborateurs (2004) ont trouvé un lien prédictif

de cette capacité, mais ils ont utilisé la connaissance du son des lettres et non pas la connaissance du nom des lettres comme nous l'avons fait. D'autres (Catts et al., 2016) se sont basés sur une tâche de fluence et donc, cela diffère là aussi de ce que nous avons fait. De plus, lors de cette tâche, il était simplement demandé aux enfants de désigner 5 lettres et d'en dénommer 5 autres. Ainsi, cette connaissance est testée sur la base d'une dizaine de lettres, ce qui n'est pas forcément représentatif.

Concernant le vocabulaire, les mêmes explications que pour le décodage peuvent être avancées. En effet, les études des auteurs sur lesquelles nous nous sommes basés (Catts et al., 2016) prennent également en compte des mesures de vocabulaire expressif. L'autre étude que nous mentionnons (Schatschneider et al., 2004) ne prend en compte que le vocabulaire réceptif et ces auteurs ne trouvent pas non plus de lien entre vocabulaire et compréhension en lecture. À nouveau, cela questionne sur l'importance du vocabulaire expressif (et non réceptif) sur les capacités de compréhension en lecture. De plus, les mesures de compréhension effectuées par certains auteurs (Catts et al., 2016) sont plus tardives que les nôtres. En effet, elles ont été faites deux ans plus tard (à 9 ans), ce qui pourrait expliquer pourquoi nous n'observons pas de lien entre vocabulaire et compréhension au début de son apprentissage formel. En effet, comme nous l'avons vu avec le modèle de Coltheart (1978) ou le modèle de Frith (1986), le jeune lecteur utilise d'abord sa voie d'assemblage afin de développer son lexique orthographique, puis lorsque la lecture devient plus automatisée, la voie d'adressage prend le dessus. Nous pensons donc que la taille du vocabulaire serait plus en lien avec la mise en place de cette voie d'adressage (qui a lieu plus tard) qui permet un accès plus direct au sens, et donc à la compréhension.

Puis, l'absence de lien entre la tâche de DRA et la compréhension ne peut pas s'expliquer par des différences de méthodologie. Cependant, les différentes études qui tentent d'expliquer son lien avec la lecture font surtout référence aux processus par lesquels la forme phonologique des mots est récupérée (Wimmer & al., 2000, cité par Lervag & Hulme, 2009). Cette récupération est certes importante dans la lecture, mais le processus de compréhension est plus complexe que cela car il faut aussi pouvoir accéder au sens, mettre en lien différents éléments et aussi pouvoir s'appuyer sur des connaissances générales du monde (Gough & Tunmer, 1986). Nous n'avons d'ailleurs connaissance aucune étude qui permette d'expliquer directement le lien entre cette tâche et la compréhension.

Une deuxième hypothèse que nous avons en relation avec la compréhension était que, parmi les prédicteurs mentionnés, c'est le vocabulaire qui permet prédire le mieux la compréhension et donc qui en explique une plus grande part de variance. Étant donné que le vocabulaire n'est pas un prédicteur de la compréhension dans notre étude, cette hypothèse est invalidée. Comme le montrent les résultats, c'est la conscience phonologique qui est le meilleur prédicteur de la compréhension en lecture puisque cette mesure permet d'expliquer 12.1% de la variance de la compréhension. Ce résultat peut notamment s'expliquer par le fait que, comme les enfants sont au début de l'apprentissage de la lecture, ils utilisent encore beaucoup la voie d'assemblage pour décoder les mots et ainsi accéder au sens. Et comme nous l'avons vu, le décodage des mots via la voie d'assemblage demande de mettre en œuvre certains processus et certaines compétences dont fait partie la conscience phonologique. Il est certes important que l'enfant connaisse les mots qu'il lit pour comprendre, mais il est encore plus important qu'il parvienne à les décoder avant. Certains auteurs (Engen & Høien, 2002) sont d'ailleurs arrivés aux mêmes résultats avec, eux aussi, des enfants qui commencent à apprendre à lire. Dans leur étude, la conscience phonologique permet d'expliquer 15% de la variance de compréhension de texte. Ils mettent ces deux capacités en lien, notamment par le fait que certaines compétences nécessaires pour comprendre un texte sont déterminées en partie par la conscience phonologique. C'est d'ailleurs le cas du vocabulaire. En effet, le développement de la conscience phonologique faciliterait l'acquisition du vocabulaire, et faciliterait ainsi indirectement la compréhension. Les auteurs spécifient néanmoins que les enfants de leur étude sont au début de l'apprentissage de la lecture, et que le lien entre ces deux compétences peut évoluer lorsque les enfants deviennent plus âgés. D'autres auteurs (Vellutino, Tunmer, Jaccard & Chen, 2007) ont d'ailleurs trouvé que la relation entre conscience phonologique et compréhension est plus forte chez un groupe d'enfants plus jeunes que chez un groupe d'enfants plus âgés.

4.2.2. Liens entre vocabulaire dans la petite enfance et lecture

Pour rappel, nos hypothèses en lien avec le vocabulaire avant 2 ans se basent sur deux études menées en anglais et dont la méthodologie varie un peu de la nôtre. En effet, l'étude de Lee (2011) ne prend en compte que le vocabulaire expressif, et celle de Duff et al. (2015) tient compte du vocabulaire expressif et réceptif. Comme nous tenons à explorer ce lien et non pas à répliquer leurs résultats, nous n'avons choisi qu'une seule mesure de vocabulaire réceptif car il s'agit d'une mesure objective qui nous semble plus pertinente pour le processus de compréhension. De plus, nous avons choisi de mener deux analyses séparées (une sans les

prédicteurs classiques et l'autre avec) afin de voir si le vocabulaire seul dans la petite enfance a un lien fort avec la lecture ou non. Puis, nous avons décidé d'inclure les prédicteurs classiques au modèle afin de voir si le vocabulaire avant 2 ans continuait de prédire une part de variance de la lecture.

Les résultats ont mis en évidence qu'il y a un lien prédictif du vocabulaire réceptif dans la petite enfance tant sur le décodage que sur la compréhension en lecture. Ainsi, cela va dans le sens de nos hypothèses et des études précédemment mentionnées (Lee, 2011 ; Duff et al., 2015). Cela vient renforcer l'idée selon laquelle une mesure plus précoce de langage oral permet de prédire les capacités futures de lecture 5 ans plus tard. De plus, tout comme Duff et collaborateurs (2015), nous avons trouvé que le vocabulaire réceptif à 22 mois est un meilleur prédicteur de la compréhension que du décodage, puisqu'il permet d'expliquer respectivement 31.6% et 13.4% de leur variance. Néanmoins, nous observons une différence de prédiction plus importante que ces auteurs car dans notre étude, le vocabulaire dans la petite enfance permet de prédire plus de deux fois plus de variances de compréhension que de décodage. Ainsi, le vocabulaire avant 2 ans a à lui seul un lien assez fort avec la lecture.

Lorsque nous tenons compte des prédicteurs classiques de la lecture, le vocabulaire à 22 mois conserve son pouvoir prédictif, et en plus, les prédicteurs ajoutent de la force à cette mesure sans la diminuer. En effet, à lui seul le vocabulaire avant 2 ans n'explique que 13.4% du décodage et lorsque les prédicteurs classiques sont pris en compte, ils expliquent ensemble 49.2% du décodage (vocabulaire à 22 mois (13.4%) + connaissance des lettres (28.7%) + DRA (7.1%)). De même, le vocabulaire à 22 mois explique 31.6% de la compréhension à lui seul et lorsqu'on ajoute la conscience phonologique, ce pourcentage augmente jusqu'à 38.4%. Ainsi, cela va dans le sens de notre hypothèse selon laquelle le vocabulaire dans la petite enfance permet de renforcer la part de variance expliquée des mesures de décodage et de compréhension en lecture, et cela même lorsque les prédicteurs classiques sont pris en compte.

Néanmoins, même si les modèles globaux prenant en compte le vocabulaire à 22 mois et les différents prédicteurs sont significatifs, il faut s'attarder sur celui expliquant le décodage. En effet, comme nous l'avons vu dans les résultats, lorsque nous ajoutons au modèle la mesure de DRA, le vocabulaire à 22 mois ne contribue plus significativement au modèle. L'hypothèse que nous pouvons faire est que le vocabulaire réceptif à 22 mois et la DRA sont liés car ils sont tous deux en lien avec la récupération lexicale d'informations en mémoire (Wimmer & al., 2000, cité par Lervag & Hulme, 2009). C'est donc la variable la plus proche en termes d'âge (et donc

la DRA) qui explique mieux le décodage. Le modèle global est certes significatif, mais nous avons forcé l'entrée de cette variable (CCT) au modèle dans nos analyses. Ainsi, le vocabulaire à 22 mois prédit bien le décodage vers 7 ans, mais cette relation est médiatisée par la tâche de DRA à 5 ans. Contrairement au décodage, la compréhension en lecture continue d'être expliquée significativement par le vocabulaire à 22 mois, même lorsque la conscience phonologique est ajoutée au modèle. Cela renforce l'importance et la force de cette mesure dans la petite enfance comme prédicteur de la compréhension.

Pour conclure, nous pouvons voir que la prise en compte de la mesure de vocabulaire dans la petite enfance n'a pas beaucoup d'influence sur la part de variance expliquée du décodage (qui passe de 47.2% à 49.2%), mais qu'elle a une grande influence sur la part de variance expliquée de la compréhension en lecture. En effet, l'ajout au modèle d'une mesure de vocabulaire réceptif avant 2 ans permet d'expliquer trois fois plus de variances de la compréhension qui passe de 12.1% à 38.4%. Ces résultats nous laissent ainsi penser que le vocabulaire dans la petite enfance est une mesure qui est plus en lien avec la compréhension que le décodage. En effet, il est important de pouvoir comprendre le sens des mots pour avoir accès au sens plus global d'un texte par exemple. De plus, ces résultats encouragent les cliniciens et les professionnels à faire des mesures précoces. En effet, le fait de faire une seule et unique mesure à un instant « t » renseigne moins bien que plusieurs mesures prises à différents moments du développement.

4.3. Limites de la présente étude

Quelques limites concernant notre étude sont toutefois à prendre en compte lors de la lecture de ce document. Premièrement, notre étude se base sur un petit échantillon de 44 enfants, ce qui peut compromettre la représentativité des résultats ainsi que leur fiabilité. De plus, certains de ces enfants habitent en France et suivent leur scolarité là-bas, alors que la majeure partie des autres enfants suivent leur scolarité en Suisse. Il existe en effet des différences dans les programmes scolaires et dans les enseignements donnés, et cela peut aussi biaiser nos résultats. Il est aussi important de prendre en compte que cette étude s'est faite sur de nombreuses années et que les passations n'ont pas toujours été réalisées par les mêmes personnes et dans les mêmes conditions expérimentales, ce qui peut là aussi influencer les résultats ainsi que les performances des enfants. Ensuite, les tâches utilisées ne sont pas toujours écologiques et ne sont donc pas toujours réellement représentatives du niveau des enfants.

Finalement, certaines durées de passations ont pu être longues, ce qui a pu entraîner une certaine fatigabilité et un manque de concentration chez certains enfants.

Comme nous l'avons mentionné dans notre discussion, les tâches utilisées n'étaient pas toujours optimales et ne permettent pas toujours d'avoir une mesure représentative de la compétence réelle de l'enfant. Par exemple, la tâche de connaissance de lettres ne teste que 10 lettres (en désignation et dénomination). La conscience phonologique n'est testée qu'à l'aide d'une tâche sur la rime et d'une sur le phonème. Puis, la tâche de vocabulaire ne mesure que le vocabulaire réceptif. Nous aurions ainsi pu utiliser des tâches qui prennent en compte à la fois le son et le nom de toutes les lettres, varier et augmenter le nombre de tâches mesurant la conscience phonologique et ajouter une tâche mesurant le vocabulaire expressif à 60 mois. Cela dit, la longueur de la passation s'en serait également ressentie. Finalement, la tâche de compréhension que nous avons utilisée pose certaines questions. En effet, il s'agit de notre seule mesure de compréhension qui se compose d'une lecture de texte suivie de 8 questions. De plus, 34% des participants de notre étude ont atteint le score maximal à cette épreuve, ce qui peut conduire à un effet plafond. Nous n'avons donc pas de mesure de compréhension au niveau du mot et/ou de la phrase. Il aurait été intéressant d'ajouter ce type d'épreuve pour obtenir un score de compréhension plus global et ainsi plus représentatif.

Finalement, nous avons fixé le seuil de significativité à $p < 0.05$ pour nos analyses. Néanmoins, nos résultats sont à prendre avec précaution car nous n'avons pas procédé aux corrections (de Bonferonni) nécessaires demandées par les multiples comparaisons effectuées sur nos données.

4.4. Perspectives futures et ouverture clinique

Les recherches dans le domaine anglophone ont déjà clairement démontré que certaines capacités telles que la conscience phonologique, la connaissance des lettres, le vocabulaire, la dénomination rapide automatisée et la mémoire à court terme phonologique permettent de prédire des compétences futures de lecture. Notre étude permet d'amener des résultats dans le domaine francophone et suggère que vers l'âge de 5 ans, ce sont des compétences de conscience phonologique, de connaissance des lettres et de DRA qui permettent de prédire les compétences futures de lecture. Ainsi, il nous semble pertinent qu'un accent particulier soit mis à l'école sur le développement de la conscience phonologique ainsi que sur la connaissance des lettres. Comme nous l'avons vu, ces deux domaines de compétences sont essentiels pour une bonne mise en place de la lecture, même s'ils ne sont pas suffisants. La DRA est une tâche complexe

qui met en place de nombreux processus et il est compliqué d'en faire un objet de travail, c'est pourquoi nous ne recommandons pas un travail à ce niveau. Finalement, même si le vocabulaire n'a pas démontré de rôle prédictif vers 5 ans, nous avons pu mettre en évidence un rôle prédictif de celui-ci à 22 mois déjà. De ce fait, nous recommandons de mettre l'accent sur l'enrichissement du vocabulaire chez les enfants, et ce, dès que possible. Finalement, la contribution nouvelle qu'apporte notre étude est qu'elle concerne une population francophone et que nos mesures ainsi que nos analyses ont pour base des données objectives récoltées de manières standardisées.

De plus, cette étude pourrait s'inscrire dans la continuité d'autres domaines d'étude qui concernent les modèles de réponse à l'intervention (*responsiveness-to-intervention (RTI)* en anglais) et qui s'inscrivent dans une visée de prévention. Ce domaine est encore peu étudié en français et, de ce fait, pas ou encore peu mis en place dans nos écoles suisses. Cependant, en anglais, cette méthodologie est déjà appliquée dans certaines écoles américaines et largement étudiée (Fuchs & Vaughn, 2012). Le but principal de cette approche est de fournir à tous les enfants un enseignement jugé efficace par la recherche (en lien avec l'*Evidence Based Practice*) et d'évaluer régulièrement tous les élèves à l'aide d'un dépistage universel. Le but est d'identifier les enfants à risque de développer des difficultés d'apprentissage. Ce dépistage permet ensuite de fournir aux élèves potentiellement à risque un autre enseignement intensif et ciblé qui a pour but de remédier à leurs difficultés avant que celles-ci ne les handicapent pour la suite (Desrochers, Laplante & Brodeur, 2015). De plus, nous sommes d'avis que cette approche serait à considérer dans nos écoles suisses qui se veulent de plus en plus inclusives. Ainsi, les résultats que nous avons obtenus fournissent une première base d'éléments scientifiques utilisables pour la création d'un outil de dépistage qui s'inscrirait dans une approche de réponse à l'intervention. La mise en place d'une telle approche permettrait aux logopédistes, ainsi qu'aux divers professionnels de l'éducation de pouvoir faire de la prévention et de prévenir certains troubles, tels que ceux liés au langage écrit.

5. Conclusion

Cette étude sur les liens entre vocabulaire dans la petite enfance et compétences futures de lecture avait pour but d'apporter de nouvelles pistes dans un domaine qui est encore peu développé au niveau francophone. Elle avait notamment pour objectif de confirmer le rôle prédictif de certaines compétences à 60 mois qui sont avérées dans de nombreuses langues comme des prédicteurs de la lecture. De plus, cette étude avait comme deuxième objectif d'investiguer si des compétences de langage oral plus précoces avaient un rôle prédictif des capacités de lecture lors de son acquisition. Les résultats montrent que des mesures de connaissance des lettres et de DRA à 60 mois permettent de prédire le décodage vers 7 ans. De plus, la conscience phonologique à 60 mois permet de son côté de prédire la compréhension en lecture à 7 ans. Puis, le vocabulaire réceptif mesuré à 22 mois permet de prédire tant le décodage que la compréhension à 7 ans, même si son pouvoir prédictif est plus important pour la compréhension. Finalement, prendre à la fois en compte des mesures de vocabulaire avant 2 ans ainsi que les prédicteurs mentionnés permet d'expliquer près de 50% du décodage et près de 40% de la compréhension en lecture vers 7 ans.

Cette étude amène donc des résultats intéressants qu'il faudrait toutefois reproduire sur la base d'un échantillon plus grand afin de pouvoir les généraliser et en tirer des conclusions. Ainsi, des études futures dans ce domaine sont nécessaires. De plus, il serait intéressant d'ajouter à ce modèle des données concernant la réponse à l'intervention (voir notamment Catts et al., 2015) et d'autres données concernant le risque familial de difficultés en lecture (tel que Duff et al., 2015). En effet, nous pensons que l'ajout de ces composantes au modèle de prédiction de la lecture permettrait d'expliquer davantage de variance des composantes de décodage et de compréhension.

6. Bibliographie

- Alario, F.-X. & Ferrand, L. (1999). A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, 31(3), 531-552.
- Araújo, S., Reis, A., Petersson, K. M. & Faísca, L. (2015). Rapid Automated Naming and Reading Performance: A Meta-Analysis. *Journal of Educational Psychology*, 3, 868-883.
- Bandura, A. (2007). *Auto-efficacité : le sentiment d'efficacité personnelle*. Bruxelles : De Boeck Université.
- Bernicot, J.L., Bert-Erboul, A., (2009). *L'acquisition du langage par l'enfant*. Paris: Editions in Press.
- Bonin, P., Peereman, R., Malardier, N., Méot, A. & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, 35(1), 158-167.
- Briquet-Duhazé, S. (2015). Définir les rituels à l' école maternelle : un paradoxe institutionnel, pédagogique et scientifique. *Recherches en éducation, HS8*, 27- 38.
- Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, 36, 948-958.
- Catts, H. W., Fey, M., Zang, X. & Tomblin, B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, 3(4), 331-361.
- Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y., & Bontempo, D. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, 48(3), 281-297.
- Catts, H. W., Nielsen, D. C., Bridges, M. S. & Liu, Y.-S. (2016). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities*, 49(5), 451-465.

- Chew, A. L. (1981). *The Lollipop Test: A Diagnostic Screening Test of School Readiness*. Atlanta, GA: Humanics Limited.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (ed). *Strategies of Information Processing* (pp. 151-216). Londres : Academic Press.
- Desrochers, A., Laplante, L. et Brodeur, M. (2016). Le modèle de réponse à l'intervention et la prévention des difficultés d'apprentissage de la lecture au préscolaire et au primaire. In M.-F. Morin, D. Alamargot et C. Gonçalves (dir.), *Perspectives actuelles sur l'apprentissage de la lecture et de l'écriture / Contributions about learning to read and write* (p. 290-314). Sherbrooke : Éditions de l'Université de Sherbrooke.
- De Weck, G., & Marro, P. (2010). *Les troubles du langage chez l'enfant : description et évaluation*. Paris: Elsevier Masson.
- Duff, F. J., Reen, G., Plunkett, K. & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848-856.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4thed.). Circle Pines: American Guidance Service.
- Dunn, L. M., Theriault-Whalen, C. M., & Dunn, L. M. (1993). Adaptation française du peabody picture vocabulary test – révisée [The French adaptation of the peabody vocabulary scale - revised]. Toronto: Psycan.
- Engen, L., & Høien, T. (2002). Phonological skills and reading comprehension. *Reading and writing: an interdisciplinary Journal*, 15, 613-631.
- Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, 18, 129-155.
- Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments and Computers*, 35, 302–309.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of dyslexia*, 36, 69-81.

- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195-203.
- Furnes, B. & Samuelsson, S. (2011). Phonological Awareness and Rapid Automatized Naming Predicting Early Development in Reading and Spelling: Results from a Cross-Linguistic Longitudinal Study. *Learning and Individual Differences, 21*(1), 85-95.
- Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and Language*. HoveUK: Lawrence Erlbaum.
- Giasson, J. (1997). *La lecture : de la théorie à la pratique*. Bruxelles : De Boeck.
- Giasson, J., Vandecasteele, G. (2012). *La lecture : apprentissage et difficultés* (1^e éd.). Bruxelles, Belgique: De Boeck.
- Gough, P. & Tunmer, W. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6-10.
- He, Q., Xue, G., Chen, C., Chen, C., Lu, Z. L., & Dong, Q. (2013). Decoding the neuroanatomical basis of reading ability: a multivoxel morphometric study. *The Journal of Neuroscience, 33*(31), 12835-12843
- Hulme, C., Browyer-Crane, C., Carroll, J. M., Duff, F J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572-577.
- Hulmes, C. & Snowling, M. J. (2013). The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society, 369*:20120395.
- Jaffré, J.-P. (1991). Compétence orthographique et système d'écriture. *Repères, 4*, 35-48.
- Kamhi, A., & Catts, H. (Eds.). (2012). *Language and reading disabilities* (3rd ed.). Boston, MA: Allyn & Bacon.
- Lieberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

- Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201–212.
- Lee, J. (2011). Size matters: early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, 32, 69-92.
- Lervag, A., & Hulme, C. (2009). Rapid Automatized Naming (RAN) Taps a Mechanism That Places Constraints on the Development of Early Reading Fluency. *Psychological Science*, 20(8), 1040-1048.
- Maltais, C., & Herry, Y. (1997). Le concept de soi des élèves éprouvant des difficultés d'apprentissage. *Éducation et Francophonie*, 25(2), 64-76.
- Melby-Lervag, M., Halaas Lyster, S.-A., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *American Psychological Association*, 138(2), 322-352.
- McArthur G. M., Hogben J. H., Edwards V. T., Heath S. M. & Mengler E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41, 869–874.
- Muter, V., Hulme, C., Snowling, M. J. & Stevenson, J. (2004). Phoneme, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, 40(5), 665-681.
- Norton, E. S. & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, 63, 427-452.
- Ouellette, G. P. (2006). What's Meaning got to do with it : the role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 3, 554-566.
- Ouzoulias, A. (2004). *Favoriser la réussite en lecture : les maclé. Modules d'approfondissement des compétences en lecture-écriture*. Paris : Retz-Scéren.
- Pech-Georgel, C., George, F. (2006). *BELO : Batterie d'évaluation de la lecture et de l'orthographe*. Marseille : Solal.

- Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235-257.
- Roth, F. P., Speece, D. L. & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The journal of Educational Research, 95*(5), 259-272.
- Rvachew, S. & Savage, R. (2006). Preschool foundations of early reading acquisition. *Paediatrics & Child Health, 11*(9), 589-93.
- Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: contributions of phonemic awareness, verbal memory, rapid naming and IQ. *Annals of dyslexia, 48*, 115-136.
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook for research in early literacy* (pp. 97-110). New York: Guilford Press.
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*, 265-282.
- Schoenhals, L., Zesiger, P., Mounir, D.G., & Friend, M. (2009). *Can we assess infants' comprehension in an objective way?* Presented at the 7th European Congress of Speech and Language Therapists and Logopedists, Ljubjana, Slovenia.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.
- Share, D.L. (1995). Phonological recoding and self-teaching : Sine qua non of reading acquisition. *Cognition, 55*, 151-218.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143-174.
- Sprenger-Charolles, L. & Colé, P. (2006). *Lecture et dyslexie*. Paris : Dunod.

- Sprenger-Charolles, L., Colé, P., Kipffer-Piquard, A., & Leloup, G. (2010). EVALEC, *Une batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture*. Ortho-Edition.
- Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling: A four-year longitudinal study. *Journal of experimental child psychology*, 84(3), 194-217.
- Suortti, O., & Lipponen, L. (2016). Phonological Awareness and Emerging Reading Skills of Two- to Five-Year-Old Children. *Early Child Development and Care*, 186(11), 1703-1721.
- Treillet, E. (2004). *Dyslexie développementale: déficit d'automatisation, de conscience phonologique ou de traitement des lettres*. Mémoire d'orthophonie.
- Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability : multivariate evidence for a convergent skills model of reading development. *Scientific studies of reading*, 11(1), 3-32.
- Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research*, 50, 1093-1109.
- Zagar, D. (1992). L'approche cognitive de la lecture : de l'accès au lexique au calcul syntaxique. In M. Fayol, J.E. Gombert, P. Lecocq, L. Sprenger-Charolles et D. Zagar (Eds), *Psychologie cognitive de la lecture* (pp. 15-72). Paris : PUF.
- Zesiger, P. E., Brun, M. & Patrucco-Nanchen, T. (2004). Les relations entre l'oral et l'écrit dans l'acquisition du langage. *Approche neuropsychologique des apprentissages chez l'enfant*, 76-77, 80-85.
- 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), 395-421.

COMPREHENSION

Questions	Réponse correcte ✓ ou fausse ✗	Score
1. Quelle est la couleur de la petite poule ? (blanche)		Compréhension littérale
2. Quelle est la couleur de la grande poule ? (noire)		
3. Qui est tombé en dernier dans la mare ? (la grande poule noire)		
4. Quel objet important a vu la petite poule ? (un tronc d'arbre)		
5. Pourquoi un tronc d'arbre peut les sauver ? a. Parce que le tronc d'arbre est en bois b. <u>Parce que le tronc d'arbre peut servir de bateau</u> c. Parce que le tronc d'arbre est à côté de la mare		Inférences
6. Quel est le meilleur titre pour cette histoire ? a. Au bord de la mare b. Les poules mouillées c. <u>Le sauvetage des deux poules</u>		
7. Dans « La grande poule se précipite pour le secourir », quel est le sens de « se précipiter » ? a. Aller lentement b. Tomber dans un précipice c. <u>Aller rapidement</u>		Vocabulaire
8. Le rivage, c'est : a. Le marécage b. <u>Le bord de la mare</u> c. La rivière		

Annexe II : cahier de passation des épreuves de Lecture en Une Minute (LUM) de mots familiers et de non-mots

Lecture en Une minute – Fluence (mots familiers et non-mots)

Sujet # _____

DOT: ____/____/____ **DOB:** ____/____/____ **SEX:** ____ **AGE:** _____

Montrer les exemples et corriger si besoin. Durant l'épreuve, pas de correction.

/!\ Chronométrer après les exemples /!\

Si reste bloqué sur un mot, lui demander de continuer après 3 secondes.

Après une minute, mettre un crochet (]) après le dernier mot lu, et lui demander de s'arrêter.

Consigne pour l'enfant

Regarde cette feuille, il y a des mots écrits. Tu vas lire tous les mots que tu peux lire [montrer le premier exemple] : comment se lit ce mot ['ta'] ?

Si répond correctement, dire « très bien », s'il n'a pas répondu ou s'il a fait faux, lire le mot à sa place.

Faire de même avec le second exemple.

Maintenant, tu vas travailler tout seul et tu vas lire le mieux que tu peux tous les mots [ne pas oublier le chronomètre].

Mots familiers

Essai : « ta » et « bol »					Total	Mots avec graphème contextuel	Mots irréguliers
Début chronomètre (rappel : arrêt après une minute)							
à	où	la	au	tu	/5		
un	il	été	on	mur	/5		
ni	sur	qui	vélo	par	/5		
feu	ce	peur	ami	moto	/5	/1	
peau	lune	gare	lire	bon	/5	/1	
mardi	col	avril	roi	faire	/5	/1	
facile	cheval	vrai	ligne	porte	/5	/1	
autre	loup	soir	page	raisin	/5	/2	/1
car	sucre	chat	matin	trésor	/5	/2	/1
soixante	lundi	rose	visage	six	/5	/2	/2
ciseau	août	pays	balais	filles	/5	/1	/3
sept	lourd	femme	garage	hibou	/5	/1	/3
1. Nbr items correctement lus en 1 minute					/60	/12	/10
2. Si terminé en moins d'1 min, temps exact en sec.					sec.		
3. Nbr items correctement lus en 1 min. = score ligne 1 divisé par le temps exact (sec.) multiplié par 60							

Mots inventés

Essai : « ti » et « buk »					Total	Items avec graphème contextuel	Items avec digraphes
Début chronomètre (rappel : arrêt après une minute)							
o	i	bi	ul	til	/5		
ja	ol	ata	dik	nar	/5		
vaf	zou	dul	lévo	tur	/5		/1
veur	co	noil	opa	jal	/5	/1	/2
neau	timé	cal	tibe	lupe	/5	/1	/1
pirde	gor	ablir	sar	vaïpe	/5	/1	/1
cifale	sapir	vro	rigne	rosan	/5	/2	/2
aubre	talpe	gur	louré	ragife	/5	/2	/2
noge	tope	uché	truson	ranli	/5	/2	/3
tosir	poitade	ousir	bapre	gigne	/5	/3	/3
gésin	kraulir	cibe	maïpe	ligase	/5	/3	/3
gofi	cïpre	tasin	cïpé	punse	/5	/5	/2
1. Nbr items correctement lus en 1 minute					/60	/20	/20
2. Si terminé en moins d'1 min, temps exact en sec.							sec.
3. Nbr items correctement lus en 1 min. = score ligne 1 divisé par le temps exact (sec.) multiplié par 60							

Annexe III : cahier de passation de l'épreuve de lecture de mots irréguliers**Lecture 12 mots irréguliers courts de la BELO**

Sujet # _____

DOT: ____/____/____ **DOB:** ____/____/____ **SEX:** ____ **AGE:** _____**Consigne pour l'enfant :***« Je vais te présenter des mots, et tu vas me les lire à haute voix. Tu es prêt(e) ? On commence. »*

/!\ Chronométrer /!\

	Transcription	Définition de l'enfant
Coup		
Outil		
Monsieur		
Village		
Porc		
Femme		
Blanc		
Bord		
Hiver		
Ours		
Sept		
Fils		
NOTE	/12	/12 Temps :

Annexe IV : tableau complet des analyses de corrélations entre toutes les VIs et les VD

Corrélations

		CCT	CP	Conn_Lettres	RAN	EVIP_Lex	MCTp	Score_déc	Score_comp
CCT	Corrélation de Pearson	1	.159	.195	-.379*	.336*	.066	.366*	.562**
	Sig. (bilatérale)		.301	.205	.011	.026	.670	.015	.000
	N	44	44	44	44	44	44	44	44
CP	Corrélation de Pearson	.159	1	.348*	.000	.201	.128	.323*	.348*
	Sig. (bilatérale)	.301		.021	1.000	.190	.408	.033	.021
	N	44	44	44	44	44	44	44	44
Conn_Lettres	Corrélation de Pearson	.195	.348*	1	-.102	.157	.237	.596**	.235
	Sig. (bilatérale)	.205	.021		.511	.310	.121	.000	.125
	N	44	44	44	44	44	44	44	44
RAN	Corrélation de Pearson	-.379*	.000	-.102	1	-.378*	-.237	-.400**	-.283
	Sig. (bilatérale)	.011	1.000	.511		.011	.122	.007	.063
	N	44	44	44	44	44	44	44	44
EVIP_Lex	Corrélation de Pearson	.336*	.201	.157	-.378*	1	.020	.409**	.283
	Sig. (bilatérale)	.026	.190	.310	.011		.895	.006	.063
	N	44	44	44	44	44	44	44	44
MCTp	Corrélation de Pearson	.066	.128	.237	-.237	.020	1	.369*	.259
	Sig. (bilatérale)	.670	.408	.121	.122	.895		.014	.090
	N	44	44	44	44	44	44	44	44
Score_déc	Corrélation de Pearson	.366*	.323*	.596**	-.400**	.409**	.369*	1	.430**
	Sig. (bilatérale)	.015	.033	.000	.007	.006	.014		.004
	N	44	44	44	44	44	44	44	44
Score_comp	Corrélation de Pearson	.562**	.348*	.235	-.283	.283	.259	.430**	1
	Sig. (bilatérale)	.000	.021	.125	.063	.063	.090	.004	
	N	44	44	44	44	44	44	44	44

*. La corrélation est significative au niveau 0.05 (bilatéral).

**. La corrélation est significative au niveau 0.01 (bilatéral).