

DEPARTEMENT ORTHOPHONIE
FACULTE DE MEDECINE
Pôle Formation
59045 LILLE CEDEX
Tél : 03 20 62 76 18
departement-orthophonie@univ-lille.fr



MEMOIRE

En vue de l'obtention du Certificat de Capacité d'Orthophoniste
présenté par

Laure MECQUINION

qui sera soutenu publiquement en juin 2022

**Impact du niveau de vocabulaire sur l'identification de
mots écrits chez les enfants allophones.**

MEMOIRE dirigé par

Matthieu BIGNON, Orthophoniste et doctorant, SCALab UMR 9193, Lille

Sandrine MEJIAS, Maître de conférences, SCALab UMR 9193, Lille

Lille – 2022

Remerciements

Je voudrais tout d'abord remercier les membres du jury pour la lecture de ce mémoire et son évaluation. Je remercie mes directeurs de mémoire, Mme Mejias et M. Bignon, pour leur grande disponibilité, leurs précieux conseils et leurs encouragements durant ces deux années de travail. Merci aux écoles de m'avoir accueillie, aux enseignantes d'UPE2A de Dunkerque pour leur aide et leur enthousiasme et enfin merci à tous ces élèves rencontrés pour leur participation et leur bonne humeur.

Je tiens également à remercier mes maîtres de stage Agathe, Laurence et Rachel, qui m'ont conseillée et aidée à donner le meilleur de moi-même durant cette dernière année d'étude.

Plus personnellement, je remercie mes parents, ma sœur, Grégoire, ma grande et belle famille ainsi que mes amis de toujours pour m'avoir soutenue et motivée sans relâche depuis la préparation des concours d'entrée en école jusqu'à ce jour. Enfin, merci à mes amies de promotion Louise, Amandine et Joséphine pour ces cinq belles années.

Résumé :

Les enfants allophones sont des enfants qui n'ont jamais été en contact avec la langue française orale et écrite avant leur arrivée en France. Des classes spécialisées accueillent ces élèves afin de répondre à l'enjeu majeur que représente l'apprentissage de notre langue et cet enseignement est primordial pour leurs apprentissages scolaires futurs. Peu d'articles scientifiques décrivent les compétences nécessaires au développement du langage écrit dans le cadre de l'allophonie. Cette étude a pour objectif d'évaluer et de déterminer le poids des différents prédicteurs cognitifs et linguistiques essentiels au développement de la lecture chez les enfants allophones et de les comparer au poids des prédicteurs des enfants tout-venant en émettant l'hypothèse que le rôle du vocabulaire sera plus important chez les enfants allophones que chez les enfants tout-venant. Dans le cadre de cette recherche, 186 enfants tout-venant et 115 enfants allophones ont été rencontrés individuellement en milieu scolaire afin de réaliser des épreuves évaluant les prédicteurs cognitifs et langagiers et les habiletés en lecture. Ces résultats montrent que tous les prédicteurs expliquent la lecture de mots chez les enfants allophones sauf la répétition de pseudo-mots et que le vocabulaire est mieux corrélé à la lecture chez les enfants allophones que chez les enfants tout-venant. Cependant, le vocabulaire et la lecture de mots ne sont pas corrélés significativement chez les enfants tout-venant ce qui pourrait s'expliquer par la transparence de la langue française qui nécessite un recours moins important au stock lexical pour corriger les erreurs de lecture.

Mots-clés :

Allophonie, langue seconde, langage écrit, prédicteurs, vocabulaire

Abstract :

Allophone children are children who have never been in contact with the oral and written French language before their arrival in France. Specialized classes welcome these children in order to meet the major challenge of learning our language and this teaching is essential for their future schooling. Few scientific articles describe the skills necessary for the development of written language in the context of allophonia. The aim of this study is to evaluate and determine the weight of the different cognitive and linguistic predictors essential to the development of reading in allophone children and to compare them to the weight of the predictors of mainstream children by hypothesising that the role of vocabulary will be more important for allophone children than for mainstream children. In the context of this research, 186 mainstream children and 115 allophone children were individually interviewed in a school setting to perform cognitive and language tests. These results show that all predictors explain word reading in allophone children except for the repetition of pseudowords and that vocabulary is better correlated with reading in allophone children than in mainstream children. However, vocabulary and word reading were not significantly correlated in mainstream children, which could be explained by the transparency of the French language, which requires less recourse to the lexical stock to correct reading errors.

Keywords :

Allophony, second language, written language, predictors, vocabulary

Table des matières

Introduction	1
Contexte théorique, buts et hypothèses	2
1. La lecture.....	2
1.1. Les deux composantes de la lecture.....	2
1.1.1. L'identification des mots écrits	2
1.1.2. La compréhension.....	3
1.2. Les prédicteurs de la lecture	4
1.2.1. La conscience phonologique	4
1.2.2. La mémoire à court terme.....	5
1.2.3. La dénomination rapide automatisée	6
1.2.4. Le vocabulaire	6
2. Compétences en lecture dans le cadre de l'allophonie	7
2.1. Identification des mots écrits	8
2.2. Qualité de prédiction des compétences cognitives et linguistiques étudiées.....	8
3. Buts et hypothèses.....	10
Méthode	10
1. Population.....	11
2. Procédure	11
3. Matériel et collecte des données.....	12
3.1. Analyse phonologique, EVALEC (Sprenger-Charolles et al., 2010).....	12
3.2. Dénomination rapide automatisée (Plaza & Robert-Jahier, 2006)	13
3.3. Répétition de pseudo-mots (Korkman et al., 2012).....	13
3.4. Désignation d'images, EVIP (Dunn et al., 1993)	13
3.5. Lecture de mots et de non-mots, EVALEC (Sprenger-Charolles et al., 2010)	14
4. Analyse des données	14
Résultats	14
1. Présentation des scores obtenus aux différentes épreuves.....	15
1.1. Prédicteurs cognitifs et linguistiques	15
1.2. Épreuves de lecture de mots	16
2. Corrélations entre les variables étudiées	16
2.1. Enfants tout-venant.....	17
2.2. Enfants allophones.....	17
3. Contribution des prédicteurs cognitifs et linguistiques sur la réussite en lecture	21
3.1. Réussite en lecture des enfants tout-venant.....	21
3.2. Réussite en lecture des enfants allophones.....	22
3.2.1. Réussite en lecture de mots réguliers.....	22
3.2.2. Réussite en lecture de mots irréguliers	23
Discussion	24
1. Rappel des objectifs du mémoire	24
2. Interprétation des résultats.....	24
3. Limites de l'étude.....	26

Conclusion	27
Bibliographie.....	28
Liste des annexes.....	32
Annexe n°1 : Lettre d'information aux parents ou tuteurs légaux	32
Annexe n°2 : Lettre d'information aux enfants à lire par leurs parents ou tuteurs légaux	32
Annexe n°3 : Consentement de participation libre et éclairée.....	32
Annexe n°4 : Questionnaire (version destinée aux titulaires de l'autorité parentale et/ou aux tuteurs légaux des enfants allophones)	32
Annexe n°5 : Questionnaire (version destinée aux titulaires de l'autorité parentale et/ou tuteurs légaux des enfants tout-venant)	32
Annexe n°6 : Description des épreuves utilisées.....	32

Introduction

Les élèves allophones nouvellement arrivés (EANA) sont des enfants qui n'ont pas le français comme langue première et qui sont arrivés en France depuis moins de deux ans (Armagnague & Rigoni, 2018). Ils sont alors confrontés à l'apprentissage simultané du français oral et écrit à l'école. Le terme « allophonie » permet de supprimer la connotation négative que portait la terminologie « élèves non-francophones » et qui reflétait un manque, une incapacité. L'allophone maîtrise en effet une ou plusieurs autres langues qui ne sont pas la langue de scolarisation du pays d'arrivée. La scolarisation des EANA est encadrée par les Centres académiques pour la scolarisation des enfants allophones nouvellement arrivés et des enfants issus de familles itinérantes et de voyageurs (CASNAV). En 2018-2019, 67909 « élèves nouvellement arrivés » avec une scolarité plus ou moins régulière avant leur arrivée ont été enregistrés dans les données du CASNAV. Depuis la nouvelle circulaire de 2012 publiée par le ministre de l'Éducation Nationale, le CASNAV répartit ces élèves au sein d'unités pédagogiques qui ont pour objectif l'apprentissage et la maîtrise de la langue française. La scolarisation dans ces Unités pédagogiques pour élèves allophones arrivant (UPE2A) a une durée maximale de deux ans. Les UPE2A se basent sur le profil scolaire et le niveau de maîtrise du français de l'élève pour déterminer les temps d'inclusion en classe ordinaire et l'accompagnement par un enseignant personnalisé du français langue seconde (FLE) nécessaires.

Durant cette scolarisation, l'acquisition du langage écrit constituera un enjeu de taille puisqu'il sera ensuite la base de tous les apprentissages scolaires de ces élèves. De plus, au moment d'aborder l'écrit, l'enfant monolingue a pu s'imprégner du système phonologique de la langue, acquérir un lexique étendu ainsi que des structures syntaxiques précises. Les données actuelles indiquent que les compétences orales préexistantes jouent un rôle déterminant sur l'apprentissage du langage écrit (Content & Zesiger, 1999). Or pour l'enfant allophone, le français oral ne pourra pas soutenir aussi fortement l'acquisition du langage écrit. En effet, il présente de faibles compétences dans la langue de scolarisation au moment de sa rencontre avec l'écrit. C'est la raison pour laquelle nous nous intéresserons dans cette étude à l'acquisition du langage écrit des enfants allophones et plus particulièrement à leurs compétences en identification de mots écrits. Nous observerons et comparerons le poids des prédicteurs cognitifs et linguistiques des compétences en lecture chez les enfants allophones et les enfants monolingues. Nous chercherons à comprendre lesquels de ces prédicteurs influencent le plus les compétences en identification de mots écrits chez ces enfants en portant une attention particulière au rôle du vocabulaire. Pour cela, nous effectuerons une étude observationnelle multicentrique pour évaluer les différentes compétences linguistiques orales et écrites de ces enfants.

Nous commencerons par expliquer le contexte théorique dans lequel s'inscrit cette étude afin de décrire dans un premier temps les compétences en lecture chez des enfants « tout-venant » apprenant le français écrit. Nous verrons ensuite les particularités de l'apprentissage du langage écrit des enfants allophones. Nous détaillerons la méthodologie ainsi que les résultats observés. Pour finir, nous mènerons une discussion concernant les résultats obtenus puis nous conclurons.

Contexte théorique, buts et hypothèses

Nous étudierons dans un premier temps les composantes essentielles pour une lecture réussie ainsi que les compétences cognitives et linguistiques sous-jacentes nécessaires à l'entrée dans l'écrit pour tout enfant apprenant à lire dans un système orthographique alphabétique. Nous analyserons ensuite les caractéristiques de l'apprentissage de la lecture pour les enfants allophones ainsi que l'influence des prédicteurs cognitifs et linguistiques présentés cette fois dans le cadre de l'allophonie.

1. La lecture

Nous nous intéresserons ici particulièrement aux compétences impliquées dans la lecture puis aux prérequis essentiels à l'entrée dans l'écrit.

1.1. Les deux composantes de la lecture

Selon le modèle simple de la lecture de Gough et Tunner proposé en 1986, la compréhension en lecture est liée aux deux facteurs suivants : l'identification de mots écrits et la compréhension du langage. Ce modèle est présenté sous forme d'un produit, soit $R = D \times LC$ (Reading = Decodage $\times$ Linguistic Comprehension) traduit en français tel que : $L = D \times LC$. Le « L », qui est le résultat de cette opération, correspond ainsi à la lecture dont l'objectif est la compréhension de ce qui est lu. Cette compréhension écrite dépend alors de « D » qui signifie « décodage », c'est-à-dire la maîtrise des procédures de lecture pour identifier les mots, et de « LC » qui est la « compréhension du langage ». Pour éviter toute confusion, nous utiliserons dans cette étude le terme « identification de mots écrits », qui est le processus de lecture, à la place du terme « décodage ».

Nous allons maintenant détailler chacune des composantes du modèle simple de lecture, soit l'identification de mots écrits et la compréhension du langage.

1.1.1. L'identification des mots écrits

L'identification de mots écrits nécessite un apprentissage complexe qui va mettre en jeu deux formes de traitement décrits dans le modèle à double voie de Coltheart (1978) : la procédure phonologique et la procédure lexicale.

La procédure phonologique permet le passage de l'information orthographique à une information phonologique grâce aux correspondances grapho-phonologiques (Valdois, 2010). Elle permet la lecture de mots réguliers et de pseudo-mots, ou encore de mots qui ne font pas partie du stock orthographique de l'individu.

La procédure lexicale permet d'identifier directement le mot écrit en activant sa représentation orthographique dans le lexique orthographique interne de l'individu. Elle est utilisée notamment pour les mots réguliers ou irréguliers connus qui sont présents dans le stock orthographique de l'individu. Ce dernier est enrichi grâce à un apprentissage explicite mais également grâce aux lectures personnelles de l'enfant. C'est le phénomène d'auto-apprentissage conceptualisé par Share en 1995 (Ginestet et al., 2020) qui permet un apprentissage de traces orthographiques de manière implicite.

Ces deux processus de traitement des mots écrits peuvent être évalués par une épreuve de lecture de mots et de pseudo-mots afin d'analyser chacune des voies de la lecture. En effet, la lecture de pseudo-mots nous permet de mettre en avant le fonctionnement de la procédure phonologique grâce à des mots que l'individu n'a jamais rencontrés et qu'il n'aura pas dans son stock orthographique. Il est alors obligé de décomposer le mot en faisant correspondre chaque graphème à un son et de fusionner ensuite mentalement ces sons. L'épreuve de lecture de mots contient à la fois des mots réguliers et irréguliers. Les mots réguliers sont des mots qui respectent les règles de conversion graphème-phonème (ex : « minute », « pilote) et qui seront identifiés soit via la procédure phonologique si le mot a été peu ou jamais rencontré à l'écrit, soit par la procédure lexicale si le mot est connu du lecteur et qu'il fait partie de son stock orthographique. Les mots irréguliers ne respectent pas ces règles de conversion grapho-phonologiques et l'utilisation de la procédure phonologique pour identifier ces mots, lus phonétiquement, donnerait lieu à des erreurs de régularisation. Il est ainsi nécessaire d'utiliser la procédure lexicale faisant appel aux connaissances orthographiques, c'est-à-dire aux règles d'usage du français (ex : « monsieur », « femme ») pour les identifier et les lire correctement.

L'identification de mots écrits est donc le résultat d'une utilisation simultanée de ces deux procédures dont l'influence de chacune varie au cours du développement des compétences en lecture. Cependant, la reconnaissance précise et rapide des mots est une condition nécessaire mais non suffisante pour une compréhension écrite efficace (Stanké, 2006).

1.1.2. La compréhension

La notion de compréhension associée au domaine de la lecture fait référence au sens que l'individu donne à ce qu'il lit grâce à l'identification de mots écrits. La compréhension écrite est fortement liée aux compétences orales comme l'indique le modèle simple de la lecture (Gough et Tunner, 1986). En effet, les compétences en compréhension mises en jeu à l'oral sont réinvesties au contact de textes écrits (Lecocq & Casalis, 1996). La composante « CL » ou « compréhension du langage » a été détaillée car elle implique des compétences linguistiques orales mais aussi certaines compétences propres à l'écrit. Goigoux et al. (2016) précisent que la compréhension d'un texte lu ne fait pas appel à l'ensemble des compétences langagières orales (compréhension du contexte d'énonciation, éléments non verbaux de la communication par exemple). Dans le même sens, cette compréhension induit la maîtrise de compétences spécifiques à l'écrit telles qu'un vocabulaire plus soutenu et une syntaxe différente de celle employée à l'oral. Il semble donc important de prendre en compte que le facteur « compréhension du langage » implique de nombreuses compétences sous-jacentes.

Par ailleurs, la capacité du sujet à comprendre un texte lu nécessite qu'il se crée une représentation mentale de ce qu'il est en train de décoder grâce aux deux voies de lecture. Cette représentation mentale dépend en grande partie de la « base de connaissances » du lecteur (Fayol, 1992). La base de connaissances est constituée, selon l'auteur, d'aspects linguistiques et conceptuels. En effet, il s'agit d'un bloc de connaissances concernant à la fois les concepts du monde qui l'entoure mais également des notions linguistiques nécessaires pour comprendre des phrases et intégrer du texte (Avanthey-Granges & Botteron, 2003). Ainsi, une base de notions conceptuelles large et étendue permet une meilleure compréhension des textes abordés. Goigoux et al. (2016) concluent leur étude en citant trois facteurs responsables de la compréhension écrite en fin de première année du primaire: « le décodage, la compréhension du langage (phrases ou textes entendus) et le vocabulaire. »

Enfin, il est important de préciser que l'impact de ces deux facteurs, identification de mots écrits et maîtrise de la langue orale, varie au cours de l'apprentissage de la lecture (Raudzus et al., 2021). Le rendement en compréhension de l'écrit apparaît en effet largement déterminé par les compétences en identification des mots écrits durant les premières années d'apprentissage alors qu'il repose beaucoup plus sur la maîtrise de la langue orale ensuite (Ouellette & Beers, 2010).

Nous allons maintenant étudier les compétences cognitives et linguistiques qui vont déterminer les progrès ultérieurs des enfants en lecture.

1.2. Les prédicteurs de la lecture

Dans ce mémoire, nous utiliserons le terme de « prédicteurs » pour renvoyer aux différents prérequis du langage écrit. Ce sont des compétences susceptibles de prédire un apprentissage de la lecture réussi. Negro et Genelo (2009) distinguent les prédicteurs de l'identification des mots écrits et les prédicteurs de la compréhension. Nous ne détaillerons pas ici les prédicteurs de la compréhension car nous focalisons notre recherche sur l'identification de mots écrits.

La reconnaissance des mots est impactée par la conscience phonologique (Negro & Genelo, 2009), la mémoire à court terme (Kim et al., 2018), la dénomination rapide automatisée (Lervag, 2009) ainsi que le vocabulaire oral (Ricketts et al., 2007). Ce sont ces prédicteurs que nous allons maintenant présenter.

1.2.1. La conscience phonologique

La conscience phonologique est la capacité de l'individu à manipuler des unités phonologiques de la langue orale (phonèmes, syllabes et rimes). Elle comprend la conscience phonémique, la conscience syllabique et la conscience de la rime. De plus, on distingue l'épiphonologie de la métaphonologie. L'épiphonologie est la manipulation inconsciente des unités de la langue. Elle est implicite et constitue un pré-requis à la métaphonologie. La métaphonologie correspond quant à elle à la prise de conscience des unités de la langue qui sont traitées. De plus, la conscience phonologique est étroitement liée aux les compétences mnésiques de l'individu et plus particulièrement l'empan phonologique de la mémoire à court terme (Demont & Botzung, 2003).

Cette capacité de traitement des unités phonologiques a un rôle déterminant au début de l'apprentissage de la lecture. En effet, des études ont révélé une corrélation longitudinale relativement forte entre la conscience phonologique et la capacité de lire des mots isolés chez des apprentis lecteurs du français et ont ainsi montré que la conscience phonologique constitue un prédicteur significatif de la réussite ultérieure en lecture (Desrochers et al., 2009).

Comme nous l'avons évoqué, la conscience phonologique se décompose en trois catégories : la conscience phonémique, la conscience syllabique et la conscience de la rime. Chacune peut être testée isolément grâce à des tâches spécifiques. La conscience phonémique, qui est le niveau d'abstraction de la conscience phonologique le plus élevé, est l'aptitude à extraire dans une syllabe les déterminants phonémiques de celle-ci. Les tâches les plus utilisées pour évaluer la conscience phonémique sont des exercices d'identification du phonème initial, de localisation phonémique ou encore de suppression du phonème initial. La conscience syllabique recouvre la capacité à découper les mots en syllabes et à les manipuler. Les tâches les plus utilisées pour l'évaluer sont des exercices d'identification syllabique, d'inversion syllabique ou encore de fusion syllabique. Pour finir, la rime constitue la deuxième partie de la syllabe. Elle complète l'attaque de la syllabe et se compose d'une

voyelle et des phonèmes qui suivent. L'épreuve la plus communément utilisée pour évaluer la conscience de la rime est celle de jugement de rimes.

Des études ont révélé que la conscience phonémique est le prédicteur le plus puissant des compétences en lecture et notamment la segmentation phonologique qui est liée au développement des compétences en lecture et en orthographe (Muter et al., 1998). C'est la raison pour laquelle une tâche de suppression de phonème initial a été choisie pour évaluer la conscience phonologique dans cette étude.

1.2.2. La mémoire à court terme

La mémoire à court terme ou « Short-term memory » (STM) permet de retenir une quantité d'informations limitée durant une courte durée. Cette mémoire est décrite dans le modèle de Baddeley et Hitch (1974) selon quatre composantes. Au centre de ce modèle se trouve l'administrateur central qui sélectionne, coordonne et contrôle les opérations de traitement. Selon Baddeley, l'administrateur central correspond à l'ensemble des processus attentionnels. Le calepin visuo-spatial est la deuxième procédure de traitement de la mémoire à court terme évoquée dans le modèle. Il permet la manipulation d'informations visuelles et spatiales. La boucle phonologique est la composante qui permet la rétention d'informations verbales ou prononçables (Ehrlich & Delafoy, 1990). Pour finir, le buffer épisodique est un processus qui a été ajouté au modèle par Baddeley en 2000. Il permet l'intégration en mémoire à long terme des informations provenant des deux systèmes précédents (boucle phonologique et calepin visuo-spatial).

La boucle phonologique est la composante la plus connue de ce modèle. Celle-ci est en réalité composée de deux processus sous-jacents : le stock phonologique permettant de stocker temporairement des informations verbales et la boucle de récapitulation articulatoire permettant d'actualiser les informations dans le stock phonologique (Majerus, 2010). C'est la raison pour laquelle les études évoquent le terme de « mémoire phonologique à court terme ». Des recherches ont montré une relation directe entre la mémoire phonologique à court terme et les compétences en lecture de mots (Kim et al., 2018). En effet, les auteurs expliquent ce lien par la nécessité de maintenir en mémoire, durant une courte période, les sons identifiés grâce à la conversion grapho-phonologique lors du processus d'identification de mots écrits. Ainsi, les sons peuvent ensuite être combinés pour récupérer la forme phonologique correcte du mot.

L'évaluation de la mémoire à court terme peut s'effectuer au moyen de tâches telles que des empans de chiffres, de mots et de séquences visuo-spatiales. Cependant, ce type de tâches ne permet pas d'évaluer l'un des sous-systèmes de la mémoire à court terme de manière spécifique (Majerus, 2010). C'est la raison pour laquelle certaines études évoquent la tâche de répétition de non-mots (ou pseudo-mots) pour évaluer la mémoire phonologique à court terme et plus précisément le stock phonologique qui est l'une des composantes de la boucle phonologique (Baddeley et al., 1998). En effet, cette épreuve consiste à répéter, immédiatement après présentation orale, des items inconnus et de longueur variable. Les pseudo-mots font alors appel uniquement au stock phonologique de l'individu puisqu'ils ne peuvent induire aucune représentation lexicale en mémoire à long terme comme pourraient le faire les chiffres et les mots grâce au buffer épisodique (Poncelet et Van der Linden, 2003). Ainsi, pour évaluer les compétences en mémoire à court terme, cette étude a ciblé une tâche de répétition de pseudo-mots qui est généralement moins bien réussie par les enfants dyslexiques que les enfants tout venant (Melby-Lervag, 2012).

1.2.3. La dénomination rapide automatisée

Le dernier prédicteur non langagier des compétences en langage écrit est la dénomination rapide automatisée ou RAN (Rapid Automatized Naming). Il s'agit de dénommer des symboles familiers présentés visuellement tels que des chiffres, des lettres, des couleurs ou encore des objets aussi rapidement que possible (Georgiou et al., 2013).

La littérature a établi que la tâche de dénomination rapide est un prédicteur important des compétences en lecture dans différentes orthographe. Plusieurs arguments théoriques se recouvrent pour expliquer le lien entre ces deux variables (Kirby et al., 2010). En effet, certains auteurs ont mis en avant la piste du traitement phonologique comme explication du lien entre lecture et dénomination rapide. Ces deux tâches nécessitent un accès rapide aux informations phonologiques de chaque item et une récupération de celles-ci en mémoire à long terme. D'autres auteurs expliquent le lien entre ces deux tâches par un traitement orthographique, c'est-à-dire une reconnaissance de chaque item comme unité indépendante qui ne nécessite pas de réaliser des conversions grapho-phonologiques. Un autre argument théorique avance que de façon plus globale, ces deux tâches sont liées par la vitesse générale de traitement, c'est-à-dire le résultat de la vitesse d'exécution des différents processus sous-jacents. Pour finir, d'autres auteurs lient ces compétences car elles font toutes deux appel à des compétences exécutives telles que l'inhibition (pour supprimer la prononciation des items précédents), la mémoire de travail (récupérer alternativement les bonnes représentations phonologiques en mémoire notamment pour la dénomination rapide où ce sont les mêmes items présentés à plusieurs reprises) ou encore l'attention. Ces différentes preuves se complètent sans se contredire puisque chacune fait référence à des processus sous-jacents nécessaires au mécanisme d'identification de mots écrits. Enfin, les études évoquées par Kirby et al. (2010) s'accordent pour définir un lien solide entre la dénomination rapide et la lecture et ce indépendamment des autres prédicteurs.

C'est la raison pour laquelle notre étude propose une tâche de dénomination rapide chronométrée basée sur la lecture de stimuli connus et redondants. Selon les auteurs, les performances en dénomination de stimuli alphanumériques (des chiffres et des lettres) seraient plus fortement corrélées aux capacités en lecture que la dénomination d'objets et de couleurs à partir de 6 ans (Norton et Wolf, 2012). Dans le cadre de notre étude, l'épreuve se présente ainsi sous la forme d'une grille de chiffres et d'une grille de lettres à dénommer rapidement.

1.2.4. Le vocabulaire

L'apprentissage de la lecture est également favorisé par le développement des habiletés langagières orales telles que le niveau de vocabulaire. Ouellette (2006) a établi deux dimensions du lexique. L'étendue du lexique est reliée à l'identification de mots écrits tandis que la qualité des représentations sémantiques apparaît mieux associée à la compréhension écrite. Le niveau de vocabulaire a donc un effet à la fois sur la capacité à identifier des mots écrits mais également sur les compétences en compréhension (Labat, 2013). Cependant, nous nous intéresserons ici plus particulièrement au lien entre le niveau de vocabulaire et la capacité à identifier des mots écrits.

Le niveau de vocabulaire peut être évalué de deux façons : en réception et en expression. Le vocabulaire réceptif est constitué de l'ensemble des mots que le sujet est en capacité de comprendre. Ces mots font ainsi partie de son stock lexical. Le vocabulaire en réception est évalué en général par une tâche de désignation d'images sur présentation d'un mot oral. Le vocabulaire en expression est

constitué des mots que le sujet peut produire. Il est évalué par une tâche de dénomination orale qui consiste à dire le mot correspondant à l'image présentée.

Plusieurs modèles théoriques permettent d'expliquer le poids du niveau de vocabulaire, et plus particulièrement de l'étendue du lexique, sur la lecture de mots (Wegener et al., 2018). En effet, les auteurs expliquent le lien entre vocabulaire et lecture par la capacité à prédire la représentation orthographique de mots connus à l'oral mais jamais rencontrés à l'écrit. Ils montrent que lorsqu'on connaît un mot à l'oral, on réalise une prédiction sur sa forme orthographique, ce qui facilite sa lecture lorsqu'il est rencontré à l'écrit pour la première fois. De plus, Tunmer et Chapman (2012) décrivent la compétence à générer des représentations phonologiques approximatives des mots inconnus qui se rapprochent de la forme phonologique correcte. L'enfant crée ainsi des prononciations alternatives du mot cible jusqu'à produire une prononciation correspondant à un mot dans la mémoire lexicale de l'individu. Leur étude a permis d'établir un lien entre cette compétence et le développement des compétences d'identification de mots écrits. Un vocabulaire oral étendu est ainsi un bon avantage lors des premières expositions au langage écrit. Ces deux hypothèses permettent d'expliquer le lien entre le vocabulaire et l'identification de mots écrits grâce aux représentations phonologiques des mots et à leur forme auditive présente en mémoire.

A cette contribution des représentations phonologiques, Laing et Hulme (1999) ajoutent également la notion de représentation sémantique qui correspond au sens du mot écrit. Les auteurs évoquent l'utilisation d'une voie phonologique et d'une voie sémantique par les jeunes lecteurs pour apprendre à lire des mots nouveaux. En effet, cette composante sémantique fait partie du modèle en triangle à traitement parallèle distribué (Seidenberg et McClelland, 1989). Ce modèle théorique de la lecture suggère que les représentations phonologiques, orthographiques et sémantiques sont interconnectées et propose deux voies pour identifier un mot, la voie phonologique et la voie sémantique. Ainsi, la voie sémantique permet, si l'individu ne parvient pas à retrouver la forme phonologique correcte du mot, de s'appuyer sur sa représentation orthographique pour accéder au sens et ainsi récupérer la forme auditive correspondante. Ce processus est particulièrement intéressant pour la lecture des mots irréguliers pour lesquels la forme orthographique ne correspond pas strictement à la forme phonologique puisque ces mots ne respectent pas les règles de conversion graphème-phonème.

Dans cette étude, nous avons fait le choix d'évaluer uniquement le vocabulaire réceptif grâce à une tâche de désignation d'images. Cette tâche nous permet ainsi d'avoir un aperçu des connaissances lexicales des enfants. La tâche de dénomination, évaluant le vocabulaire en production, risquerait de biaiser l'interprétation du lien entre l'étendue du vocabulaire et les compétences en lecture. En effet, la tâche de dénomination met en jeu les habiletés phonologiques en production de l'individu qui risquent d'être inférieures chez les enfants allophones par rapport aux enfants monolingues.

Nous allons maintenant développer les particularités de l'apprentissage de la lecture du français dans le cadre de l'allophonie.

2. Compétences en lecture dans le cadre de l'allophonie

Peu de recherches concernent l'allophonie et l'apprentissage du français écrit en tant que langue seconde. Nous avons donc pour objectif ici de décrire les particularités du développement du langage écrit dans une langue seconde, lorsque l'enfant est confronté simultanément au français oral

et écrit. Nous nous intéresserons plus particulièrement au développement des compétences en identification de mots écrits chez les enfants allophones ainsi qu'aux prédicteurs du langage écrit que nous évaluerons.

2.1. Identification des mots écrits

L'identification de mots écrits est la première étape nécessaire pour une lecture réussie. Le développement de l'identification de mots écrits dans une langue seconde implique que l'individu ait saisi le principe alphabétique, commun à toutes les langues alphabétiques. Il pourra ainsi comprendre et appliquer les correspondances nécessaires entre les graphèmes (lettre ou groupe de lettres) et les phonèmes (unités de parole). La compétence de déchiffrage acquise dans la langue maternelle sera ainsi bénéfique lors de l'apprentissage de la langue seconde (L2) (Nicot-Guillorel, 2012). Cependant, un apprentissage spécifique de nouvelles constructions graphophonologiques (phonèmes et graphèmes n'appartenant pas à la langue maternelle) sera également nécessaire pour que l'identification de mots s'automatise dans la langue seconde. Par la suite, le lecteur en L2 pourra s'imprégner de la forme orthographique des mots et reconnaître les mots fréquents grâce à la procédure lexicale dont nous avons parlé précédemment. C'est l'efficacité de cette reconnaissance orthographique qui va permettre au lecteur d'alléger le coût cognitif lié à l'identification de mots et de lire couramment. La charge cognitive pourra ainsi se reporter sur les processus nécessaires à la compréhension.

De plus, Geva et al. (2019) évoquent de bonnes compétences pour l'identification de mots écrits chez les enfants apprenant une L2 depuis plusieurs années même si leur maîtrise de la langue orale se développe encore. Ils ajoutent qu'une maîtrise précise et fluide de la lecture de mots peut être similaire chez les apprenants de L1 et de L2 avec le même niveau d'expérience de lecture en L2.

Notre étude cible cependant des enfants allophones, population plus spécifique que celle évoquée dans le cadre de l'étude de Geva et al. (2019) car pour ces enfants l'apprentissage de la L2 vient seulement de débuter et leur temps d'exposition à la langue française est par conséquent restreint. C'est la raison pour laquelle nous chercherons dans cette étude à explorer le développement des compétences en identification de mots écrits des enfants allophones grâce à l'évaluation des facteurs cognitifs et linguistiques sous-jacents à cette compétence. En effet, ces prédicteurs cognitifs et linguistiques jouent un rôle primordial dans l'acquisition de la lecture et leur évaluation permettra d'établir un lien entre chaque compétence sous-jacente et le niveau en identification de mots écrits des enfants allophones. Ainsi, il sera possible d'identifier les compétences cognitives et linguistiques ayant un poids prédictif plus important lors du développement de la lecture des enfants allophones.

Nous détaillerons par la suite les compétences préalables nécessaires à l'apprentissage de la lecture dans une L2.

2.2. Qualité de prédiction des compétences cognitives et linguistiques étudiées

Caravolas et al. (2012) ont conclu, grâce à leur étude longitudinale, que l'apprentissage de l'écrit dans n'importe quelle langue, à condition que celle-ci soit alphabétique, est conditionné par les prédicteurs suivants : conscience phonologique et dénomination rapide automatisée. Ces résultats suggèrent que les compétences cognitives qui sous-tendent le développement de la lecture seraient

universelles pour toutes les orthographes alphabétiques. A ces processus cognitifs sous-jacents, Geva et al. (2019) ajoutent la mémoire à court terme et précisent que ce sont ces compétences cognitives de base qui prédisent la lecture de mots, la compréhension en lecture ainsi que la fluidité en lecture et ce dans n'importe quelles langues.

C'est la raison pour laquelle nous nous attendons à ce que ces compétences (dénomination rapide automatisée, conscience phonologique et mémoire à court terme) impactent toutes la réussite future en lecture des enfants allophones. Il est cependant important de prendre en compte que le niveau de maîtrise du français pourra impacter les compétences dans certaines épreuves telles que la conscience phonologique ou la dénomination rapide automatisée. En effet, le résultat en conscience phonologique dépendra de la capacité de l'enfant à reproduire certains sons du français qui sera différente selon la langue maternelle des enfants. De même, les temps de réponse en dénomination rapide automatisée augmenteront si les items ne sont pas connus par l'enfant.

Ricketts et al. (2007) ont montré que la connaissance du vocabulaire oral profite à l'apprentissage de la lecture. Le vocabulaire apparaît alors comme un critère à prendre en compte lors de l'apprentissage d'une langue seconde. Par ailleurs, des études montrent que les enfants bilingues auront un niveau inférieur, et ce de manière constante, en ce qui concerne le vocabulaire réceptif et expressif par rapport aux enfants monolingues (Thordardottir, 2019). C'est la raison pour laquelle nous nous attendons à observer une corrélation importante entre le niveau de vocabulaire et les compétences en identification de mots écrits des enfants allophones.

Si nous reprenons la notion de prévisibilité orthographique des mots avancée par Wegener et al. (2018), l'enfant allophone ne sera pas en mesure de faire un transfert entre les informations phonologiques d'un mot entendu à l'oral au préalable et sa possible représentation orthographique puisqu'il n'aura pas été exposé à la langue française orale en amont. Il ne pourra alors pas prédire les représentations orthographiques de ces mots. De même en reprenant le concept exposé par Tunmer et Chapman (2012), l'enfant allophone ne sera pas en mesure de générer des représentations phonologiques approchant celle du mot cible écrit afin de sélectionner la représentation correspondant à un mot connu dans sa mémoire lexicale. Selon le modèle triangulaire de Seidenberg et McClelland (1989), la voie sémantique permettant d'accéder au sens du mot écrit ne pourra pas suppléer la voie phonologique notamment lors de la lecture de mots irréguliers chez les enfants allophones puisque ceux-ci n'ont pas encore créé suffisamment de liens entre les représentations orthographiques des mots en français et leur signification. Enfin, Ricketts et al. (2007) évoquent l'importance du vocabulaire oral en particulier lorsque l'orthographe du mot écrit est incohérente avec sa prononciation comme c'est le cas pour la lecture de mots irréguliers. Les enfants allophones ne pourront pas s'appuyer sur ce stock lexical oral qui est très pauvre en français étant donné leur faible exposition à la langue.

Pour ces différentes raisons, la contribution du vocabulaire sur les compétences en identification de mots écrits sera susceptible de varier entre le groupe d'enfants monolingues et celui des enfants allophones. Nous nous attendons également à ce que les enfants allophones aient plus de difficultés à identifier les mots irréguliers que les mots réguliers.

Nous nous sommes intéressés au développement des compétences en identification de mots écrits chez les enfants allophones et à l'impact du vocabulaire sur ces compétences. Nous allons maintenant développer notre objectif de recherche puis les hypothèses qui en découleront.

3. Buts et hypothèses

L'objectif de recherche de cette étude est de déterminer le poids des différents prédicteurs cognitifs et linguistiques sur les compétences en identification de mots écrits des enfants allophones.

Nous nous interrogeons plus particulièrement sur le lien entre les compétences d'identification de mots écrits des enfants allophones et leur niveau de vocabulaire. La contribution du vocabulaire aux compétences ultérieures en lecture est inconsistante selon les études et est expliquée par différents modèles théoriques évoqués précédemment. De plus, il est important de relever qu'au début de l'apprentissage de la lecture, le vocabulaire employé dans les phrases et dans les textes est un vocabulaire fréquent et relativement simple. Cela est observable dans la littérature jeunesse mais également dans des épreuves d'évaluation plus spécifiques telles que les épreuves orthophoniques. Ce vocabulaire fréquent est ainsi intégré dans le stock lexical des enfants tout-venant puisque ce sont des mots qu'ils entendent depuis leur naissance. Nous supposons ainsi que pour ces enfants, le niveau de décodage dépendra moins du stock lexical oral qui est connu et maîtrisé, que d'autres prédicteurs tels que la mémoire à court terme, la dénomination rapide ou la conscience phonologique qui seront plus susceptibles de faire varier les résultats en identification de mots écrits.

En revanche, les enfants allophones ne possèdent pas ce stock lexical de mots fréquents puisqu'ils n'ont pas entendu ces mots appartenant à la langue française depuis leur naissance. Le vocabulaire simple et fréquent utilisé dans les épreuves est donc peu connu ou très partiellement. C'est la raison pour laquelle nous nous attendons à ce que le niveau de vocabulaire soit mieux corrélé aux compétences en identification de mots que pour les enfants tout-venant. La vitesse de traitement ainsi que la précision de la lecture peuvent ainsi varier puisque les mots identifiés ne déclencheront pas ou peu les représentations lexicales que l'enfant a en mémoire. Dans ce sens, nous pensons également observer des différences de résultats au sein du groupe d'enfants allophones entre la lecture de mots réguliers et la lecture de mots irréguliers avec une corrélation plus importante entre le vocabulaire et la lecture de mots irréguliers. En effet, les mots réguliers peuvent être identifiés par le biais de la procédure phonologique sans faire appel au stock lexical de l'individu.

Nous émettons ainsi l'hypothèse que le niveau de vocabulaire sera plus fortement corrélé aux compétences d'identification de mots écrits pour les enfants allophones que pour les enfants monolingues. Nous avançons également que les résultats en lecture de mots irréguliers seront mieux corrélés au niveau de vocabulaire que ceux en lecture de mots réguliers chez les enfants allophones.

Méthode

Cette étude s'intègre dans le projet TANMALL (Tâche d'Apprentissage de Nouveaux Mots chez des enfants ALLOphones) porté par M. Bignon, doctorant en psychologie cognitive. Il s'agit d'une étude observationnelle multicentrique menée en collaboration avec le CASNAV de Lille, Amiens et Nancy-Metz. Ce projet a reçu un avis favorable du comité d'éthique de Lille n° 2021-466-S90 en février 2021.

Nous décrirons ici la population concernée, la procédure mise en place ainsi que le matériel utilisé dans le cadre de cette étude.

1. Population

Dans le cadre de cette étude, 115 enfants monolingues, 71 enfants bilingues ainsi que 115 enfants allophones ont été rencontrés. Les élèves monolingues et bilingues scolarisés en CP ou en CE1 formaient le groupe contrôle. Les enfants allophones étaient scolarisés du CP au CM2 afin d'obtenir un échantillon assez important d'élèves allophones. Les questionnaires d'informations transmis aux parents nous ont permis de déterminer la moyenne du temps d'exposition à l'écrit, la moyenne du temps de séjour en France ainsi que le terrain de scolarisation pour les deux groupes d'enfants. Un troisième échantillon d'enfants bilingues a été créé afin de respecter la diversité des statuts linguistiques rencontrés et de déterminer s'il existe des différences de résultats aux épreuves entre ces groupes. Ces données concernant les trois échantillons d'élèves apparaissent dans le Tableau 1 proposé ci-dessous. Certains critères pris en compte tels que l'âge, la classe ou le terrain de scolarisation montrent d'importantes différences entre les groupes d'enfants. Les comparaisons entre ces groupes pourraient être biaisées par ces critères.

Tableau 1.
Caractéristiques de l'échantillon.

	Monolingues			Bilingues			Allophones		
	N	M(ET)	min-max	N	M(ET)	min-max	N	M(ET)	min-max
Âge	114	85.29 (6.03)	77-99	71	87.1 (7.8)	75-122	115	112 (20.89)	75-191
Classe¹	115	1.26 (0.44)	1-2	71	1.42 (0.65)	1-5	115	2.84 (1.34)	1-5
Temps exposition à l'écrit	115	11 (3.96)	7.7-19.18	71	12.94 (6.06)	6.46-31.05	115	12.05 (7.33)	1.61-30.95
Terrain²	115	2.76 (0.49)	1-3	71	2.58 (0.58)	1-3	115	1.77 (0.84)	1-3
Temps de séjour	113	85.26 (6.05)	77-99	70	84.38 (14.55)	9.53-110	114	17.78 (13.42)	0.79-116

Note. ET : écart-type ; M : moyenne ; N : effectif

¹ 1 = CP, 2 = CE1, 3 = CE2, 4 = CM1, 5 = CM2

² 1 = REP+, 2 = REP, 3 = tout venant

Nous avons vu les caractéristiques propres aux trois échantillons d'élèves rencontrés. Nous allons maintenant décrire la procédure employée pour réaliser les passations.

2. Procédure

Une étude pilote a eu lieu de mars à juillet 2021 afin de tester le matériel auprès d'une vingtaine d'enfants monolingues ayant été sélectionnés de la même manière que les enfants pour l'étude observationnelle multicentrique. Sur ce même temps, nous avons commencé à contacter les Inspecteurs de l'Éducation Nationale (IEN) à la fin de l'année scolaire 2020-2021 ainsi qu'en septembre

2021 afin d'obtenir leur accord pour intervenir dans les circonscriptions dans lesquelles nous pourrions nous déplacer. Nous avons ensuite contacté les enseignants UPE2A ainsi que les écoles élémentaires correspondants aux circonscriptions des IEN afin de leur exposer l'étude et d'obtenir leur consentement pour intervenir auprès de leurs élèves.

Les familles des enfants allophones et monolingues ont alors reçu une lettre d'information (annexe n°1), des formulaires de consentement (annexe n°3) ainsi qu'un questionnaire (annexes n°4 et n°5) pour que nous puissions interpréter les données recueillies. Les enfants ont également reçu une lettre d'information qui leur a été lue par leurs parents ou tuteurs légaux (annexe n°2).

Après avoir reçu le consentement des parents, nous avons pu construire un planning de passations en accord avec la direction des établissements ainsi que les enseignants concernés afin de rencontrer les élèves. Les entretiens avec les élèves ont été conduits de manière individuelle dans une salle calme et isolée de chaque école durant le temps scolaire. Chaque élève a été vu durant deux périodes de 30 minutes afin d'effectuer tous les tests nécessaires.

Nous allons maintenant évoquer les outils utilisés lors des passations et le traitement des données effectué ensuite.

3. Matériel et collecte des données

Chaque élève s'est vu administrer 9 épreuves dans le cadre du projet TANMALL. Cependant, nous ne nous intéresserons qu'à cinq d'entre elles dans le cadre de notre question d'étude qui cible les prédicteurs du langage écrit et leur impact sur l'identification de mots. C'est pourquoi nous détaillerons la tâche d'analyse phonologique de l'EVALEC (Sprenger-Charolles et al., 2010), la tâche de dénomination rapide (Plaza & Robert-Jahier, 2006), la tâche de répétition de pseudo-mots (Korkman et al., 2012), la tâche de désignation d'images de l'EVIP (Dunn et al., 1993) ainsi que la lecture de mots et de pseudo-mots de l'EVALEC (Sprenger-Charolles et al., 2010). Ces épreuves ont été passées grâce à un ordinateur via le programme PsychoPy qui code et conserve les résultats de chaque passation. Pour certains tests, l'examineur devait également prendre en note les productions des élèves et enregistrer l'épreuve pour pouvoir les analyser et les coter par la suite. Ces outils sont détaillés dans l'annexe n°6.

Nous allons maintenant détailler chacune de ces épreuves ainsi que leurs modalités de passation.

3.1. Analyse phonologique, EVALEC (Sprenger-Charolles et al., 2010)

EVALEC est un logiciel d'évaluation des troubles spécifiques d'apprentissage de la lecture du CP à la 3^{ème}. La tâche d'analyse phonologique qui en provient permet d'évaluer la conscience phonologique de l'individu et plus particulièrement ses compétences en segmentation phonémique. Le participant a réalisé trois épreuves successives :

- Une tâche de suppression syllabique initiale contenant 2 essais et 10 items tels que « pajomi ». Le score maximum pouvant être obtenu sur cette liste est de 10.
- Une tâche de suppression de phonème initial avec une structure syllabique correspondant à CVC (Consonne-Voyelle-Consonne) contenant 2 essais et 12 items tels que « fur ». Le score maximum pouvant être obtenu sur cette liste est de 12.

- Une tâche de suppression de phonème initial avec une structure CCV (Consonne-Consonne-Voyelle) contenant 2 essais et 12 items tels que « gron ». Le score maximum total pouvant être obtenu est de 12.

Pour chaque liste de mots, l'examineur présente au préalable un exemple avec manipulation de jetons pour s'assurer de la compréhension de l'enfant. Il montre la manipulation à réaliser avec un premier item d'exemple puis demande à l'enfant de faire de même avec un deuxième item d'exemple. Enfin, un troisième exemple est demandé à l'enfant sans la manipulation des jetons. A chaque début de liste, deux items d'entraînement sont proposés.

Sur cette épreuve, l'enfant est enregistré. La cotation est informatique et manuelle. Un point est attribué par réponse correcte. Les mauvaises réponses sont transcrites. Les autocorrections sont acceptées et doivent être signalées par la transcription « AC ». Le nombre de points maximum pouvant être obtenu sur cette tâche est ainsi de 34.

3.2. Dénomination rapide automatisée (Plaza & Robert-Jahier, 2006)

Pour cette tâche, l'enfant doit dénommer aussi vite que possible un certain nombre de stimuli présentés de manière répétée et dans un ordre aléatoire.

La tâche est composée de deux grilles à dénommer. La première grille comporte une série de 4 items à dénommer tels que les chiffres 1, 2, 7 et 9. La deuxième grille comporte une série de 4 autres items à dénommer telles que les lettres B, A, U et C. Avant le début du test, l'examineur vérifie que l'enfant connaît chacun des items qui seront proposés. Une série d'essai avec quelques items est ainsi proposée pour les chiffres puis pour les lettres.

L'épreuve est enregistrée et chronométrée. La notation correspond au temps mis pour lire tous les chiffres puis toutes les lettres.

3.3. Répétition de pseudo-mots (Korkman et al., 2012)

Cette tâche de répétition de pseudo-mots a pour but de tester la mémoire à court terme de l'enfant qui est l'un des prédicteurs des compétences en identification de mots. Les pseudo-mots de ce test sont enregistrés et énoncés par l'ordinateur. L'épreuve contient 31 items tels que « fikuspa ». L'enfant doit répéter ce qu'il entend et est enregistré.

Concernant la notation, un point est accordé par pseudo-mots correctement répété. Les autocorrections sont acceptées. Les mauvaises réponses doivent être transcrites par l'examineur. Le score maximum pouvant être obtenu sur cette épreuve est de 31.

3.4. Désignation d'images, EVIP (Dunn et al., 1993)

L'échelle de vocabulaire en images Peabody (EVIP) a été créée par Dunn et al. en 1993. Elle a été adaptée en français du « Peabody Picture Vocabulary Test (PPVT) ». Cette échelle permet d'étudier le stock lexical en réception de sujets âgés de 2 ans 6 mois à 18 ans. L'EVIP est composée de 170 items, tels que « kangourou » classés par difficulté croissante.

Cette épreuve est implémentée sur le programme PsychoPy. Des étiquettes ont été posées au préalable sur les touches du clavier de l'ordinateur utilisé. Chaque étiquette est numérotée de 1 à 4. C'est l'enfant qui appuie sur la touche correspondant au numéro de l'image qu'il veut désigner. La

cotation est ainsi automatique et un point est attribué par item réussi. L'épreuve s'arrête au bout 5 erreurs consécutives. Le nombre d'items a été réduit à 50 afin de limiter le temps de passation. Le score maximum pouvant être obtenu sur cette épreuve est de 50.

3.5. Lecture de mots et de non-mots, EVALEC (Sprenger-Charolles et al., 2010)

La tâche de lecture à voix haute du logiciel EVALEC permet l'évaluation des procédures d'identification de mots écrits.

Pour cette épreuve, l'enfant doit lire à haute voix dans un premier temps une série de 48 mots tels que « montre » puis une série de 36 pseudo-mots tels que « énoure ». Avant de commencer, l'examineur prévient l'enfant qu'il devra d'abord lire des petits mots puis des mots qui n'existent pas. L'enfant est enregistré.

La cotation est manuelle. Les mauvaises réponses sont transcrites. Un point est attribué par item correctement lu. A la fin de la réponse du participant, l'examineur appuie sur une touche pour arrêter le chronomètre. Le score maximum pouvant être obtenu pour la lecture de mots est de 48 points et celui pour la lecture de pseudo-mots est de 36 points.

Nous avons décrit les épreuves et les modalités de passation sélectionnées dans le cadre de cette étude. Nous allons à présent expliquer la méthode utilisée pour analyser les données recueillies.

4. Analyse des données

Suite aux passations dans les écoles, nous avons encodé les données pour chaque épreuve dans un tableur. Concernant la cotation des épreuves, nous avons choisi d'effectuer une cotation souple afin de ne pas pénaliser les enfants allophones sur d'éventuelles déformations phonologiques liées aux caractéristiques typologiques de leur langue maternelle. Cette cotation nous permet ainsi d'évaluer uniquement les processus responsables des compétences évaluées.

Nous avons ensuite réalisé des tableaux contenant les moyennes et les écart-types des enfants monolingues, des enfants bilingues et des enfants allophones pour chaque épreuve. Par la suite, nous avons réalisé des matrices de corrélations pour les enfants tout-venant et allophones.

Nous avons effectué des régressions linéaires multiples en intégrant dans les modèles les variables corrélées de façon significative avec la lecture de mots afin d'obtenir la contribution unique de chaque variable sur les compétences en identification de mots écrits.

Pour les statistiques inférentielles, la significativité retenue est $p \leq .05$ comme cela est pratiqué dans la littérature.

Nous avons présenté la méthodologie de cette étude. Nous allons maintenant détailler les résultats obtenus grâce aux analyses statistiques.

Résultats

Les résultats obtenus suite aux différentes analyses statistiques seront présentés ci-dessous. Un état des lieux des scores obtenus aux différentes épreuves par les trois groupes d'enfants sera ainsi proposé dans un premier temps. Nous détaillerons ensuite les corrélations observées grâce aux matrices de corrélation réalisées chez les enfants tout-venant et chez les enfants allophones. Pour finir, nous évoquerons les données résultant des régressions linéaires multiples réalisées.

1. Présentation des scores obtenus aux différentes épreuves

Nous présenterons dans un premier temps les résultats obtenus aux épreuves évaluant les prédictors cognitifs et linguistiques puis nous détaillerons les scores obtenus aux tâches de lecture de mots par les deux groupes d'enfants.

1.1. Prédictors cognitifs et linguistiques

Le Tableau 2 présenté ci-dessous détaille les résultats obtenus par les enfants monolingues, bilingues et allophones aux épreuves de vocabulaire, de dénomination rapide automatisée, de conscience phonologique et de répétition de pseudo-mots.

Les moyennes des enfants monolingues sont supérieures à celles des enfants bilingues et des enfants allophones sur l'ensemble des épreuves y compris pour l'épreuve de dénomination rapide automatisée. Celle-ci est cotée selon le temps de réalisation de la tâche et les enfants monolingues sont ainsi plus rapides. Sur la tâche de vocabulaire, nous notons un écart plus important entre les résultats des enfants monolingues ($M = 30.97$, $ET = 6.49$) et ceux des enfants bilingues ($M = 27.48$, $ET = 6.55$). La moyenne des enfants allophones sur cette épreuve de vocabulaire est nettement inférieure ($M = 17.03$, $ET = 7.6$). Cependant, les moyennes des enfants bilingues restent très proches des moyennes obtenues par les enfants monolingues sur les autres épreuves. L'écart entre les moyennes est plus important sur l'ensemble des tâches pour les enfants allophones par rapport aux deux autres groupes.

Tableau 2.

Résultats aux différents prédictors cognitifs et linguistiques.

	Monolingues			Bilingues			Allophones		
	N	M (ET)	min-max	N	M (ET)	min-max	N	M (ET)	min-max
EVIP	115	30.97 (6.49)	14-45	69	27.48 (6.55)	10-39	115	17.03 (7.6)	0-31
DRA (vitesse)	114	32.57 (6.04)	19.53-48	68	35.04 (10.77)	20.5-79	114	35.55 (13.05)	17.5-95.5
CP	113	26.97 (5.6)	4-34	65	26.59 (5.23)	8-34	113	21.66 (9.56)	0-34
RPM	114	27.03 (3.21)	9-31	69	26.86 (3.1)	16-31	113	25.18 (4.68)	4-31

Note. CP : conscience phonologique ; DRA : dénomination rapide automatisés ; ET : écart-type ; EVIP : tâche de vocabulaire ; M : moyenne ; N : effectif ; RPM : répétition de pseudo-mots.

Les résultats nous montrent un écart important entre les moyennes obtenues par les enfants monolingues et bilingues et les moyennes obtenues par les enfants allophones. Nous allons maintenant comparer les moyennes obtenues aux épreuves de lecture de mots pour les trois groupes d'enfants.

1.2. Épreuves de lecture de mots

Les résultats obtenus aux épreuves de lecture sont présentés dans le Tableau 3. La tâche de lecture de mots a été divisée en trois indices : la lecture de mots réguliers, la lecture de mots irréguliers et la lecture de pseudo-mots. Les moyennes des enfants monolingues en lecture de mots sont supérieures à celles des enfants bilingues et allophones. En effet, nous retrouvons une moyenne de 40,35 ($ET = 6.49$) pour les enfants monolingues, une moyenne de 39,28 ($ET = 7.44$) pour les enfants bilingues et une moyenne de 25,77 ($ET = 12$) pour les enfants allophones. Cependant, comme pour les résultats concernant l'évaluation des prédicteurs cognitifs et linguistiques, l'écart entre les moyennes des enfants monolingues et celles des enfants bilingues est très faible et ce pour chaque tâche de lecture.

Tableau 3.

Scores aux épreuves de lecture.

	Monolingues			Bilingues			Allophones		
	N	M(ET)	min-max	N	M(ET)	min-max	N	M(ET)	min-max
LM (total)	114	40.35 (6.49)	21-48	69	39.28 (7.44)	20-48	113	25.77 (12)	0-46
LM 1	114	32.58 (3.94)	18-36	69	31.68 (4.68)	17-36	113	21.64 (9.6)	0-35
LM 2	114	7.77 (3.09)	0-12	69	7.59 (3.17)	0-12	113	4.13 (3.04)	0-11
LPM	112	27.18 (5.39)	10-36	68	26.93 (5.65)	11-34	111	18.41 (8.5)	0-32

Note. ET : écart-type ; LM : lecture mots ; LM 1 : lecture de mots réguliers ; LM 2 : lecture de mots irréguliers ; LPM : lecture de pseudo-mots

Suite aux comparaisons effectuées sur les épreuves des prédicteurs cognitifs et linguistiques et de la lecture de mots, nous constatons que les écarts entre les moyennes des enfants monolingues et des enfants bilingues ne sont pas très marqués. Dans ce sens, nous rassemblerons ces deux groupes d'enfants pour la suite des analyses afin d'augmenter la puissance statistique de nos résultats grâce à un échantillon plus important. Nous appellerons ce groupe ainsi créé « enfant tout-venant » afin de les différencier des enfants allophones.

Nous avons présenté les moyennes obtenues aux différentes épreuves par les différents groupes d'enfants. Nous allons maintenant effectuer des analyses statistiques en nous basant sur ces résultats.

2. Corrélations entre les variables étudiées

Afin de réaliser ces analyses, des corrélations de Pearson ont été effectuées entre les critères socio-démographiques renseignés grâce aux formulaires d'information, les prédicteurs cognitifs et linguistiques et les tâches de lecture de mots pour chaque groupe d'enfants tout-venant et allophones.

Nous décrirons dans un premier temps les corrélations retrouvées au sein de l'échantillon d'enfants tout-venant puis nous procéderons de la même façon pour le groupe d'enfants allophones.

2.1. Enfants tout-venant

La matrice de corrélation faite sur les données des enfants tout-venant est présentée dans le Tableau 4.

Nous nous intéressons plus particulièrement, dans le cadre de notre étude, aux corrélations entre les épreuves évaluant les prédicteurs cognitifs et linguistiques et les tâches de lecture de mots. Nous constatons dans un premier temps que les quatre épreuves évaluant ces prédicteurs sont toutes corrélées à la lecture de mots (score total). La conscience phonologique est corrélée positivement et modérément à la lecture de mots et cette corrélation est significative ($r = .52, p \leq .001$), la corrélation entre la lecture de mots et la répétition de pseudo-mots est positive, faible et significative ($r = .38, p \leq .001$) et la corrélation entre la lecture de mots et l'EVIP, qui est la tâche évaluant le vocabulaire en réception, est positive, faible et significative ($r = .27, p \leq .001$). Pour finir, la corrélation entre la lecture de mots et la dénomination rapide automatisée est négative, modérée et significative ($r = -.42, p \leq .001$).

Concernant les facteurs socio-démographiques, nous n'observons pas de corrélation significative entre les différentes épreuves et le temps d'exposition au langage écrit. L'âge est corrélé positivement mais faiblement avec la lecture de mots irréguliers et cette corrélation est significative ($r = .16, p \leq .05$) et il est également corrélé négativement et faiblement à la dénomination rapide automatisée et cette corrélation est significative ($r = -.17, p \leq .05$). Pour finir, le terrain de scolarisation est corrélé de façon significative à la dénomination rapide automatisée, à la lecture de mots et à la lecture de mots irréguliers.

Nous avons vu les corrélations établies par l'étude entre les différents prédicteurs cognitifs et linguistiques et la lecture de mots pour les enfants tout-venant. Nous allons maintenant réaliser ces observations sur la matrice de corrélation des enfants allophones.

2.2. Enfants allophones

La matrice de corrélation réalisée sur les données des enfants allophones est présentée dans le Tableau 5.

L'objectif de notre étude est de déterminer le poids des différents prédicteurs cognitifs et linguistiques sur les compétences en lecture de mots. C'est la raison pour laquelle nous allons observer dans un premier temps les corrélations existantes entre la tâche de lecture de mots et chacun de ces prédicteurs. La matrice de corrélation met en avant une corrélation positive forte et significative entre la lecture de mots et la conscience phonologique ($r = .75, p \leq .001$) ainsi qu'entre la lecture de mots et la répétition de pseudo-mots pour lesquels la corrélation est modérée ($r = .50, p \leq .001$) et entre la lecture de mots et l'EVIP pour lesquels la corrélation est faible ($r = .32, p \leq .001$). Comme pour les enfants tout-venant, la tâche de dénomination rapide automatisée est corrélée négativement et de façon modérée avec la tâche de lecture de mots et ce résultat est significatif ($r = -.66, p \leq .001$).

De plus, nous nous intéressons, dans le cadre de notre deuxième hypothèse, aux corrélations entre la tâche de vocabulaire en réception (EVIP) et les tâches de lecture de mots réguliers et irréguliers. Lecture de mots réguliers et vocabulaire sont corrélés positivement et ce score est

significatif ($r = .30, p \leq .05$). C'est également le cas pour la corrélation entre la lecture de mots irréguliers et le vocabulaire en réception ($r = .32, p \leq .001$).

Concernant les facteurs socio-démographiques, la scolarité antérieure ainsi que l'apprentissage de la lecture dans une autre langue sont corrélés significativement avec l'ensemble des épreuves proposées à l'exception de la tâche de vocabulaire. En revanche, la durée du séjour en France n'est corrélée significativement avec aucune des épreuves. La durée de scolarisation en France ainsi que la durée d'exposition au langage écrit sont corrélées de façon positive, faible et significative avec la tâche de vocabulaire. La tâche de vocabulaire est corrélée de façon positive, faible et significative avec la durée de scolarisation en France ($r = .22, p \leq .05$) ainsi qu'avec la durée d'exposition au langage écrit ($r = .26, p \leq .05$). Pour finir, le terrain de scolarisation est corrélé de façon positive, faible et significative avec la conscience phonologique ($r = .25, p \leq .05$) et la répétition de pseudo-mots ($r = .21, p \leq .05$).

Tableau 4.*Matrice de corrélation des enfants tout-venant.*

	1	2	3	4	5	6	7	8	9	10	11	12
1. Terrain	—											
2. Age	-.63***	—										
3. Exposition à l'écrit	-.77***	.81***	—									
4. Matrices de Raven	-.03	.13	.06	—								
5. CP	.01	-.11	-.14	.36***	—							
6. DRA	.23*	-.17*	-.14	-.19*	-.21*	—						
7. RPM	-.08	.03	.01	.13	.34***	-.32***	—					
8. EVIP	-.01	.07	-.05	.35***	.26***	-.14	.21*	—				
9. LM (total)	-.17*	.07	.05	.38***	.52***	-.42***	.37***	.28***	—			
10. LM 1	-.10	.00	-.01	.33***	.54***	-.41***	.35***	.28***	.95***	—		
11. LM 2	-.23*	.16*	.13	.38***	.41***	-.37***	.34***	.24*	.91***	.73***	—	
12. LPM	-.14	.05	.03	.29***	.37***	-.44***	.38***	.13	.78***	.78***	.66***	—

Note. CP : conscience phonologique ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire en réceptions ; LM : lecture mots ; LM 1 : lecture mots réguliers ; LM 2 : lecture mots irréguliers ; LPM : lecture pseudo-mots ; RPM : répétition pseudo-mots ;

* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$.

Tableau 5.*Matrice de corrélation des enfants allophones.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Scolarité antérieure	—															
2. Lecture antérieure	.54***	—														
3. Durée séjour en France	-.29*	-.02	—													
4. Scolarité en France	-.26*	-.21*	.53***	—												
5. Exposition à l'écrit	-.01	-.06	.35***	.79***	—											
6. Terrain	-.11	.12	.24*	.14	-.02	—										
7. Age	.31*	.29*	-.06	-.04	.25*	-.10	—									
8. Classe	.41***	.40***	-.02	.02	.34***	-.17	.64***	—								
9. Matrices de Raven	.34***	.45***	.10	-.03	.09	.13	.15	.33***	—							
10. CP	.49***	.59***	.07	.01	.08	.25*	.12	.33***	.55***	—						
11. DRA	-.49***	-.45***	-.01	.07	-.15	.06	-.35***	-.54***	-.41***	-.59***	—					
12. RPM	.48***	.44***	.07	.04	.16	.21*	.21*	.37***	.37***	.61***	-.4***	—				
13. EVIP	.05	.14	.11	.22*	.26*	-.01	.09	.09	.30*	.20*	-.20*	.11	—			
14. LM (total)	.39***	.61***	.11	.03	.14	.08	.19*	.39***	.63***	.75***	-.66***	.50***	.32***	—		
15. LM 1	.37***	.60***	.12	.05	.14	.09	.17	.35***	.60***	.76***	-.64***	.52***	.30*	.99***	—	
16. LM 2	.38***	.52***	.07	-.03	.12	.01	.21*	.44***	.58***	.55***	-.59***	.33***	.32***	.84***	.73***	—
17. Lecture PSM	.34***	.51***	.06	-.02	.08	-.06	.10	.28*	.51***	.63***	-.59***	.42***	.17	.88***	.89***	.67***

Note. CP : conscience phonologique ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire en réceptions ; Lecture antérieure : apprentissage de la lecture dans une autre langue ; LM : lecture mots ; LM 1 : lecture mots réguliers ; LM 2 : lecture mots irréguliers ; LPM : lecture pseudo-mots ; RPM : répétition pseudo-mots ; Séjour : période écoulée depuis l'arrivée en France (mois).

* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$.

Nous avons analysé les matrices de corrélation pour les deux échantillons. Nous nous intéressons désormais à déterminer la part prédictive de chacune de ces variables à la réussite en lecture de mots.

3. Contribution des prédicteurs cognitifs et linguistiques sur la réussite en lecture

Afin de voir la part unique de chaque variable dans la réussite en lecture de mots des deux groupes d'enfants, nous avons réalisé des régressions linéaires multiples pour chaque échantillon. Nous détaillerons dans un premier temps le tableau de régression linéaire des enfants tout-venant puis celui des enfants allophones.

3.1. Réussite en lecture des enfants tout-venant

Concernant les enfants tout-venant, les régressions linéaires permettent de faire ressortir la contribution unique de quatre variables prédisant la réussite en lecture de mots. Ces variables sont les matrices de Raven testant l'intelligence non-verbale, la dénomination rapide automatisée, la conscience phonologique et la répétition de pseudo-mots. En revanche, le pourcentage de variance expliquée par le vocabulaire (EVIP) est nul et la pente de régression n'est pas significative ($p = .13$). Ces régressions linéaires multiples sont représentées dans le Tableau 6 ci-dessous.

Tableau 6.

Modèle de régression linéaire multiple expliquant la réussite en lecture chez les enfants tout-venant.

Modèle	Variable	b	β	t	p	ΔR^2
$R^2 = 0.44$						
H ₀	(Intercept)	39.95	.00	77.36	< .001	
H ₁	(Intercept)	12.79	.00	2.53	.01	
	Matrices de Raven	.22	.17	2.63	.01	.02
	DRA	-.21	-.25	-4.16	< .001	.05
	EVIP	.10	.10	1.54	.13	.00
	CP	.37	.29	4.54	< .001	.08
	RPM	.59	.25	3.94	< .001	.04

Note. la valeur R^2 est la valeur ajustée ; CP : conscience phonologique ; b : coefficient brut ; β : coefficient bêta ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire ; p : significativité ; RPM : répétition de pseudo-mots ; t : test t de Student ; ΔR^2 : pourcentage de variance expliqué par le modèle.

La réussite en lecture de mots chez les enfants tout-venant peut ainsi être prédite par les matrices de Raven, la dénomination rapide automatisée, la conscience phonologique et la répétition de pseudo-mots.

Nous allons maintenant détailler les résultats obtenus aux régressions linéaires des enfants allophones.

3.2. Réussite en lecture des enfants allophones

Les régressions linéaires multiples portant sur les performances des enfants allophones mettent en évidence la contribution unique des matrices de Raven, de la dénomination rapide automatisée et de la conscience phonologique. De plus, le vocabulaire contribue également de façon significative à la réussite en lecture de mots des enfants allophones ($\beta = .12, p = .03, \Delta R^2 = .01$). La contribution des performances en répétition de pseudo-mots n'est pas significative ($p = .43$). Ces régressions linéaires multiples sont représentées dans le Tableau 7 ci-dessous.

Tableau 7.

Modèle de régression linéaire multiple expliquant la réussite en lecture chez les enfants allophones

Modèle	Variable	b	β	t	p	ΔR^2
$R^2 = .68$	(Intercept)	8.96	.00	1.64	.10	
	Matrices de Raven	.43	.22	3.34	.00	.03
	DRA	-.30	-.29	-4.38	< .001	.05
	EVIP	.19	.12	2.18	.03	.01
	CP	.51	.40	4.87	< .001	.06
	RPM	.14	.05	.79	.43	.00

Note. la valeur R^2 est la valeur ajustée ; CP : conscience phonologique ; b : coefficient brut ; β : coefficient bêta ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire ; p : significativité ; RPM : répétition de pseudo-mots ; t : test *t* de Student ; ΔR^2 : pourcentage de variance expliqué par le modèle.

Nous avons détaillé la contribution de chaque variable à la réussite en lecture de mots chez les enfants allophones. Nous allons maintenant nous intéresser à la contribution de ces mêmes variables à la réussite en lecture de mots réguliers puis en lecture de mots irréguliers au sein du groupe d'enfants allophones.

3.2.1. Réussite en lecture de mots réguliers

Tout comme les modèles précédents, les matrices de Raven, la dénomination rapide automatisée et la conscience phonologique contribuent de façon unique et significative à la lecture de mots réguliers. La répétition de pseudo-mots n'apparaît pas comme une variable explicative de la réussite en lecture de mots réguliers car la pente n'est pas significative ($p = .31$). Pour finir, le vocabulaire quant à lui y contribue de façon unique et significative ($\beta = .11, p = .05, \Delta R^2 = .02$). Ces régressions linéaires multiples apparaissent dans le Tableau 8 ci-dessous.

Tableau 8.

Modèle de régression linéaire multiple expliquant la réussite en lecture de mots réguliers chez les enfants allophones.

Modèle	Variable	b	β	t	p	ΔR^2
R ² = .66	(Intercept)	6.50	.00	1.45	.15	
	Matrices de Raven	.28	.18	2.61	.01	.02
	DRA	-.20	-.25	-3.60	< .001	.04
	EVIP	.14	.11	1.95	.05	.02
	CP	.46	.46	5.37	< .001	.07
	RPM	.15	.07	1.02	.31	.00

Note. la valeur R² est la valeur ajustée ; CP : conscience phonologique ; b : coefficient brut ; β : coefficient bêta ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire ; p : significativité ; RPM : répétition de pseudo-mots ; t : test *t* de Student ; ΔR^2 : pourcentage de variance expliqué par le modèle.

Nous avons détaillé la contribution de chaque variable à la réussite en lecture de mots réguliers chez les enfants allophones. Nous allons maintenant faire de même pour la réussite en lecture de mots irréguliers chez les enfants allophones.

3.2.2. Réussite en lecture de mots irréguliers

Les régressions linéaires pour la lecture de mots irréguliers mettent en avant une contribution unique et significative des matrices de Raven et de la dénomination rapide automatisée. En revanche, la contribution des autres variables telles que la conscience phonologique, la répétition de pseudo-mots et le vocabulaire ne sont pas significatives. Ces régressions linéaires sont détaillées dans le Tableau 9 ci-dessous.

Tableau 9.

Modèle de régression linéaire multiple expliquant la réussite en lecture de mots irréguliers chez les enfants allophones.

Modèle	Variable	b	β	t	p	ΔR^2
R ² = .50	(Intercept)	2.46	.00	1.42	.16	
	Matrices de Raven	.15	.31	3.74	< .001	.06
	DRA	-.10	-.37	-4.45	< .001	.09
	EVIP	.05	.13	1.79	.08	.01
	CP	.05	.15	1.42	.16	.01
	RPM	-.01	-.01	-.15	.88	.00

Note. la valeur R² est la valeur ajustée ; CP : conscience phonologique ; b : coefficient brut ; β : coefficient bêta ; DRA : dénomination rapide automatisée ; EVIP : tâche de vocabulaire ; p : significativité ; RPM : répétition de pseudo-mots ; t : test *t* de Student ; ΔR^2 : pourcentage de variance expliqué par le modèle.

Nous avons décrit ci-dessus les analyses statistiques réalisées dans le cadre de cette étude et les résultats qui en découlaient. Nous allons maintenant discuter des résultats obtenus au cours de la partie suivante.

Discussion

Après un rappel des objectifs de cette étude, les résultats exposés ci-dessus seront détaillés et discutés dans cette partie. Nous évoquerons ensuite les limites de cette étude.

1. Rappel des objectifs du mémoire

L'objectif de recherche de la présente étude est de déterminer le poids des différents prédicteurs cognitifs et linguistiques sur les compétences en identification de mots écrits des enfants allophones. Nous cherchons ainsi à évaluer si le vocabulaire est plus fortement corrélé aux compétences en identification de mots écrits chez les enfants allophones que chez les enfants tout-venant. Dans ce sens, nous avons également émis l'hypothèse que le vocabulaire contribuerait plus à prédire les compétences sur la tâche de lecture de mots irréguliers que sur la lecture de mots réguliers.

Nous allons maintenant interpréter les résultats obtenus sur les différentes épreuves évaluant les prédicteurs cognitifs et linguistiques et les compétences en lecture de mots.

2. Interprétation des résultats

Dans le cadre de cette étude, nous avons cherché à déterminer les contributions uniques de chaque prédicteur cognitif et linguistique sur les compétences en identification de mots écrits.

Nous retrouvons grâce aux régressions linéaires une contribution unique et significative de la dénomination rapide automatisée, de la conscience phonologique et de la mémoire à court terme, évaluée par la tâche de répétition de pseudo-mots, chez les enfants tout-venant. Ces résultats sont ainsi en accord avec ceux de la littérature scientifique qui place la conscience phonologique, la dénomination rapide automatisée et la mémoire à court terme au rang de prédicteurs cognitifs et linguistiques universaux du développement des compétences en lecture de mots (Caravolas et al., 2012 ; Geva et al., 2019). Concernant le vocabulaire, sa contribution à la réussite en lecture de mots n'est pas significative pour les enfants tout-venant. Nous pensons que cette absence de significativité pourrait être liée à un manque de précision de l'épreuve de vocabulaire en réception. Ce résultat est contradictoire au vu de la littérature existante. En effet, Ouellette (2006) a trouvé que le vocabulaire contribuait significativement à la réussite en lecture de mots car les compétences en identification de mots sont liées à l'étendue du stock lexical. Or, un enfant tout-venant possède à son arrivée en CP, un stock lexical en réception déjà extrêmement étendu. Nous pouvons donc penser que les items utilisés pour cette épreuve ne sont pas suffisamment représentatifs de l'étendue du vocabulaire en réception d'un enfant tout-venant.

Pour les enfants allophones, nous avons trouvé une contribution unique et significative de la dénomination rapide automatisée, de la conscience phonologique et du vocabulaire sur la réussite en lecture de mots. Les résultats de l'épreuve de répétition de pseudo-mots, évaluant la mémoire à court terme, ne sont en revanche pas significatifs. Or, Geva et al. (2019) avaient cité la mémoire à court terme comme étant une compétence cognitive de base qui prédit la lecture de mots, la compréhension en lecture ainsi que la fluidité en lecture dans n'importe quelle langue. De plus, la matrice de corrélation met en avant une relation positive et significative entre la tâche de répétition de pseudo-mots et la lecture de mots. D'autres variables ont par conséquent dû biaiser nos résultats sur cette tâche. En

effet, l'épreuve de conscience phonologique fait également appel à la mémoire à court terme et cette tâche a ainsi pu capter la part de variance de la mémoire à court terme de la tâche de répétition de pseudo-mots. Cependant concernant les régressions linéaires des enfants allophones, nous retenons que le vocabulaire est un prédicteur significatif des compétences en identification de mots.

Nous voulons mettre en avant les différences entre les coefficients de corrélation obtenus sur les matrices de corrélation des enfants tout-venant et des enfants allophones pour la corrélation entre la lecture de mots et chacune des quatre variables évoquées précédemment. Les coefficients de corrélation présentés dans le tableau des enfants allophones montrent une intensité plus importante de la relation entre chaque variable et la lecture de mots que pour les enfants tout-venant. Nous notons dans un premier temps que les enfants allophones ont globalement un niveau inférieur aux enfants tout-venant sur l'ensemble des épreuves évaluant les prédicteurs cognitifs et langagiers et la réussite en lecture en français, comme nous l'ont montré les analyses descriptives. De plus, le rôle des prédicteurs cognitifs et langagiers est plus important au début de l'apprentissage de la lecture. En effet, l'étude de Parrila et al. (2004) montre par exemple une contribution plus importante de la conscience phonologique lors de la première année d'apprentissage de la lecture et cette contribution diminue au cours de la deuxième et troisième année d'apprentissage. Les prédicteurs du langage écrit mis en avant dans cette étude peuvent ainsi apparaître plus prégnants sur les analyses de corrélation des enfants allophones car ils sont aux prémices de l'apprentissage de la lecture en français. Le faible niveau des enfants allophones sur les tâches de lecture ainsi que leur récent apprentissage du français écrit peuvent peut-être expliquer les intensités de corrélation plus fortes entre les prédicteurs cognitifs et langagiers et la lecture de mots. De plus, les caractéristiques de socio-démographiques de l'échantillon, telles que la scolarité antérieure, sont également à prendre en compte. Certains enfants n'ont pas ou peu appris à lire dans une autre langue que le français. Or, nous avons vu que les compétences d'identification de mots écrits développées dans la langue maternelle sont bénéfiques à l'apprentissage d'une langue seconde (Nicot-Guillourel, 2012). La matrice de corrélations des enfants allophones révèle en effet une corrélation forte et significative entre l'apprentissage de la lecture dans une autre langue et la lecture de mots en français. Pour ces enfants qui n'ont pas ou peu été confrontés à l'apprentissage de l'écrit dans une autre langue alphabétique en amont, les compétences sur les prédicteurs cognitifs et linguistiques n'ont de ce fait pas ou peu été entraînées et celles-ci seraient d'autant plus sollicitées lors de l'apprentissage du français écrit comme nous l'avons vu précédemment.

La première hypothèse de cette recherche évoquait une contribution plus forte du vocabulaire sur l'identification de mots des enfants allophones que des enfants tout-venant. Les résultats évoqués précédemment nous permettent d'avancer que le vocabulaire est plus fortement corrélé aux compétences en lecture de mots pour les enfants allophones que pour les enfants tout-venant mais pour ces derniers le lien entre ces deux variables n'est pas significatif. Notre première hypothèse est validée avec une part de variance expliquée par le vocabulaire sur la lecture de mots très faible en ce qui concerne le groupe des enfants allophones.

Notre deuxième hypothèse concernait l'échantillon des enfants allophones uniquement. Nous avons supposé que le vocabulaire serait plus fortement corrélé à la lecture de mots irréguliers qu'à la lecture de mots réguliers. Les régressions linéaires nous montrent une contribution unique et significative du vocabulaire sur la lecture de mots réguliers avec une part de variance expliquée faible. En revanche, les résultats ne sont pas significatifs concernant la contribution du vocabulaire à la réussite en lecture de mots irréguliers. Or, nous avons vu que le vocabulaire oral est particulièrement important lorsque l'orthographe du mot écrit et sa prononciation ne sont pas cohérentes comme c'est le cas pour la lecture de mots irréguliers (Ricketts et al., 2007). Nous notons également, pour la réussite en

lecture de mots irréguliers une absence de significativité de la conscience phonologique et de la mémoire à court terme. Nous pensons que ces résultats sont en partie dus à la construction de l'épreuve de lecture de mots. En effet, celle-ci contenait 48 items dont 37 mots réguliers et 11 mots irréguliers. Le faible nombre de mots irréguliers sur l'épreuve pourrait expliquer l'absence de significativité de trois variables sur cinq évaluées dans le modèle.

3. Limites de l'étude

Nous avons identifié plusieurs limites et biais à cette étude. Tout d'abord, les passations dans les écoles auraient dû commencer dès le mois d'octobre 2021. Or, plusieurs éléments ont repoussé ces passations au mois de mars 2022. Dans un premier temps, le délai de traitement des démarches administratives, signature des conventions entre les différents partenaires et recueil des formulaires d'informations et de consentement des familles, a considérablement retardé le début des passations. De plus, les conditions sanitaires actuelles (COVID-19) ont obligé les écoles à réorganiser les classes et les emplois du temps. Les enseignants d'UPE2A n'avaient plus que quelques élèves allophones dans leur classe et devaient organiser en parallèle des évaluations suite à un nombre important de nouvelles arrivées d'enfants allophones à la fin de notre premier semestre. Pour finir, suite à ces nouvelles arrivées, peu d'élèves allophones répondaient ensuite aux critères imposés par l'étude. Malgré cela, nous avons réussi à constituer des échantillons composés de 186 enfants tout-venant et 115 enfants allophones.

De plus, un léger biais a été mis en évidence lors des passations. En effet, les passations se déroulaient dans une salle isolée pour éviter les perturbations durant les épreuves. Or selon les classes, certains élèves de la classe étaient susceptibles de repasser par la salle dans laquelle nous réalisions la passation ce qui dissipait l'élève qui passait les épreuves et pouvait interférer avec ses réponses.

Une autre limite réside dans l'échantillonnage du groupe d'enfants allophones. En effet, les deux échantillons d'enfants n'ont pas pu être appariés selon l'âge ou le niveau scolaire car afin d'avoir un échantillon d'enfants allophones suffisamment important, nous avons inclus tous les enfants allophones rencontrés à condition qu'ils soient arrivés en France depuis plus de six mois et moins de deux ans. Ainsi des enfants allophones scolarisés jusqu'au CM2 ont été intégrés à l'étude alors que l'échantillon des enfants tout-venant n'inclus que des élèves de CP-CE1. Cependant, les deux échantillons sont appariés selon le temps d'exposition à l'écrit. Les caractéristiques de chaque enfant allophone ont été prises en compte dans les questionnaires d'information. Ces informations concernaient notamment la scolarité antérieure, l'apprentissage de la lecture dans une autre langue ou encore la durée du séjour depuis l'arrivée en France de l'enfant. Cependant, toutes les données ont été analysées ensemble et les différences d'âge entre les groupes ainsi que des variables liées au développement de l'enfant ont pu biaiser les résultats.

Pour finir, des limites d'ordre méthodologique concernant la constitution des épreuves ont été relevées et discutées lors de l'interprétation des résultats. Le nombre réduit d'items pour certaines tâches telles que la lecture de mots irréguliers pourrait être à l'origine du manque de significativité de nos résultats.

Conclusion

Les enfants allophones sont des enfants qui n'ont jamais été confrontés au français et qui sont arrivés en France depuis moins de deux ans. Ces enfants doivent alors développer leurs compétences en français à l'oral et à l'écrit simultanément. L'objectif de recherche de cette étude consiste à déterminer le poids des différents prédicteurs cognitifs et linguistiques dans le développement du langage écrit des enfants allophones. Pour ce faire, des élèves tout-venant et des élèves allophones scolarisés en classe d'UPE2A ont été rencontrés sur les temps scolaires afin de réaliser des épreuves cognitives et langagières proposées par un examinateur.

Les hypothèses avancées au cours de la recherche supposent que le vocabulaire possède une force de prédiction des compétences en identification de mots écrits plus importante chez les enfants allophones que chez les enfants tout-venant. De plus, la deuxième hypothèse de l'étude propose que le vocabulaire contribue de façon plus importante à la lecture de mots irréguliers qu'à la lecture de mots réguliers.

Les résultats de cette étude vont dans le sens d'une contribution unique et significative du vocabulaire sur les compétences en identification de mots écrits des enfants allophones. La contribution du vocabulaire sur l'identification de mots écrits des enfants tout-venant n'est cependant pas interprétable dans le cadre de notre étude. Le vocabulaire a donc sa place auprès des autres prédicteurs cognitifs et linguistiques du langage écrit dans un contexte d'allophonie.

Les résultats concernant une contribution unique du vocabulaire plus importante pour la lecture de mots irréguliers que pour la lecture de mots réguliers ne sont pas probants. Ils valident cependant les résultats retrouvés dans le cadre de la première hypothèse en appuyant le rôle prédictif du vocabulaire sur les compétences en lecture de mots réguliers chez les enfants allophones.

L'allophonie est un statut linguistique peu décrit spécifiquement dans la littérature scientifique francophone. De manière générale, les enfants allophones sont compris dans une catégorie plus large d'enfants de langue « minoritaire ». Cette recherche a donc permis d'abonder la littérature sur le développement de la lecture dans une langue seconde en testant, chez les enfants allophones, les prédicteurs cognitifs de la réussite en lecture connus chez les enfants tout-venant. Nous avons souligné qu'une différence importante entre les deux groupes se situait sur la relation entre le niveau de vocabulaire et la réussite en lecture qui semble plus forte dans la population allophone que la population tout-venant.

Cette étude présente un intérêt pour la pratique clinique de l'orthophonie pour laquelle peu de données scientifiques sont fournies concernant la prise en charge de ces enfants. Ce mémoire permet ainsi d'améliorer la compréhension des processus en jeu lors de l'apprentissage du français écrit pour ces enfants allophones.

Bibliographie

- Araújo, S., Reis, A., Petersson, K. M., & Faisca, L. (2015). Rapid automatized naming and reading performance: a meta-analysis. *Journal of Educational Psychology*, 107(3), 868.
- Armagnague, M., & Rigoni, I. (2018). *Étude sur la scolarisation des élèves allophones nouvellement arrivés (EANA) et des enfants issus de familles itinérantes et de voyageurs (EFIV)*. Défenseur des droits <https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/synth-evascolnum-21.12.18.pdf>
- Avanthey-Granges, N., & Botteron, S. (2003). Evaluation de la compréhension orale et écrite auprès d'enfants avec et sans troubles de la lecture. *Revue Tranel (Travaux neuchâtelois de linguistique)*, 38, 153-166.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, 105(1), 158.
- Blanquer, J.-M. (2012). *Circulaire n° 2012-141*. Ministère de l'Education Nationale de la Jeunesse et des Sports.
- Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological science*, 23(6), 678-686.
- Clayton, F. J., Sears, C., Davis, A., & Hulme, C. (2018). Verbal task demands are key in explaining the relationship between paired-associate learning and reading ability. *Journal of experimental child psychology*, 171, 46-54.
- Content, A., & Zesiger, P. (1999). L'acquisition du langage écrit. *Troubles du Langage: bases théoriques, diagnostic et rééducation*, 181-209.
- Demont, E., & Botzung, A. (2003). Contribution de la conscience phonologique et de la mémoire de travail aux difficultés en lecture : étude auprès d'enfants dyslexiques et apprentis lecteurs. *L'année psychologique*, 103(3), 377-409.
- Desrochers, A., Kirby, J. R., Thompson, G. L., & Fréchette, S. (2009). Le rôle de la conscience phonologique dans l'apprentissage de la lecture. *Revue du Nouvel-Ontario*, 34, 59-82.
- Dunn, L. M., Dunn, L. M., & Theriault-Whalen, C. (1993). *EVIP: échelle de vocabulaire en image Peabody*. Pearson Canada Assessment.
- Ehrlich, M. F., & Delafoy, M. (1990). La mémoire de travail : structure, fonctionnement, capacité. *L'année Psychologique*, 90(3), 403-427.

Fayol, M. (1992). La compréhension lors de la lecture : un bilan provisoire et quelques questions. *La lecture : Processus, apprentissage, troubles*, 79-101.

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading ?. *Journal of experimental child psychology*, 115(1), 218-225.

Geva, E., Xi, Y., Massey-Garrison, A., & Mak, J. Y. (2019). Assessing Reading in Second Language Learners : Development, Validity, and Educational Considerations. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Éds.), *Reading Development and Difficulties* (p. 117-155). Springer International Publishing.

Ginestet, E., Valdois, S., Diard, J., & Bosse, M. L. Comprendre l'apprentissage de l'orthographe lexicale et ses difficultés pour mieux l'enseigner: apports et limites des dernières modélisations computationnelles. *ANAE—Approche Neuropsychologique des Apprentissages chez L'Enfant*.

Goigoux, R., Cèbe, S., & Pironom, J. (2016). Les facteurs explicatifs des performances en lecture-compréhension à la fin du cours préparatoire. *Revue française de pédagogie*, (3), 67-84.

Kim, Y. S. G., Cho, J. R., & Park, S. G. (2018). Unpacking direct and indirect relationships of short-term memory to word reading: evidence from korean-speaking children. *Journal of learning disabilities*, 51(5), 473-481

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, 45(3), 341-362.

Korkman, M., Kirk, U., & Kemp, S. (2012). *NEPSY-II—Bilan neuropsychologique de l'enfant - 2nde édition*. Pearson.

Labat, H., Farhat, S. L., Andreu, S., Rocher, T., Cros, L., Magnan, A., & Ecalle, J. (2013). *Évaluation des connaissances précoces prédictives de l'apprentissage de la lecture en grande section de maternelle* (No. 184, pp. 41-54). ENS Éditions.

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of experimental child psychology*, 73(3), 183-207.

Lecocq, P., & Casalis, S. (1996). *Apprentissage de la lecture et compréhension d'énoncés*. Presses Univ. Septentrion.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: a Norwegian latent variable longitudinal study. *Developmental psychology*, 45(3), 764.

Majerus, S. (2010). Les multiples déterminants de la mémoire à court terme verbale: Implications théoriques et évaluatives. *Développements*, (1), 5-15.

- Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1998). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of experimental child psychology*, 71(1), 3-27.
- Negro, I., & Genelot, S. (2009). Les prédicteurs en grande section maternelle de la réussite en lecture en fin de première année d'école élémentaire : l'impact du nom des lettres. *Bulletin de psychologie*, (3), 291-306.
- Nicot-Guillorel, M. (2012). Bilinguisme et apprentissage de la lecture à Madagascar : quelle place pour la phonologie ? *Revue française de pédagogie*, 178, 97114. <https://doi.org/10.4000/rfp.3573>
- 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*, 98(3), 554-566.
<https://doi.org/10.1037/0022-0663.98.3.554>
- Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading : How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, 23(2), 189-208.
- Plaza, M., & Robert-Jahier, A.-M. (2006). *DRA Enfants—Test de Dénomination Rapide*. Adeprio.
- Poncelet, M., & Van der Linden, M. (2003). L'évaluation du stock phonologique de la. *Revue de neuropsychologie*, 13(3), 377-407.
- Raudszus, H., Segers, E., & Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *Journal of Research in Reading*, n/a(n/a).
- Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11(3), 235-257.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4), 523.
- Sprenger-Charolles, L., Colé, P., Piquard-KipSer, A., & Leloup, G. (2010). *EVALEC*. Ortho Edition.
- Stanké, B. (2006). La compréhension de textes. *Rééducation orthophonique*, 227, 45-54.
- Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, 23(1), 236-255.

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills?. *Scientific Studies of Reading*, 16(2), 122-140.

Valdois, S. (2010). Evaluation des difficultés d'apprentissage de la lecture. *Revue française de linguistique appliquée*, 15(1), 89-103.

Wegener, S., Wang, H. C., de Lissa, P., Robidoux, S., Nation, K., & Castles, A. (2018). Children reading spoken words: Interactions between vocabulary and orthographic expectancy. *Developmental Science*, 21(3), e12577.

Ziegler, J., 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.

Liste des annexes

Annexe n°1 : Lettre d'information aux parents ou tuteurs légaux

Annexe n°2 : Lettre d'information aux enfants à lire par leurs parents ou tuteurs légaux

Annexe n°3 : Consentement de participation libre et éclairée

Annexe n°4 : Questionnaire (version destinée aux titulaires de l'autorité parentale et/ou aux tuteurs légaux des enfants allophones)

Annexe n°5 : Questionnaire (version destinée aux titulaires de l'autorité parentale et/ou tuteurs légaux des enfants tout-venant)

Annexe n°6 : Description des épreuves utilisées