

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



Université  
de Lille



# MEMOIRE

En vue de l'obtention du  
Certificat de Capacité d'Orthophoniste  
présenté par  
**Camille Lovichi**  
soutenu publiquement en juin 2023

## Importance du vocabulaire comme prédicteur de l'entrée dans la lecture chez les élèves allophones

MEMOIRE dirigé par  
**Matthieu BIGNON**, Orthophoniste et doctorant, SCALab UMR 9193, Lille  
**Séverine CASALIS**, Professeure des Universités SCALab, CNRS et Université de Lille

Lille –2023

# Remerciements

Je remercie infiniment Monsieur Matthieu Bignon et Madame Séverine Casalis d'avoir accepté de m'encadrer dans la rédaction de mon mémoire, et de m'avoir confié ce sujet. Je tiens à leur exprimer toute ma reconnaissance pour leur disponibilité à mes questions, leur pédagogie et l'élan qu'ils m'ont donné.

Mes plus vifs remerciements vont aussi à Madame Ingrid Gibaru, ma lectrice.

Je souhaite également remercier Madame Bergere, d'avoir soutenu ce travail à travers le CASNAV, et tous les enseignants qui m'ont accueillie, pour leur investissement crucial dans cette étude. Merci à tous les enfants qui y ont prêté leur concours.

Merci aux maîtres de stage qui m'auront permis de me projeter avec joie dans un nouvel avenir professionnel.

Merci à ma belle promo avec qui j'ai traversé ces cinq longues années.

Merci à mes ami(e)s qui m'ont aidée à avancer. Merci d'avoir été là.

## **Résumé**

La lecture est une compétence clé dans nos sociétés contemporaines où l'écrit est omniprésent. L'étude des facteurs cognitifs sous-jacents à son apprentissage permet de favoriser sa réussite.

L'objectif de notre étude a été d'étudier l'apport du vocabulaire comme prédicteur de l'entrée dans la lecture chez les élèves allophones. Accueillis au début de leur scolarisation dans les Unités pédagogiques pour élèves allophones arrivants (UPE2A), ceux-ci présentent des profils variés. Ils n'ont pas tous bénéficié d'une acculturation à l'écrit dans leur pays d'origine et ont un vocabulaire limité en français.

Nous leur avons fait passer des tests sélectionnés pour évaluer très succinctement le fonctionnement cognitif général, le niveau de langage oral et de langage écrit. Des régressions linéaires multiples nous ont permis d'estimer l'importance du stock lexical sur la réussite en lecture après avoir contrôlé le QI non-verbal ainsi que les compétences phonologiques.

Les résultats de notre étude suggèrent qu'en deçà d'un certain seuil lexical, le mécanisme d'auto-apprentissage de la lecture par auto-correction ne parvient pas à s'enclencher. Ils encouragent à poursuivre l'exploration des liens entre vocabulaire et automatisation de la lecture au moyen d'un autre design, ou en accompagnant la stimulation du stock lexical de mesures complémentaires.

## **Mots-clés**

Vocabulaire, Prédicteurs de la lecture, Reconnaissance des mots écrits, EANA.

## **Abstract**

Reading is a key skill in our modern societies where the written word is omnipresent. The study of the cognitive factors underlying its learning helps to promote its success.

The aim of our study was to investigate the contribution of vocabulary as a predictor of successful learning to read in allophone students. These students are usually enrolled in the Pedagogical Units for Newly Arrived Allophone Students (UPE2A) at the beginning of their schooling. They present a variety of profiles. They have not all benefited from an acculturation to writing in their country of origin and have a limited vocabulary in French.

We gave them selected tests to evaluate very briefly the general cognitive functioning, the level of oral and written language. Multiple linear regressions allowed us to estimate the importance of lexical stock on reading achievement after controlling for non-verbal IQ and phonological skills.

The results of our study suggest that below a certain lexical threshold, the self-learning mechanism of reading by self-correction fails to kick in. They encourage further exploration of the links between vocabulary and automatic reading by means of another design, or by accompanying the stimulation of the lexical stock with complementary measures.

## **Key words**

Vocabulary, Predictors of reading success, Written word recognition, Allophone pupils.

# Table des matières

|  |    |
|--|----|
| Introduction .....   | 5  |
| Contexte théorique, buts et hypothèses .....   | 6  |
| 1 Contexte théorique .....   | 6  |
| 1.1 Mécanismes de la lecture .....   | 6  |
| 1.1.1 Modèle simple de la lecture de Gough et Tunmer (1986).....                         | 6  |
| 1.1.2 Lexique mental et stock visuo-lexical.....   | 6  |
| 1.1.3 Modèle à double voie de Coltheart (1978) .....                                     | 6  |
| 1.1.4 Modèles connexionnistes comme le modèle triangulaire.....                          | 7  |
| 1.2 Mécanismes d'apprentissage et d'automatisation de la lecture.....                    | 7  |
| 1.2.1 Acculturation à l'écrit.....   | 7  |
| 1.2.2 Stratégies du lecteur débutant .....   | 8  |
| 1.2.3 Hypothèse d'auto-apprentissage de Share.....                                       | 8  |
| 1.2.4 Importance du langage oral et <i>set for variability</i> .....                     | 8  |
| 1.3 Prédicteurs de l'apprentissage de la lecture .....                                   | 9  |
| 1.3.1 Prédicteurs phonologiques .....  | 10 |
| 1.3.2 Apport du vocabulaire comme prédicteur de la lecture de mots à voix haute          | 11 |
| 1.3.3 Influence de l'environnement sur l'apprentissage de la lecture.....                | 12 |
| 1.4 Spécificités de l'allophonie et conséquences sur l'apprentissage de la lecture ..... | 13 |
| 1.4.1 Définitions.....   | 13 |
| 1.4.2 Hétérogénéité des profils .....  | 13 |
| 2 Buts .....   | 15 |
| 3 Hypothèse.....   | 15 |
| Méthode.....   | 15 |
| 1 Procédure .....  | 15 |
| 2 Population .....   | 17 |
| 3 Matériel .....   | 18 |
| 3.1 Evaluation des compétences .....   | 18 |
| 3.1.1 Mesure des compétences cognitives sous-jacentes .....                              | 18 |
| 3.1.2 Mesure des compétences phonologiques sous-jacentes .....                           | 19 |
| 3.1.3 Mesure du vocabulaire réceptif .....   | 19 |
| 3.1.4 Mesure de la précision de la capacité de lecture de mots à voix haute .....        | 20 |
| Résultats .....  | 20 |
| 1 Performances aux tests .....   | 20 |
| 2 Modélisation des relations entre les variables .....                                   | 21 |
| 2.1 Matrices de corrélation.....   | 21 |
| 2.2 Modèles de régressions linéaires .....   | 22 |
| Discussion et Conclusion.....  | 24 |
| Bibliographie.....   | 26 |

# Introduction

« La lecture a été érigée en grande cause nationale de l'année 2021-2022. Elle est en effet reconnue comme un vecteur de transmission du savoir, de culture, d'égalité des chances et, au-delà, d'épanouissement personnel pour les élèves » (BO, 2021).

Dans le monde contemporain, la maîtrise du langage écrit est un fondement de l'estime de soi, de l'autonomie, elle conditionne en particulier les apprentissages, l'insertion professionnelle et l'intégration sociale (Agence nationale de lutte contre l'illettrisme).

Son apprentissage est secondaire à celui du langage oral, qui s'acquiert lui, de manière implicite, grâce au « bain de langage » dans lequel sont plongés les individus neuro-typiques dès leur naissance. Si l'écriture est un code qui permet de transcrire la parole, la lecture permet de décrypter les mots écrits et de retrouver le sens derrière leur trace écrite. Or, nombre de jeunes adultes, qui bénéficient pourtant d'une scolarité gratuite et obligatoire dans la plupart des pays industrialisés ne maîtrisent pas correctement la lecture à l'issue de leur scolarité (Fluss & al., 2009). Les mêmes auteurs citent des études selon lesquelles l'échec scolaire et les difficultés en lecture sont fortement liés à des facteurs socio-économiques individuels et environnementaux lorsqu'ils n'ont pas une origine neurobiologique. Ils mettent également en exergue la forte corrélation entre les troubles de l'apprentissage de la lecture et les facteurs comportementaux et émotionnels tels l'inattention ou l'anxiété. Cependant, parmi les facteurs de risque ou de protection de l'apprentissage de la lecture, les chercheurs étudient préférentiellement les facteurs cognitifs spécifiques de cet apprentissage. Kirby et al. (2008) expliquent ce choix en invoquant 1) la difficulté, non seulement d'établir un lien de causalité entre les facteurs socio-émotionnels et socio-économiques et l'apprentissage de la lecture, mais aussi d'améliorer cet apprentissage en agissant sur ces facteurs 2) le fait que ces facteurs non cognitifs influencent l'apprentissage de la lecture par le biais des mécanismes cognitifs sous-jacents à celui-ci.

C'est dans ce contexte que nous nous proposons d'étudier l'entrée dans la lecture des élèves allophones. Le prisme de l'allophonie permet en effet d'envisager les moyens de prévention de l'échec de l'apprentissage de la lecture sous l'angle du vocabulaire, un prédicteur nettement moins exploré, dans la littérature, que celui de la phonologie.

En partant des modèles neurocognitifs de référence, nous examinerons quels sont les traitements spécifiques de la lecture, les rouages d'une lecture experte et les ressorts d'une lecture efficiente. Les modèles nous aideront également à comprendre comment ceux-ci se mettent en place au cours de l'apprentissage. Nous dresserons ensuite l'inventaire des prédicteurs de l'entrée dans la lecture selon les critères établis par les chercheurs. Enfin, nous nous pencherons sur les spécificités de la population cible de notre étude, les élèves allophones, par rapport à la population contrôle des élèves natifs et nous essaierons de déterminer en quoi celles-ci peuvent constituer un facteur de risque et partant, un levier de prévention, pour l'apprentissage de la lecture.

Nous décrirons ensuite la méthodologie employée pour mener cette étude, puis nous présenterons les résultats obtenus après analyse et nous discuterons leur interprétation. Nous achèverons notre propos en évoquant les limites de l'étude, mais aussi les perspectives qu'elle ouvre.

# Contexte théorique, buts et hypothèses

## 1 Contexte théorique

### 1.1 Mécanismes de la lecture

#### 1.1.1 Modèle simple de la lecture de Gough et Tunmer (1986)

Pour préciser le cadre de notre étude, nous nous appuyerons sur le modèle simple de la lecture de Gough et Tunmer (1986), qui se présente sous la forme d'une équation :  $L = R \times C$ . Dans cette formule, L représente la compréhension du texte lu, qui est la finalité de la lecture, R la reconnaissance des mots écrits et C la compréhension orale (Castles & al. 2018).

Cette équation montre ainsi que deux types de processus mentaux sont à l'œuvre dans la lecture. Certains seulement sont spécifiques du traitement de l'information écrite (Alegria & Morais, 1989). Ce sont des processus de bas niveau, qui entrent en jeu au niveau du mot (Kirby & Savage, 2008), et sont liés à R. Les processus qui relèvent de la composante C sont pour leur part des processus de haut niveau liés au langage oral. Ils nécessitent d'importantes ressources attentionnelles qui, pour une lecture efficace, ne peuvent être dévolues simultanément à la reconnaissance des mots écrits.

Plus encore que la maîtrise des processus de reconnaissance des mots écrits (R), c'est donc son automatiser qui détermine l'efficacité de la lecture (L) (Bowey, 2005). Aussi est-ce sur les mécanismes en jeu à ce stade que nous porterons notre attention dans le cadre de ce travail.

#### 1.1.2 Lexique mental et stock visuo-lexical

D'un point de vue psycholinguistique, le lexique mental est défini comme l'ensemble des représentations des mots de la langue emmagasinées dans la mémoire à long terme, avec toutes leurs propriétés (phonologiques, orthographiques, morphologiques, syntaxiques et sémantiques ; Nicolas, 2003). Ces connaissances se développent avec le langage oral, qui s'acquiert par imitation et imprégnation au fil des interactions avec les locuteurs.

La reconnaissance d'un mot écrit, autrement dit « l'accès au lexique », désigne les opérations qui permettent d'associer une représentation sensorielle provenant du stimulus (sonore ou graphique) à une ou plusieurs (dans le cas d'un stimulus ambigu) représentation(s) mentale(s) correspondant à un mot de la langue (Nicolas, 2003).

#### 1.1.3 Modèle à double voie de Coltheart (1978)

Dans le cadre de notre étude, le modèle à double voie de Coltheart (1978), représenté à la Figure 1 ci-dessous, aide à comprendre comment la forme écrite du mot peut être mise en relation avec la représentation stockée dans le lexique mental ou lue ex nihilo. Il s'agit toutefois d'un modèle théorique de lecture experte, selon lequel les procédures de la lecture sont automatisées.

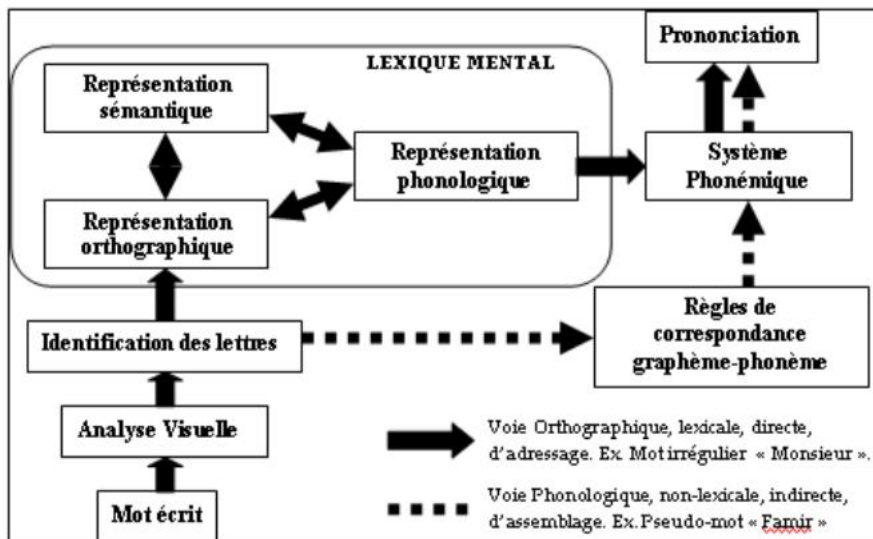


Fig. 1 Modèle à double voie de Coltheart, 1978, in Tossonian (2019)

### 1.1.4 Modèles connexionnistes comme le modèle triangulaire

L'implémentation des modèles de la lecture sur ordinateur et les enseignements des neurosciences ont présidé à une autre approche, dite « connexionniste », de l'apprentissage de la lecture. Les modèles qui en découlent, tel le modèle triangulaire, envisagent la lecture comme l'activation d'un ensemble de représentations mentales, mettant en œuvre différentes aires du cerveau. La représentation d'entrée est un mot écrit appréhendé comme un patron visuel. Elle active une représentation de sortie qui peut être phonologique – via le décodage alphabétique ou la subvocalisation – et/ou sémantique. Ce type de modèle ne présuppose pas l'accès à un lexique mental. Il montre que la spécialisation orthographique résulte du renforcement des liaisons engendré par la coactivation répétée des patrons visuels des mots et des unités sémantiques et phonologiques (Morais et al., 2003).

## 1.2 Mécanismes d'apprentissage et d'automatisation de la lecture

### 1.2.1 Acculturation à l'écrit

Contrairement aux peintures pariétales préhistoriques ou aux hiéroglyphes égyptiens par exemple, les systèmes d'écriture sont une invention culturelle qui permet de transcrire la parole et non le sens (Sprenger-Charolles & Ziegler, 2019). On pourra objecter que cette affirmation minimise la part sémiographique (marques flexionnelles, dérivationnelles, étymologiques) de langues alphabétiques tel le français écrit et la part phonologique de langues plus iconographiques, telles les langues à idéogrammes.

Néanmoins, il découle de cela que l'apprenti lecteur doit, d'une part, commencer par acquérir ce que Sénéchal et al. (2001) nomment les connaissances « conceptuelles » de la littéracie, autrement dit devenir familier des supports et de la fonction sociale de l'écrit. D'autre part, il doit s'en approprier les connaissances « procédurales » – comprendre que l'écrit code l'oral grâce à un système de signes arbitraires, les lettres, et apprendre à manipuler intentionnellement les structures linguistiques.

### **1.2.2 Stratégies du lecteur débutant**

Selon le modèle développemental de Frith (1985), l'apprenti lecteur utilise successivement trois procédures de traitement des mots écrits. Le pré-lecteur traite les mots comme de simples objets visuels, qu'il distingue par leurs traits saillants (stade « logographique »). Le stade « alphabétique » commence avec l'enseignement explicite du principe alphabétique, qui préside au « décodage phonologique ». Celui-ci comporte trois phases : la segmentation des mots écrits en graphèmes, la conversion des graphèmes en phonèmes, la fusion de ces phonèmes (Morais et al., 2003). Également appelé « recodage phonologique » (Bowey, 2005), il permet de prononcer, autrement dit de retrouver la forme phonologique des mots écrits non familiers réguliers.

L'apprenti lecteur passe ainsi progressivement d'un traitement séquentiel à un traitement simultané des lettres. Il entre alors dans le stade « orthographique », où il parvient à reconnaître les morphèmes contenus dans les mots écrits en les récupérant dans sa mémoire à long terme, sans passer par la correspondance grapho-phonémique.

La différence fondamentale entre les stades « alphabétique » et « orthographique » de ce modèle et les procédures homonymes du modèle de lecture experte est l'absence d'automatisation, qui rend le déchiffrage lent et laborieux, et très coûteux cognitivement pour le lecteur débutant. Comme le soulignent Demont & Gombert (2004), en mobilisant toute l'attention du jeune lecteur, la conversion grapho-phonémique s'automatise progressivement, contribue à créer un code orthographique utilisable, et permet l'identification de nouveaux mots jamais rencontrés.

Des données expérimentales sont également en faveur d'une contribution de la dimension sémantique à ce processus d'automatisation. Ainsi, l'apprenti lecteur qui parvient à décoder correctement des mots nouveaux introduits en contextes signifiants reconnaît ces mots beaucoup plus rapidement, par la suite, que des mots nouveaux équivalents, et a tendance à les orthographier correctement (Valdois et al., 2009).

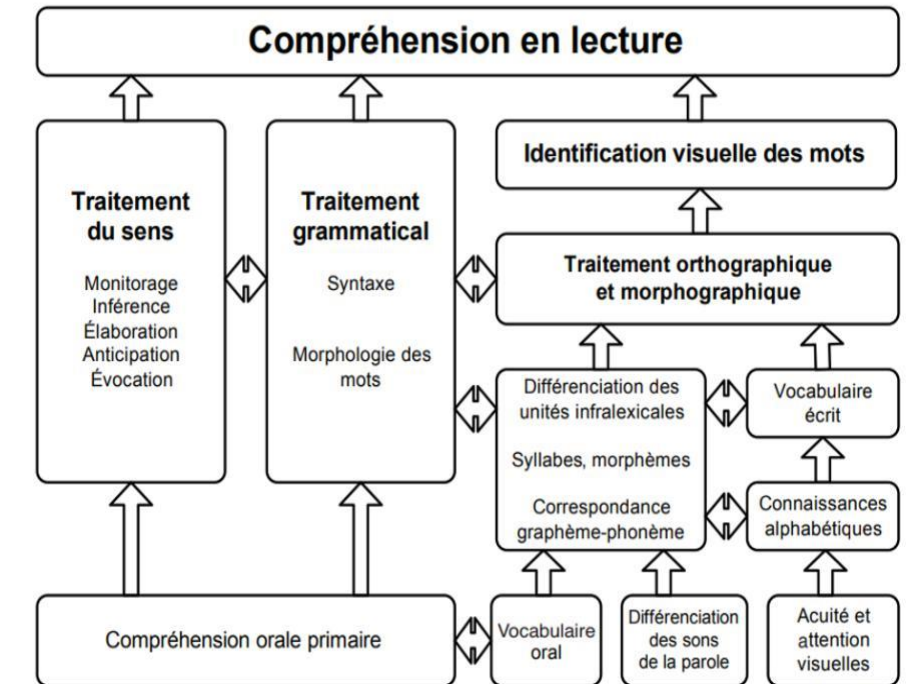
### **1.2.3 Hypothèse d'auto-apprentissage de Share**

Share (1995) a émis l'hypothèse que la conversion grapho-phonémique constituait un puissant mécanisme d'auto-apprentissage. En permettant à l'apprenti lecteur de lire tous les mots écrits qu'il n'a encore jamais rencontrés, elle le rend en effet autodidacte.

### **1.2.4 Importance du langage oral et *set for variability***

Parmi les limites du modèle de Frith, Demont & Gombert (2004) pointent notamment la non-prise en compte des compétences langagières préalables à l'entrée dans la lecture. Ils soulignent le rôle de ces habiletés de traitement du langage oral dans le versant implicite de l'apprentissage de la lecture, autrement dit dans les changements cognitifs qui s'opèrent à l'insu de l'apprenti lecteur au fil de ses confrontations avec les régularités du langage écrit.





**Fig. 2 : Représentation schématique partielle des habiletés et des bases de connaissances sollicitées par la lecture (adapté de Desrochers et al. (2012))**

Le modèle cognitif de Desrochers et al. (2012) reproduit ci-dessus (Figure 2) rend compte de la participation des habiletés de traitement du langage oral (lexique, lexique phonologique, connaissance des contrastes phonétiques, du système phonologique, des constructions syntaxiques les plus usuelles de la langue) aux processus de lecture que l'apprenti lecteur doit apprendre à automatiser.

Ainsi, les mots contenant des irrégularités ne peuvent être lus par simple application des règles de conversion grapho-phonémique (Demont & Gombert, 2004). Ils doivent l'être par analogie avec la forme phonologique des mots stockés en mémoire (vocabulaire oral) et auto-corrrections. Cette opération est désignée dans la littérature sous le terme de « *set for variability* » (Edward et al., 2020).

Après avoir observé les mécanismes de l'automatisation des procédures spécifiques de lu langage écrit, nous allons porter notre attention sur les facteurs cognitifs qui les soutiennent.

### 1.3 Prédicteurs de l'apprentissage de la lecture

Dans l'abondante littérature consacrée aux prédicteurs précoces de la lecture, Écalte et al. (2020) citent les connaissances phonologiques associées à l'acquisition des premières connaissances alphabétiques (connaissance des lettres, code alphabétique) comme des précurseurs puissants de la reconnaissance des mots dans sa dimension de décodage. La méta-analyse de Scarborough (1998), qui couvre 61 études longitudinales portant sur l'anglais, réalisées entre 1976 et 1998, met particulièrement en avant les connaissances liées aux lettres (identification et connaissance de leur nom et de leur son), les capacités d'analyse phonologique et la mémoire de travail phonologique.

Hatcher et al. (2004) identifient les trois facteurs de risques de difficultés suivants : une mauvaise connaissance des lettres, un faible niveau de conscience phonologique et un vocabulaire limité. Nous ne nous attarderons pas sur la connaissance des lettres, qui perd nettement de sa significativité dès lors que l'enfant pré-lecteur commence son apprentissage explicite de la lecture à l'entrée à l'école élémentaire.

### **1.3.1 Prédicteurs phonologiques**

La conscience phonologique désigne la capacité intentionnelle de manipulation à l'oral des unités linguistiques qui composent les mots, à savoir les syllabes et les phonèmes (Gombert & Colé, 2000). Le fait de manipuler les phonèmes aide les apprenants à décoder de nouveaux mots (...) car avec le temps, ils arrivent à faire le lien entre ceux-ci et leurs graphèmes respectifs (Ehri et al., 2001).

En tant que compétence sous-jacente à la conversion grapho-phonémique, la conscience phonologique constitue l'un des meilleurs prédicteurs de la réussite de l'entrée dans l'écrit.

Trois types de résultats en attestent : a) la forte corrélation entre la conscience phonologique des élèves pré-lecteurs et leur niveau de lecture au CP (Share et al. 1984) ; b) le fait que les enfants dont la conscience phonologique reste déficitaire alors qu'ils fréquentent l'écrit depuis plusieurs mois, voire plusieurs années sont mauvais lecteurs, plus précisément ont des difficultés de reconnaissance des mots écrits, selon le consensus scientifique rapporté par Demont (2003). Lecocq (1991) invoque également, à ce sujet, les difficultés de certains dyslexiques et des adultes illettrés dans les tâches de manipulation de sons ; enfin c) le fait que les entraînements visant à améliorer la conscience phonologique des enfants renforcent leur niveau de lecture (Bara et al., 2004). Faute de comprendre que chaque mot est constitué de plus petits éléments, un élève dont la conscience phonologique est défaillante aura du mal à décoder de nouveaux mots, ce qui pourra affecter la fluidité en lecture (Ehri et al., 2001).

La conscience phonologique couvre, chronologiquement d'un point de vue développemental, la conscience de la syllabe, la conscience de la rime et la conscience phonémique. D'après une méta-analyse de Melby-Lervåg et al. réalisée en 2012, qui portait sur 235 études, cette dernière composante est la plus fortement corrélée des trois à l'apprentissage de la lecture de mots, aussi bien chez les enfants normo-lecteurs que chez les enfants dyslexiques. Elle ressort de cette méta-analyse comme le meilleur prédicteur des différences interindividuelles au cours de l'apprentissage de la lecture. Sa significativité diminue cependant avec l'augmentation du niveau scolaire. Selon la littérature, cette habileté permet de comprendre le principe alphabétique dans les premières étapes de la mise en place de la procédure phonologique. Elle permet d'accéder à la fusion phonémique, et son automatisation conditionnerait le recours à des unités de recodage plus larges que les phonèmes (Bara et al., 2004).

La mémoire de travail est un système de mémoire active qui permet la rétention d'informations dans le cadre d'un processus de traitement. Selon le modèle de Baddeley (1986), elle se compose d'un « administrateur central », mécanisme attentionnel qui contrôle et coordonne le « calepin visuo-spatial » et la « boucle phonologique ». Ces deux sous-systèmes permettent de retenir et de manipuler respectivement des informations de nature visuelle et auditive.

La mémoire de travail phonologique permet le maintien en mémoire du résultat des conversions grapho-phonémiques et des fusions phonémiques déjà réalisées pendant que les opérations de décodage se poursuivent. Elle est reconnue comme un prédicteur de réussite de

l'apprentissage de la lecture, en particulier de la mise en place de la procédure de recodage phonologique. Nithart et al. (2009) citent, d'une part, à ce sujet, la corrélation mise en évidence par une étude longitudinale entre les performances obtenues par des enfants pré-lecteurs à des épreuves de mémoire de travail phonologique, et par ces mêmes enfants à des épreuves de lecture par décodage phonologique, en début d'apprentissage de la lecture. Selon Poncelet et al. (2001), les performances en mémoire de travail phonologique sont corrélées avec les performances en lecture de non-mots mais pas avec celles en lecture de mots irréguliers.

Majerus & Van der Linden (2003) ont par ailleurs montré l'influence automatique des connaissances phonologiques et lexicales sur les capacités en mémoire de travail phonologique. La contribution unique de la conscience phonologique et de la mémoire phonologique aux compétences en lecture ne fait toutefois pas consensus dans la littérature. Pour certains auteurs, une mémoire phonologique élevée pourrait en effet être seulement un signe de la bonne qualité des représentations phonologiques (Desrochers et al. 2012).

La latence de dénomination d'un stimulus est un très bon prédicteur du niveau de lecture. Selon diverses études citées par Ziegler et al. (2008), la vitesse de dénomination des chiffres et des lettres prédit le décodage chez les enfants de primaire. Les chiffres de 1 à 9 sont considérés comme des stimuli très automatisés parce qu'ils sont en nombre limité et que le lien entre leur forme et leur nom est sur-appris. C'est aussi le cas pour les lettres isolées, bien qu'elles soient moins nombreuses. Ainsi, de mauvaises performances en dénomination rapide et à voix haute d'une série de chiffres est souvent corrélée à des difficultés de lecture. Une possible explication de cette corrélation tient au fait que cette tâche permet de mesurer la rapidité et l'automatisation de l'accès aux représentations phonologiques. Cette hypothèse, vérifiée par les mêmes auteurs lors d'une étude sur le lien entre Dénomination rapide automatisée (DRA) et lecture dans une population d'enfants dyslexiques a été corroborée par les résultats de l'imagerie cérébrale.

Pour certains auteurs, une faible performance à une tâche de DRA traduirait une activation des phonèmes insuffisamment rapide pour permettre la fusion phonémique et partant, la lecture de mots à voix haute. D'autres voient plutôt la DRA comme un prédicteur de l'efficacité du traitement phonologique, du fait de sa corrélation avec la conscience phonologique (Desrochers et al., 2012).

### **1.3.2 Apport du vocabulaire comme prédicteur de la lecture de mots à voix haute**

Le vocabulaire est reconnu comme un prédicteur précoce de la lecture. Selon une étude réalisée par Bowey en 1995, le vocabulaire réceptif d'un échantillon de 116 enfants pré-lecteurs mesuré au « Kindergarten » prédisait 20 à 27 % de la variance de leur niveau de lecture à la fin de la première année d'apprentissage de la lecture (2005).

Nation & Cocksey (2009) montrent que l'étendue du vocabulaire de l'enfant pré-lecteur est un bon prédicteur de ses performances en lecture de mots à voix haute et ce, selon des études citées par Desrochers et al. (2012), qu'elles soient mesurées concurremment ou ultérieurement.

Dans une synthèse des recherches consacrées aux prédicteurs de succès de l'apprentissage de la lecture, Colé (2012) cite une étude de Ouellette et Beers (2010) d'après laquelle, en première année de primaire, la conscience phonémique est un prédicteur majeur des performances en lecture de pseudo-mots (37,9 %) tandis qu'en première année de collège, elle est supplantée par l'étendue du vocabulaire (13,4 %). Les auteurs en concluent que le vocabulaire capte une part croissante de la variance en décodage au fur et à mesure que cette habileté s'améliore.

Le niveau de développement lexical permet en effet de prédire le niveau de conscience phonologique des enfants (Ecalte & Magnan 2020). Ainsi, selon le modèle de restructuration lexicale de Walley et al. (2003), l'accroissement de la taille du vocabulaire expressif amène les représentations phonologiques à s'affiner et à se préciser pour faciliter la discrimination des voisins phonologiques (mots se distinguant par un seul phonème). Dans les tâches de métaphonologie (isolement et permutation de phonèmes), les meilleures performances sont réalisées par les enfants qui ont le meilleur niveau de vocabulaire (Metsala, 1999).

La lecture de mots irréguliers est prédite par la précision (sens) du vocabulaire en première année de primaire et par l'étendue de celui-ci (nombre de mots connus) en première année de collège (Ouellette et Beers, 2010, cités par Colé, 2012). Chez les plus jeunes, le décalage entre ces deux critères peut être important.

Le vocabulaire permet le *set for variability* évoqué précédemment. Plus spécifiquement, Ricketts et al. (2007) ont montré que la taille du vocabulaire oral est prédictive du niveau de lecture de mots irréguliers, à l'exclusion des mots réguliers et des non-mots. Ils expliquent ce résultat en partie par la contribution de la forme phonologique des mots au décodage, en l'absence de représentations orthographiques.

Outre les facteurs intrinsèques, d'autres facteurs, extrinsèques, influencent également l'apprentissage de la lecture.

### **1.3.3 Influence de l'environnement sur l'apprentissage de la lecture**

Une abondante littérature est consacrée à l'influence du milieu socio-économique sur le niveau de lecture orale.

L'acculturation à l'écrit se développe tout d'abord de manière informelle. La familiarisation des élèves d'âge pré-scolaire avec le langage écrit peut atteindre un degré plus ou moins élevé en fonction des habitudes culturelles familiales (Lahire, 2008 in Tiré et al., 2015). Les enfants issus d'un milieu défavorisé sont généralement de faibles lecteurs parce qu'ils ont du mal à comprendre que le code écrit est la transcription du code phonologique (Negro & Genolo, 2009).

Le niveau de lecture orale est fonction des habiletés à l'oral et de l'étendue du vocabulaire, qui sont tributaires des stimulations langagières : communications adressées, questions nécessitant une réponse articulée, etc. Ainsi, les enfants issus de milieux défavorisés ont du mal à s'approprier le recodage phonologique pour accéder à la forme phonologique des mots à partir de leur forme orthographique (Negro & Genolo, 2009).

Dans un parti-pris d'homogénéisation de ces différences, le développement du langage oral et l'acculturation à l'écrit sont deux des principaux axes de travail dévolus à l'école maternelle – une spécificité française du système éducatif.

Desrochers et al. (2009) relèvent que les répercussions d'une conscience phonologique déficitaire sur les performances en lecture peuvent être amplifiées ou compensées en fonction des ressources matérielles et humaines disponibles.

Selon Desrochers et al. (2012), le bénéfice de l'enseignement formel de la lecture à l'école est fortement conditionné par le niveau de préparation à cet apprentissage. Les auteurs citent une étude montrant qu'une intervention d'appoint intensive en littéracie a permis à 70 % des élèves en retard de lecture de rattraper leur groupe de référence et de maintenir leur rang jusqu'à la dernière évaluation de l'étude, à la fin de la 4<sup>e</sup> année.

## **1.4 Spécificités de l’allophonie et conséquences sur l’apprentissage de la lecture**

### **1.4.1 Définitions**

Un élève allophone est « un élève qui parle une ou plusieurs autres langues que le français à son arrivée en France ». Est considéré comme « élève allophone nouvellement arrivé » (EANA) tout élève venant de l’étranger, quelle que soit sa nationalité, présent sur le territoire national depuis moins d’un an (Casnav de Lille, 2021).

La Note d’information n°20.39 de novembre 2022 de la Direction de l’évaluation, de la prospective et de la performance (DEPP) précise notamment : « Est considéré comme EANA tout élève ayant des besoins éducatifs particuliers dans l’apprentissage du français langue seconde (FLS) (...) [voire] les apprentissages scolaires de manière générale. (...). Un élève allophone le demeure tout au long de sa scolarité mais il est allophone ‘nouvellement arrivé’ tant que le besoin éducatif existe. »

En application de l’article L. 111-1 du code de l’éducation, en France, l’école est un droit pour tous les enfants résidant sur le territoire national, quels que soient leur nationalité, leur statut migratoire ou leur parcours antérieur. Les EANA, qui arrivent tout au long de l’année, sont accueillis dans des Unités Pédagogiques pour élèves allophones arrivants (UPE2A) où ils reçoivent un enseignement renforcé en français, langue seconde, en fonction de leurs besoins et de leurs acquis linguistiques et langagiers. Ils bénéficient notamment d’un soutien logistique, tout en étant inclus, en parallèle, dans une classe ordinaire. L’objectif est qu’ils s’insèrent progressivement et le plus rapidement possible dans une classe ordinaire.

### **1.4.2 Hétérogénéité des profils**

La venue des EANA en France peut être motivée par diverses raisons, d’ordre politique, économique, familial ou professionnel. À l’hétérogénéité induite par le type de migration s’ajoute l’importance des variations inter-individuelles au sein d’une même population (Geva et al., 2019).

Les mêmes auteurs dressent une liste non exhaustive des facteurs pouvant contribuer aux difficultés d’apprentissage scolaire de certains apprenants d’une langue seconde tels l’âge d’apprentissage, le degré d’exposition à la L2, la qualité de l’instruction, l’accès aux ressources...

Le Casnav de Lille (2021) constate : « L’inclusion des enfants allophones (...) en classe ordinaire est particulièrement délicate en raison du décalage culturel, linguistique et scolaire, notamment lorsque les conditions de vie en France entraînent un absentéisme important (familles sans ressources vivant en habitat précaire, etc.) »

Comme le font remarquer Kirby et al. (2008), il est difficile d’isoler la contribution des facteurs socio-économiques et psycho-sociaux dans l’apprentissage de la lecture, et donc d’intervenir sur ces paramètres pour agir sur cette compétence. Ceux-ci exercent cependant sur les mécanismes cognitifs sous-jacents de l’apprentissage de la lecture une influence qui risque de retentir sur les performances des EANA dans les tâches mettant en jeu ces paramètres. Il est donc important de contrôler ces variables, notamment le niveau d’étude des parents, quand on entend comparer leurs performances à celles des enfants contrôles, dans ce domaine.

Suivant leur âge et leur pays d'origine, les EANA peuvent ou non avoir été scolarisés et avoir commencé à apprendre à lire et à écrire avant leur arrivée en France. Ce critère pourra avoir un impact sur le niveau scolaire, en particulier dans le domaine qui nous intéresse, sur l'acculturation à l'écrit et la compréhension du code alphabétique (Rafoni, 2015).

Les EANA qui ont été partiellement ou normalement scolarisés ont acquis des compétences qu'ils vont pouvoir transférer au français. En tant que lecteurs dans une écriture alphabétique ou non alphabétique, ils doivent s'approprier de nouvelles conduites d'apprentissage, apprendre les conventions qui régissent les caractères latins et les sons du français mais n'ont pas à réapprendre à lire.

Le Casnav de Lille (2021) rapporte cependant qu'un nombre croissant d'élèves allophones intègrent l'école sans avoir bénéficié d'une scolarisation ou pu suivre une scolarité régulière dans leur pays d'origine. Selon Rafoni (2015), cela induit des différences importantes dans le rythme d'acquisition des élèves de CLIN (Classe d'Initiation pour Non-francophones remplacée par l'UPE2A) : faute d'une immersion préalable dans des activités épilinguistiques, les enfants n'ont pas acquis les « réflexes de distance par rapport à la langue » que leurs camarades natifs expérimentent dans des jeux depuis plusieurs années. Le même auteur explique que les manipulations linguistiques trop fines (sur les phonèmes, les mots) sont impossibles dans une langue en cours d'acquisition : il existe un seuil au-delà duquel toute tentative de discrimination deviendrait, selon lui, inefficace. À cela s'ajoutent les nombreuses interférences avec la langue d'origine, qui font obstacle à l'identification des phonèmes.

Les apprentis lecteurs de langue seconde issus de milieux vulnérables sont un groupe à risque parce que leurs habiletés en conscience phonologique, décodage et reconnaissance de mots, qui sous-tendent la lecture, sont moins développées que celles de leurs pairs natifs.

En effet, ils apprennent à les développer en L2 sans avoir bénéficié d'un enseignement scolaire équivalent dans leur L1 (Erdos et al., 2010). C'est également le cas des EANA.

Selon le Casnav de Lille (2021), les EANA maîtrisent une ou plusieurs langues (langue maternelle, langue véhiculaire, langue[s] de scolarisation...). Certains jeunes sont issus de familles dont la culture orale est prédominante, d'autres sont originaires de pays dont les langues vernaculaires ne sont pas écrites (arabes dialectaux, langues africaines, etc.). Pour chacune des langues, le niveau de maîtrise et de compétences peut donc être très variable.

À leur arrivée en France, les EANA doivent apprendre le français comme une langue étrangère. Selon Grabe (2009), les apprenants de L2 atteignent rarement un niveau de lecture de mots à voix haute comparable à celui de leurs pairs natifs car contrairement à eux, ils doivent développer simultanément leur compréhension et ne peuvent par conséquent consacrer l'essentiel de leurs ressources cognitives au décodage.

Certains EANA maîtrisent partiellement le français, parce qu'ils le parlent en famille ou qu'il s'agit de la langue seconde du pays d'origine. Cependant, selon Rafoni (2015), une maîtrise rudimentaire du français oral en situation de communication courante est un prérequis pour entrer dans l'écrit, mais en aucun cas annonciatrice d'une réussite en lecture.

Les caractéristiques de la langue d'origine – alphabétique ou non alphabétique, à caractères latins ou non latins – ont une incidence sur la rapidité d'apprentissage du français oral, et surtout écrit. L'hypothèse contrastive permet d'anticiper les sources d'erreurs pour les apprenants en raison des transferts positifs et/ou négatifs qui s'opèrent entre la L1 et la L2.

Geva et al. (2019) recommandent de se montrer prudent dans l'interprétation des performances des élèves bilingues aux tests standardisés, normés auprès de populations plus homogènes. Les auteurs préconisent de les comparer à une population équivalente. Cependant,

au vu de la diversité des profils des EANA, dont nous venons de voir un aperçu, il est difficile d'établir des normes permettant de situer leurs performances.

## **2 Buts**

On connaît les enjeux cruciaux de la lecture dans le monde contemporain, en termes d'accès au savoir, d'égalité des chances et, au-delà, d'épanouissement personnel. La lecture, en particulier la reconnaissance des mots écrits, fait l'objet d'un enseignement scolaire explicite, secondaire à l'acquisition du langage oral. La compréhension du texte lu nécessite cependant d'importantes ressources cognitives, qui ne peuvent être recrutées qu'une fois la reconnaissance des mots écrits automatisés. Nous nous proposons donc d'investiguer l'influence que peut avoir l'étendue du vocabulaire sur ce processus de bas niveau.

Dans ce contexte, notre question de recherche peut par conséquent être formulée ainsi : « Le vocabulaire est-il un prédicteur de la lecture de mots à voix haute chez les EANA ? » Il nous semble intéressant d'explorer cette question à la lumière des performances en lecture de mots des EANA, dont le stock lexical est par nature réduit. Des éléments de réponse à cette question pourraient permettre de mieux cibler les interventions pédagogiques conçues pour cette population.

## **3 Hypothèse**

Nous nous attendons à ce que l'étendue du vocabulaire réceptif (nombre de mots connus) des EANA soit un facteur plus déterminant dans la réussite en lecture, dans ce groupe, que chez leurs homologues tout-venant. En effet, selon une étude réalisée par Bowey (2005), la compréhension de lecture et la reconnaissance de mots partagent 61 à 81 % de variance chez le lecteur débutant (tout-venant). Cela reflète le fait qu'à ce stade, les choix didactiques se portent sur des mots simples et des phrases canoniques basées sur du vocabulaire connu. Or, le vocabulaire de base d'un élève tout-venant peut être source de difficulté pour un élève allophone.

## **Méthode**

### **1 Procédure**

L'étude menée dans le cadre du présent mémoire s'inscrit dans un projet de recherche plus large, baptisé TANMALL, mené par Matthieu Bignon, orthophoniste, dans le cadre de sa thèse de doctorat. Le protocole de ce projet a été validé par le Comité d'Éthique de la faculté de Psychologie de Lille (référence : 2021-466-S90) et le recueil de données est déclaré à la CNIL. Il s'agit d'une étude transversale multicentrique, réalisée à l'échelle de plusieurs académies.

Les passations se sont déroulées avec l'accord des enseignants d'UPE2A et des directeurs concernés, après recueil de l'autorisation des inspecteurs de l'éducation nationale chargés des circonscriptions du 1er degré. Les parents des élèves inclus dans l'étude ont préalablement reçu une lettre d'information détaillant les objectifs et les modalités de l'étude et

tous ont donné leur consentement éclairé à la participation de leur enfant. Les tests avaient été sélectionnés pour évaluer très succinctement le fonctionnement cognitif général, ainsi que le langage oral (phonologie, vocabulaire réceptif) et le langage écrit (lecture de mots et de pseudo-mots). À l'exception des Matrices de Raven et de l'épreuve de DRA, ils ont été administrés sous forme informatisée avec cotation informatisée. Des séances individuelles d'une durée d'une heure à une heure vingt, parfois entrecoupées de pause (récréation, pause de midi) en fonction de la motivation et des réserves attentionnelles de l'enfant ont été organisées à l'école, sur le temps scolaire voire périscolaire. Les réponses orales des élèves aux épreuves visant à évaluer leurs compétences phonologiques ont été enregistrées pour permettre de vérifier l'encodage.

## 2 Population

Nous avons bénéficié, pour notre étude, de l'ensemble des données recueillies et traitées dans le cadre du projet TANMALL. Le Tableau 1 ci-dessous donne une vision synoptique des caractéristiques de l'échantillon.

**Tableau 1 : Caractéristiques de l'échantillon**

### *Âge et scolarisation*

| Variables   | Tout-venant (N=231)  |         | Allophones (N=157)   |         | Différences entre les groupes |        |       |            |
|---|----------------------|---------|----------------------|---------|-------------------------------|--------|-------|------------|
|   | Moyenne (écart-type) | Min-Max | Moyenne (écart-type) | Min-Max | t*                            | df     | p     | d de Cohen |
| Âge (lors de la passation)                                  | 86.52 (6.33)         | 77-110  | 113.11 (19.69)       | 75-191  | -16.36                        | 178.07 | <.001 | -1.82      |
| Nombre de mois en primaire (exposition explicite à l'écrit) | 12.22 (5.13)         | 5-31    | 13.34 (6.88)         | 6-31    | -1.74                         | 270.04 | 0.08  | -0.18      |
| Niveau scolaire numérique                                   | 1.33 (0.47)          | 1-2     | 3.05 (1.37)          | 1-5     | –                             | –      | –     | –          |
| Temps de séjour en France                                   | –                    | –       | 16.24 (8.06)         | 6-38    | –                             | –      | –     | –          |
| Temps de scolarisation en France                            | –                    | –       | 14.68 (7.26)         | 6-37    | –                             | –      | –     | –          |
| Âge lors de la scolarisation en France                      | –                    | –       | 98.42 (20.81)        | 47-181  | –                             | –      | –     | –          |

\* test t de Welch car les variances n'étaient pas homogènes.

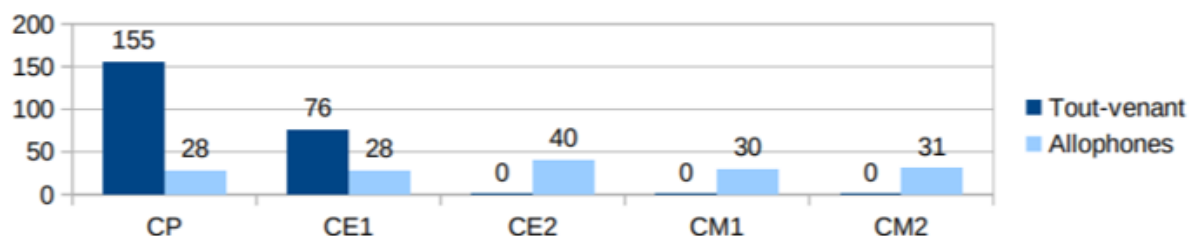
Afin de garantir la meilleure représentativité possible de la population et le maximum de puissance statistique, compte tenu de l'hétérogénéité des profils, un vaste échantillon de 157 EANA a été recruté pour ce projet. Nous avons personnellement effectué 35 passations dans 7 écoles élémentaires de 7 circonscriptions d'inspection du premier degré de Paris. 3 des participants que nous avons rencontrés ont cependant été exclus de l'étude en raison de données incomplètes (erreurs de passation, abandon).

Les élèves inclus dans le dispositif ne parlaient pas français avant leur arrivée en France et étaient scolarisés en élémentaire depuis au moins 6 mois (critère strict) et, si possible, pas plus de 2 ans (critère souple : max. 3 ans), soit en moyenne 15 mois.

Leur temps d'exposition au langage oral, mesuré indirectement à travers leur temps de scolarisation en France, soit en moyenne 16 mois, était sensiblement identique à leur temps d'exposition au langage écrit, de 13 mois en moyenne.

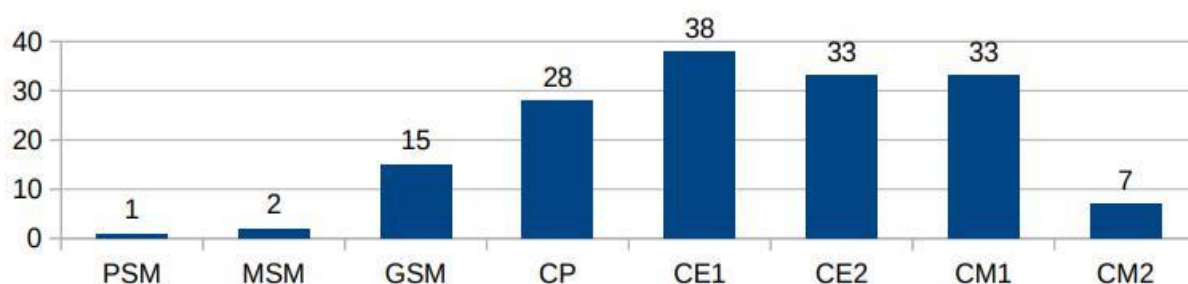


Les EANA étaient beaucoup plus âgés en moyenne que leurs homologues natifs ( $d = -1,82$ ) : par rapport à ces derniers, ils avaient commencé l'apprentissage de la lecture en français avec en moyenne 2 ans de retard (à l'âge de 8 ans). Comme le montre la Figure 3 ci-dessous, seuls 35 % des EANA étaient scolarisés en CP ou CE1, contre 100 % des élèves tout-venant. Par contre, le temps d'exposition à l'écrit de chacun des deux groupes est relativement similaire.



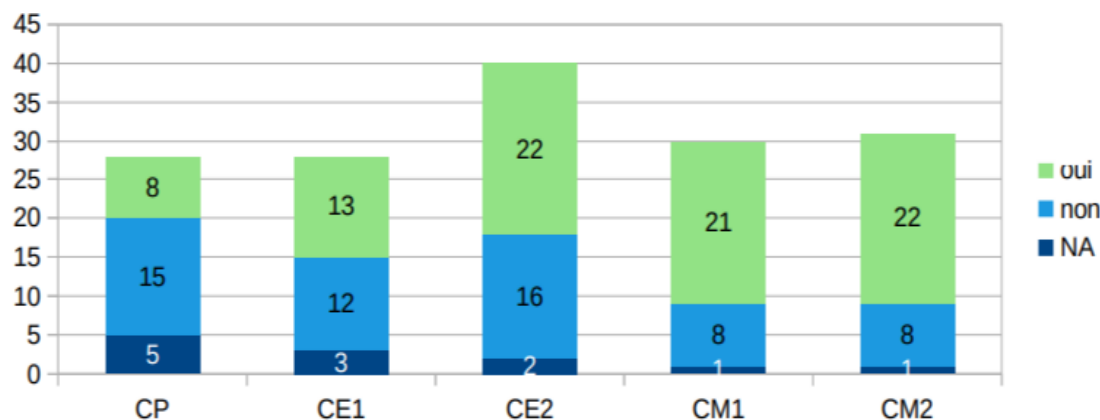
**Fig. 3 Répartition des participants par niveau de classe et par groupe (élèves tout-venant et allophones)**

La Figure 4 ci-dessous montre que seul 11 % de l'effectif des EANA avait bénéficié de la stimulation du langage oral et de l'acculturation à l'écrit dispensée aux élèves tout-venant en maternelle, en France.



**Fig. 4 : Nombre d'enfants allophones par niveau scolaire de première scolarisation en France**

Plus de la moitié cependant (55 %) avaient commencé l'apprentissage de la lecture dans leur langue maternelle, avant leur arrivée en France (cf. Figure 5 ci-dessous).



**Fig. 5 : Nombre d'enfants allophones par niveau scolaire ayant au moins commencé à apprendre à lire avant d'arriver en France**

Les EANA présentaient également une grande diversité de langues maternelles : 42 langues maternelles principales selon la Figure 6 ci-dessous, dont seulement deux langues latines, la plus représentée étant l'arabe (20 %).

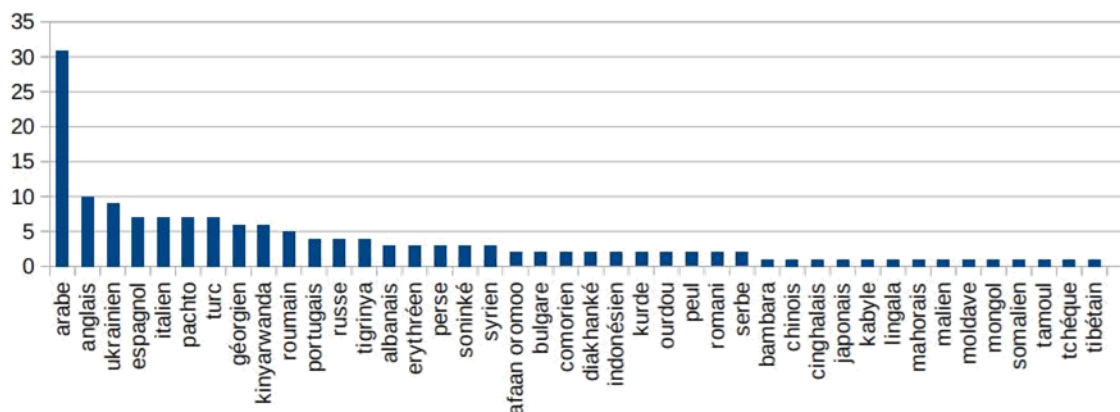


Fig. 6 : Nombre d'enfants allophones par langue maternelle principale

Leurs langues maternelles secondaires, soit 25 au total, dont trois langues latines, sont recensées à la Figure 7 ci-dessous, l'anglais étant la plus représentée (25 %).

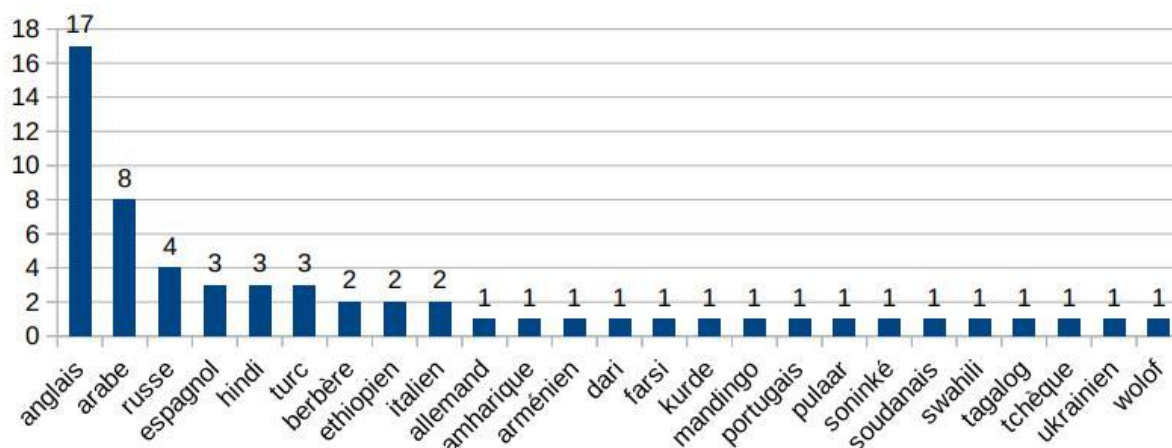


Fig. 7 : Nombre d'enfants allophones par langue secondaire

## 3 Matériel

### 3.1 Evaluation des compétences

Nous avons repris les épreuves déjà sélectionnées dans le cadre du projet TANMALL, dans lequel s'inscrit la présente étude.

#### 3.1.1 Mesure des compétences cognitives sous-jacentes

Les **matrices progressives de Raven couleur (PM 47, Raven, 1938)**, conçues pour des enfants de 4 à 11; 6 ans ont été utilisées comme test contrôle, pour mesurer le **niveau cognitif non verbal**. Elles sont composées de trois séries de 12 items. L'élève doit parvenir à donner un sens aux éléments de chaque figure afin de la compléter correctement, sans limite de temps, à l'aide de l'une des 6 propositions présentées. La cotation s'est faite manuellement. 1 point a été attribué par bonne réponse, pour un score total de 36.

### 3.1.2 Mesure des compétences phonologiques sous-jacentes

La tâche de **répétition de pseudo-mots tirée de LITMUS-NWR-31** (Ferré & Santos, 2020) a été utilisée pour mesurer la mémoire phonologique à court terme, puisque répéter des pseudo-mots implique de stocker temporairement l'information phonologique.

La contribution significative et indépendante de la mémoire de travail verbale et des représentations phonologiques à une tâche de répétition de pseudo-mots a été établie scientifiquement (Ferré & Santos, 2020). L'analyse des performances à cette tâche a également permis d'isoler des données relatives aux compétences phonologiques sous-jacentes à la reconnaissance des mots écrits des participants. La cotation était informatisée (1 point attribué par bonne réponse).

Le test de **DRA (version papier continue)** avait été choisi pour mesurer l'automatisation de la mise en relation entre la représentation visuelle et la représentation phonologique.

Il a été demandé à l'élève de dénommer successivement 48 chiffres puis 48 lettres le plus rapidement possible (tâches chronométrées) sans faire d'erreurs. Les stimuli respectifs étaient présentés sous la forme d'un tableau à lire ligne par ligne. La cotation s'est faite manuellement. Le score d'erreurs correspondant au nombre d'items ratés et le score d'auto-corrections correspondant au nombre d'erreurs corrigées ont été relevés.

Les tâches de **1) suppression de syllabe initiale, 2) suppression de phonème initial dans un cluster CVC et 3) suppression de phonème initial dans un cluster CCV** issues de la batterie informatisée **EVALEC (Sprenger-Charolles et al., 2010)** avaient été choisies pour explorer le traitement métaphonologique, autrement dit la conscience syllabique et la conscience phonémique. La cotation était informatisée (1 point attribué par bonne réponse).

### 3.1.3 Mesure du vocabulaire réceptif

L'étendue du vocabulaire réceptif, qui correspond au nombre de mots connus, a été évaluée à l'aide d'une **épreuve de désignation d'images basée sur l'EVIP (échelle standardisée de vocabulaire en image du Peabody (adaptation française de Dunn, Thiéroult-Whalen et Dunn, 1993))**. La tâche ainsi conçue utilise une sous-sélection d'items et la procédure d'administration a été simplifiée pour raccourcir le temps de passation. L'élève devait sélectionner l'image correspondant à chaque mot qu'il entendait, parmi quatre. La cotation était informatisée (1 point accordé par bonne réponse pour un score total de 50 ; arrêt automatique au bout de 5 mauvaises réponses d'affilée).

### 3.1.4 Mesure de la précision de la capacité de lecture de mots à voix haute

Trois tâches tirées d'EVALEC (Batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture, 2010) avaient été choisies pour évaluer la précision de la lecture de mots à voix haute.

La tâche de lecture de mots visait à évaluer la lecture de mots à voix haute. La cotation était informatisée (1 point attribué par bonne réponse).

La tâche de lecture de pseudo-mots visait à évaluer les habiletés de décodage par conversion grapho-phonémique. Les pseudo-mots étant par définition inconnus du lecteur, cette tâche garantit en effet l'utilisation exclusive d'une procédure de recodage phonologique. Le

test comporte 36 pseudo-mots bisyllabiques. La cotation était informatisée (1 point attribué pour chaque pseudo-mot correctement décodé).

## Résultats

Nous nous sommes basés uniquement sur la cotation souple proposée par le concepteur du projet TANMALL. Selon cette cotation, un écart d'un trait par rapport aux règles phonéto-phonologiques du français a été toléré. L'acceptation de ces distorsions a pour but de limiter l'impact des aspects phonologiques de la langue maternelle sur les scores. Comme l'expliquent en effet Bignon et al. (2022), le risque de la cotation stricte était d'imputer à une faiblesse de décodage des erreurs dues à l'impossibilité de percevoir certains contrastes absents du répertoire phonologique de la langue maternelle. De notre point de vue, l'application de cette cotation souple se justifie également par les possibles différences de perception entre locuteurs d'une même langue de diverses origines géographiques et partant, entre les examinateurs chargés de recueillir les données.

## 1 Performances aux tests

Comme le montre le Tableau 2 ci-dessous, nous observons une différence de performances significative et plus ou moins importante ( $0.22 < d < 1.8$  ;  $p < .001$ ) entre les deux groupes, aussi bien aux tâches prédictives de la lecture (excepté la DRA et la répétition de pseudo-mots) qu'à celles qui mesurent les habiletés en lecture de mots à voix haute et en décodage. Sans surprise, une forte différence entre les groupes est observée dans l'étendue du vocabulaire réceptif ( $d = 1.86$ ). L'écart entre les deux groupes est plus marqué à la tâche de lecture de mots à voix haute ( $d = 1.25$ ) qu'à la tâche de lecture à voix haute de pseudo-mots ( $d = 1.06$ ).

**Tableau 2 : Scores aux différentes épreuves**

| Variable                                  | Tout-venant (N=231)  |         | Allophones (N=157)   |         | Différences entre les groupes |        |       |            |
|---|----------------------|---------|----------------------|---------|-------------------------------|--------|-------|------------|
|   | Moyenne (écart-type) | Min-Max | Moyenne (écart-type) | Min-Max | t*                            | df     | p     | d de Cohen |
| <b>Prédicteurs</b>                        |                      |         |                      |         |                               |        |       |            |
| Matrices de Raven                         | 24.76 (5.33)         | 7-36    | 22.97 (6.18)         | 6-34    | 2.95                          | 301.30 | <.001 | 0.31       |
| Conscience phonologique (stricte)         | 26.50 (5.64)         | 3-34    | 21.80 (8.68)         | 0-34    | 5.98                          | 244.60 | <.001 | 0.64       |
| Conscience phonologique (souple)          | 26.73 (5.59)         | 4-34    | 22.94 (8.78)         | 0-34    | 4.79                          | 241.33 | <.001 | 0.51       |
| Dénomination rapide (temps moyen en sec.) | 32.97 (8.06)         | 18,5-79 | 33.31 (10.91)        | 17,5-82 | -0.34                         | 268.14 | 0.74  | -0.04      |
| Répétition de pseudo-mots (stricte)       | 26.59 (3.10)         | 13-31   | 24.22 (5.08)         | 4-31    | 5.22                          | 234.90 | <.001 | 0.56       |
| Répétition de pseudo-mots (souple)        | 26.94 (3.01)         | 14-31   | 26.15 (4.18)         | 4-31    | 2.04                          | 262.93 | 0.04  | 0.22       |
| Vocabulaire                               | 30.90 (6.41)         | 13-46   | 18.09 (7.76)         | 0-35    | 17.11                         | 291.54 | <.001 | 1.80       |
| <b>Tâches de lecture</b>                  |                      |         |                      |         |                               |        |       |            |
| Lecture de mots (stricte)                 | 39.69 (7.21)         | 11-48   | 25.45 (12.20)        | 0-46    | 13.15                         | 230.03 | <.001 | 1.42       |
| Lecture de mots (souple)                  | 40.54 (6.74)         | 14-48   | 28.34 (12.02)        | 0-47    | 11.55                         | 222.90 | <.001 | 1.25       |
| Lecture de pseudo-mots (stricte)          | 26.84 (5.57)         | 9-36    | 16.64 (8.77)         | 0-34    | 12.90                         | 240.95 | <.001 | 1.39       |
| Lecture de pseudo-mots (souple)           | 27.94 (5.44)         | 10-36   | 20.22 (8.72)         | 0-34    | 9.87                          | 238.28 | <.001 | 1.06       |

\* test t de Welch car les variances n'étaient pas homogènes.

## 2 Modélisation des relations entre les variables

### 2.1 Matrices de corrélation

**Tableau 3 : Matrice de corrélation variables démographiques et scores aux épreuves pour les allophones**

|  | 1        | 2       | 3        | 4        | 5        | 6        | 7        | 8        | 9        | 10       | 11       | 12      | 13      | 14      |
|--|----------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|---------|---------|---------|
| 1. Âge (lors de la passation)                                  | —        |         |          |          |          |          |          |          |          |          |          |         |         |         |
| 2. Nombre de mois en primaire (exposition explicite à l'écrit) | 0.26**   | —       |          |          |          |          |          |          |          |          |          |         |         |         |
| 3. Temps de scolarisation en France                            | 0.03     | 0.83*** | —        |          |          |          |          |          |          |          |          |         |         |         |
| 4. Temps de séjour en France                                   | 0.04     | 0.76*** | 0.93***  | —        |          |          |          |          |          |          |          |         |         |         |
| 5. Âge lors de la scolarisation en France                      | 0.93***  | 0.18*   | -0.03    | -0.01    | —        |          |          |          |          |          |          |         |         |         |
| 6. Niveau scolaire numérique                                   | 0.70***  | 0.31*** | 0.05     | 0.07     | 0.60***  | —        |          |          |          |          |          |         |         |         |
| 7. Scolarisation antérieure                                    | 0.24**   | -0.03   | -0.22**  | -0.28*** | 0.22**   | 0.32***  | —        |          |          |          |          |         |         |         |
| 8. Apprentissage de la lecture antérieur à l'arrivée en France | 0.15     | -0.03   | -0.14    | -0.11    | 0.18*    | 0.26**   | 0.50***  | —        |          |          |          |         |         |         |
| 9. Matrices de Raven   | 0.12     | 0.11    | 0.01     | 4.89e-3  | 0.09     | 0.32***  | 0.32***  | 0.41***  | —        |          |          |         |         |         |
| 10. Conscience phonologique (souple)                           | 0.14     | 0.10    | 0.02     | 0.02     | 0.13     | 0.35***  | 0.39***  | 0.36***  | 0.46***  | —        |          |         |         |         |
| 11. Dénomination rapide (temps en sec.)                        | -0.34*** | -0.15   | 0.09     | 0.12     | -0.32*** | -0.52*** | -0.45*** | -0.33*** | -0.38*** | -0.54*** | —        |         |         |         |
| 12. Répétition de pseudo-mots (souple)                         | 0.17*    | 0.11    | -2.59e-3 | -0.03    | 0.18*    | 0.32***  | 0.42***  | 0.32***  | 0.35***  | 0.54***  | -0.32*** | —       |         |         |
| 13. Vocabulaire  | 0.19*    | 0.30*** | 0.24**   | 0.22**   | 0.13     | 0.22**   | 0.05     | 0.12     | 0.29***  | 0.22**   | -0.25**  | 0.16    | —       |         |
| 14. Lecture de mots (souple)                                   | 0.24**   | 0.15    | 0.04     | 0.04     | 0.23**   | 0.45***  | 0.35***  | 0.41***  | 0.55***  | 0.71***  | -0.67*** | 0.44*** | 0.40*** | —       |
| 15. Lecture de pseudo-mots (souple)                            | 0.14     | 0.05    | -0.04    | -0.06    | 0.15     | 0.33***  | 0.34***  | 0.35***  | 0.48***  | 0.63***  | 0.59***  | 0.42*** | 0.28*** | 0.89*** |

\* p < .05, \*\* p < .01, \*\*\* p < .001

Le Tableau 3 ci-dessus montre que chez les EANA, le vocabulaire est moyennement corrélé à la réussite en lecture, plus fortement à la lecture de mots à voix haute ( $r = 0.4$ ,  $p < .001$ ) qu'au décodage ( $r = 0.28$ ,  $p < .001$ ). Parmi les facteurs non cognitifs de la réussite en lecture qui ont été testés, l'exposition au français oral, mesurée indirectement à travers le temps de séjour en France, est moyennement corrélée au niveau de vocabulaire ( $r = 0.22$ ,  $p < .001$ ) mais sans incidence sur le niveau de lecture. Inversement, comme on pouvait s'y attendre, le fait d'avoir appris à lire une autre langue avant l'arrivée en France, qui influe moyennement sur le décodage ( $r = 0.35$ ,  $p < .001$ ) et la lecture à voix haute de mots ( $r = 0.41$ ,  $p < .001$ ), est sans lien avec l'étendue du vocabulaire réceptif en français. On peut en dire autant du fait d'avoir commencé sa scolarité avant l'arrivée en France. Parmi les facteurs phonologiques modérément corrélés à la réussite en lecture ( $0.42 < r = 0.44$ ,  $p < .001$ ), la boucle phonologique s'avère sans lien avec le niveau de vocabulaire.

**Tableau 4 : Matrices de corrélation variables démographiques et scores aux épreuves pour les élèves tout-venant**

|  | 1        | 2       | 3       | 4        | 5        | 6        | 7       | 8       | 9       |
|--|----------|---------|---------|----------|----------|----------|---------|---------|---------|
| 1. Âge (lors de la passation)                                  | —        |         |         |          |          |          |         |         |         |
| 2. Nombre de mois en primaire (exposition explicite à l'écrit) | 0.83***  | —       |         |          |          |          |         |         |         |
| 3. Niveau scolaire numérique                                   | 0.78***  | 0.94*** | —       |          |          |          |         |         |         |
| 4. Matrices de Raven   | 0.17**   | 0.12    | 0.09    | —        |          |          |         |         |         |
| 5. Conscience phonologique (souple)                            | -0.03    | -0.07   | -0.06   | 0.39***  | —        |          |         |         |         |
| 6. Dénomination rapide (temps en sec.)                         | -0.22*** | -0.18** | -0.21** | -0.22*** | -0.27*** | —        |         |         |         |
| 7. Répétition de pseudo-mots (souple)                          | 0.07     | 0.06    | 0.06    | 0.16*    | 0.39***  | -0.30*** | —       |         |         |
| 8. Vocabulaire   | 0.11     | 0.03    | -0.01   | 0.35***  | 0.29***  | -0.15*   | 0.20**  | —       |         |
| 9. Lecture de mots (souple)                                    | 0.16*    | 0.17**  | 0.14*   | 0.37***  | 0.54***  | -0.44*** | 0.45*** | 0.36*** | —       |
| 10. Lecture de pseudo-mots (souple)                            | 0.10     | 0.11    | 0.09    | 0.28***  | 0.43***  | -0.46*** | 0.36*** | 0.21**  | 0.79*** |

\* p < .05, \*\* p < .01, \*\*\* p < .001

Comme l'indique le Tableau 4 ci-dessus, chez les élèves tout-venant, les variables non cognitives (âge, classe, exposition explicite à l'écrit) n'ont pas d'incidence sur l'étendue du vocabulaire réceptif. Parmi les variables corrélées à la lecture chez ces élèves, sont aussi corrélées au vocabulaire, par ordre décroissant, l'intelligence non verbale ( $r = 0.35$ ,  $p < .001$ ) et les facteurs cognitifs phonologiques : la conscience phonologique ( $r = 0.29$ ,  $p < .001$ ), la boucle phonologique ( $r = 0.2$ ,  $p < .01$ ), et dans une moindre mesure, la DRA ( $r = - 0.15$ ,  $p < .05$ ).

Pour tester la contribution unique du vocabulaire à la lecture, il faut donc contrôler toutes ces variables dans un modèle de régressions linéaires.

## 2.2 Modèles de régressions linéaires

Afin d'isoler la contribution unique du vocabulaire à la lecture de mots et de pseudo-mots dans chacun des deux groupes, un modèle réunissant les différentes variables corrélées au vocabulaire a été élaboré.

Seules les variables dont le lien avec le vocabulaire avait été mis en évidence par les matrices de corrélation, dans chacun des deux groupes, ont été intégrées dans ce modèle. Le temps de scolarisation en France, le temps de séjour en France, l'âge d'entrée à l'école en France, la scolarisation et l'apprentissage de la lecture avant l'arrivée en France ont donc été écartés. La répétition de pseudo-mots a été intégrée dans le modèle malgré son absence de corrélation avec le vocabulaire chez les EANA, afin de garantir l'homogénéité des modèles.

Le modèle explique une part plus importante de la variance chez les EANA que chez les élèves tout-venant, en lecture de mots à voix haute (68 % contre 49 %) et en décodage (53 % contre 34 %). Il explique une part plus importante de la variance en lecture de mots à voix haute qu'en décodage, chez les EANA (68 % contre 53 %) et chez les élèves tout-venant (49 % contre 34 %).

Le vocabulaire explique une part plus élevée de la variance en lecture de mots à voix haute chez les EANA (2.56 % ( $p < .001$ )) que chez les élèves tout-venant (1.69 % ( $p < .001$ )). Par contre, la contribution du vocabulaire au décodage n'est significative dans aucun des deux groupes.

Sans surprise, les variables phonologiques expliquent une part nettement plus importante de la variance en lecture de mots que le vocabulaire, aussi bien chez les EANA (excepté pour ce qui est de la boucle phonologique) que chez les élèves tout-venant. On relève ainsi les pourcentages suivants respectivement chez les EANA : conscience phonologique 7,84 %, DRA 6,25 %, contre 2,56 % pour le vocabulaire et chez les élèves tout-venant : conscience phonologique 7,84 %, DRA 4,41 %, boucle phonologique 3,61 % contre 1,69 % pour le vocabulaire.

**Tableau 5 : Modèle de régressions linéaires appliqué aux élèves tout-venant**

**Lecture de mots**

F(7,222)=29.95, p<.001, R<sup>2</sup> = .49

| Variables   | NS    | 95 % CI |       | S     | t     | p       | ΔR <sup>2</sup> |
|---|-------|---------|-------|-------|-------|---------|-----------------|
|   |       | Bas     | Haut  |       |       |         |                 |
| (Intercept)   | 21.22 | 5.42    | 37.01 |       | 2.65  | 8.70e-3 |                 |
| Nombre de mois en primaire (exposition explicite à l'écrit) | -0.10 | -0.28   | 0.08  | -0.10 | -1.09 | 0.28    | 0.25%           |
| Âge (lors de la passation)                                  | 0.27  | 0.04    | 0.49  | 0.20  | 2.35  | 0.02    | 1.21%           |
| Matrices de Raven   | 0.12  | -0.01   | 0.26  | 0.10  | 1.77  | 0.08    | 0.81%           |
| Conscience phonologique (souple)                            | 0.40  | 0.26    | 0.54  | 0.33  | 5.72  | < .001  | 7.84%           |
| Dénomination rapide (temps en sec.)                         | -0.19 | -0.28   | -0.10 | -0.23 | -4.35 | < .001  | 4.41%           |
| Répétition de pseudo-mots (souple)                          | 0.47  | 0.23    | 0.71  | 0.21  | 3.88  | < .001  | 3.61%           |
| Vocabulaire   | 0.16  | 0.05    | 0.27  | 0.15  | 2.78  | 5.86e-3 | 1.69%           |

Note. NS = coefficient de pente non standardisé ; S = coefficient de pente standardisé.

**Lecture de pseudo-mots**

F(7,222)=16.65, p<.001, R<sup>2</sup>= .34

| Variables   | NS    | 95 % CI |       | S     | t     | p      | ΔR <sup>2</sup> |
|---|-------|---------|-------|-------|-------|--------|-----------------|
|   |       | Bas     | Haut  |       |       |        |                 |
| (Intercept)   | 25.99 | 11.55   | 40.42 |       | 3.55  | < .001 |                 |
| Nombre de mois en primaire (exposition explicite à l'écrit) | -0.11 | -0.28   | 0.06  | -0.13 | -1.28 | 0.20   | 0.49%           |
| Âge (lors de la passation)                                  | 0.16  | -0.04   | 0.37  | 0.15  | 1.58  | 0.11   | 0.81%           |
| Matrices de Raven   | 0.08  | -0.05   | 0.20  | 0.08  | 1.21  | 0.23   | 0.49%           |
| Conscience phonologique (souple)                            | 0.24  | 0.11    | 0.36  | 0.24  | 3.75  | < .001 | 4.00%           |
| Dénomination rapide (temps en sec.)                         | -0.22 | -0.30   | -0.14 | -0.33 | -5.49 | < .001 | 9.00%           |
| Répétition de pseudo-mots (souple)                          | 0.26  | 0.04    | 0.47  | 0.14  | 2.31  | 0.02   | 1.69%           |
| Vocabulaire   | 0.05  | -0.05   | 0.15  | 0.06  | 0.93  | 0.35   | 0.25%           |

Note. NS = coefficient de pente non standardisé ; S = coefficient de pente standardisé.

**Tableau 6 : Modèle de régressions linéaires appliqué aux élèves allophones**

**Lecture de mots**

F(7,149)=45.37, p<.001, R<sup>2</sup> = .68

| Variables   | NS    | 95 % CI |       | S     | t     | p      | ΔR <sup>2</sup> |
|---|-------|---------|-------|-------|-------|--------|-----------------|
|   |       | Bas     | Haut  |       |       |        |                 |
| (Intercept)   | 11.65 | -0.91   | 24.20 |       | 1.83  | 0.07   |                 |
| Nombre de mois en primaire (exposition explicite à l'écrit) | -0.03 | -0.20   | 0.14  | -0.02 | -0.34 | 0.73   | 0.04%           |
| Âge (lors de la passation)                                  | 0.01  | -0.05   | 0.07  | 0.02  | 0.43  | 0.67   | 0.04%           |
| Matrices de Raven   | 0.37  | 0.16    | 0.58  | 0.19  | 3.52  | < .001 | 2.56%           |
| Conscience phonologique (souple)                            | 0.54  | 0.36    | 0.71  | 0.39  | 6.08  | < .001 | 7.84%           |
| Dénomination rapide (temps en sec.)                         | -0.36 | -0.49   | -0.23 | -0.32 | -5.47 | < .001 | 6.25%           |
| Répétition de pseudo-mots (souple)                          | 0.07  | -0.25   | 0.38  | 0.02  | 0.41  | 0.68   | 0.04%           |
| Vocabulaire   | 0.27  | 0.11    | 0.42  | 0.17  | 3.40  | < .001 | 2.56%           |

Note. NS = coefficient de pente non standardisé ; S = coefficient de pente standardisé.

**Lecture de pseudo-mots**

F(7,149)=23.65, p<.001, R<sup>2</sup>= .53

| Variables   | NS    | 95 % CI |       | S     | t     | p      | ΔR <sup>2</sup> |
|---|-------|---------|-------|-------|-------|--------|-----------------|
|   |       | Bas     | Haut  |       |       |        |                 |
| (Intercept)   | 13.80 | 2.71    | 24.89 |       | 2.46  | 0.02   |                 |
| Nombre de mois en primaire (exposition explicite à l'écrit) | -0.09 | -0.24   | 0.06  | -0.07 | -1.19 | 0.24   | 0.49%           |
| Âge (lors de la passation)                                  | -0.02 | -0.08   | 0.03  | -0.05 | -0.85 | 0.40   | 0.25%           |
| Matrices de Raven   | 0.22  | 0.04    | 0.41  | 0.16  | 2.38  | 0.02   | 1.69%           |
| Conscience phonologique (souple)                            | 0.32  | 0.17    | 0.48  | 0.33  | 4.15  | < .001 | 5.29%           |
| Dénomination rapide (temps en sec.)                         | -0.26 | -0.38   | -0.15 | -0.33 | -4.58 | < .001 | 6.76%           |
| Répétition de pseudo-mots (souple)                          | 0.17  | -0.11   | 0.45  | 0.08  | 1.21  | 0.23   | 0.49%           |
| Vocabulaire   | 0.11  | -0.03   | 0.25  | 0.10  | 1.61  | 0.11   | 0.81%           |

Note. NS = coefficient de pente non standardisé ; S = coefficient de pente standardisé.

## Discussion et Conclusion

Un enjeu majeur de l'enseignement du premier degré est de préparer, puis d'installer l'entrée dans l'écrit, base de l'accès au savoir. Lorsqu'ils entrent à l'école primaire en CP-CE1 à leur arrivée en France (ce qui était le cas de 35% seulement de notre échantillon), les EANA se distinguent a minima des élèves natifs de leur classe d'âge pour deux raisons : Premièrement par leur mode d'acculturation à l'écrit plus informel que scolaire (seul 11% de notre échantillon avait ainsi fait sa maternelle, du moins partiellement, en France). Deuxièmement par leur retard lexical en français. Quand ils sont scolarisés – et commencent par conséquent à apprendre à lire le français – à un âge plus avancé (avec en moyenne deux ans de retard par rapport à l'âge chronologique, dans notre échantillon), ils peuvent, dans le meilleur des cas, transférer au



français les compétences en littéracie qu'ils ont acquises pour autant qu'ils aient commencé l'apprentissage de la lecture dans une autre langue.

En raison de la pression migratoire, les EANA présentent, d'une part, des profils beaucoup plus hétérogènes que les apprentis lecteurs natifs, du point de vue des compétences langagières et scolaires, d'autre part, ils sont aussi globalement plus vulnérables aux plans psycho-affectif et socio-économique.

Compte tenu de l'état de la recherche sur les prédicteurs précoces de la lecture, notamment des conclusions de Bowey (1995) ou de Nation & Cocksey (2009) précédemment citées, nous avons supposé, et souhaité vérifier, dans le cadre de cette étude, que le stock lexical des EANA pouvait être un bon prédicteur de leur niveau de lecture à voix haute. Pour ce faire, nous avons donc comparé les performances des EANA à celles d'élèves tout-venant en lecture de mots et de pseudo-mots. Cela nous a permis d'étudier les processus spécifiques de la lecture, avec et sans l'étayage du stock lexical attendu pour l'âge.

D'après les résultats obtenus aussi bien chez les EANA que chez les élèves tout-venant, le stock lexical réceptif n'est pas un prédicteur du décodage. Nous pouvions nous y attendre au vu de la littérature : Colé (2012), citée plus haut, a fait ce constat ; Ricketts et al. (2007) ont montré que l'étendue du vocabulaire ne prédit pas la lecture des mots réguliers et des non-mots, ceux-ci étant lus par une procédure de décodage phonologique.

L'importante contribution de la triade phonologique (conscience phonologique, mémoire phonologique, DRA) à la lecture de mots et de non-mots observée ici, tant chez les élèves tout-venant que chez les EANA (sauf pour la mémoire phonologique), est sans surprise au vu du consensus scientifique sur les facteurs phonologiques sous-jacents à la lecture.

Nous nous attendions par contre à ce que le stock lexical des EANA rende compte d'une plus grande part de variance individuelle en lecture de mots à voix haute. Bignon et al. (2022) expliquent en effet que les EANA présentent des différences interindividuelles sur le stock lexical qui permet à un lecteur débutant d'identifier ses premiers mots écrits, contrairement aux élèves tout-venant qui ont un vocabulaire beaucoup plus étendu. Dans cette hypothèse, une action adaptée à cette population en dépit de sa diversité de profils, à savoir l'entraînement lexical, aurait alors permis de favoriser sa réussite dans l'apprentissage de la lecture. Or, la part de variance attribuable au vocabulaire s'avère certes plus élevée (2.56 %,  $p < .001$ ) que dans l'autre groupe (1.69 %,  $p < .001$ ), mais l'écart entre la variance expliquée dans l'un et l'autre groupe (68 % contre 49 %) oblige à relativiser ce résultat. On ne relève pas d'effet plancher en vocabulaire chez les EANA : bien que traduisant une grande variabilité, les scores bruts témoignent, chez les meilleurs élèves, de bonnes connaissances lexicales en français.

D'après les observations de terrain, les EANA semblent avoir des difficultés à s'appuyer sur leur vocabulaire oral pour s'auto-corriger selon le mécanisme d'auto-apprentissage décrit par Share. En l'occurrence, celui-ci ne semble pas porter ses fruits malgré l'apprentissage préalable de la lecture dans la langue maternelle (qui concerne 55% de l'effectif des EANA et 41% de cet effectif scolarisé en CE2-CM2) et/ou au moins 6 mois de scolarisation en UPE2A. Ces résultats étonnants seraient cohérents avec l'absence de corrélation que nous avons relevée entre le niveau de lecture et le temps de scolarisation en France, d'une part, l'exposition explicite au français écrit, d'autre part. Cependant, il convient avant tout de comparer les listes de mots et de pseudo-mots appariées entre elles pour vérifier l'absence d'effet de lexicalité recherché.

Les résultats obtenus, a priori en retrait sur ceux que nous escomptions nous confrontent

à une limite de notre étude. En s'inscrivant dans un projet de plus grande envergure, cette recherche a bénéficié de la puissance statistique d'un échantillon de grande taille. En contrepartie, elle ne s'est pas appuyée sur des épreuves taillées sur mesure.

Ainsi, le lexique testé par l'EVIP n'est pas ciblé sur les premiers mots écrits systématiquement proposés en décodage à l'apprenti lecteur pour déclencher le mécanisme du *set for variability*. Compte tenu de l'exposition variable des EANA au français oral (exclusivement scolaire ou également familiale), il aurait été intéressant de s'assurer que ces mots figuraient bien dans leur stock lexical. Cependant, les résultats quantitatifs de l'épreuve proposée ne nous fournissent pas cette information qualitative.

L'épreuve de désignation d'images traditionnellement utilisée pour mesurer l'étendue du stock lexical réceptif nous inspire une autre réserve. En effet, la modalité visuelle convoque des représentations sémantiques. Or, celles-ci contribuent davantage à la compréhension en lecture qu'à la reconnaissance des mots écrits. Une tâche sur le modèle du *go/no go*, avec signalisation des mots entendus connus d'un simple clic, mais sans critère temporel, aurait peut-être permis de mesurer plus fidèlement l'étendue du vocabulaire réceptif en comptabilisant les mots dont seule l'image phonologique est stockée dans le lexique interne.

Enfin, force est de constater qu'une étude d'une telle envergure génère un bruit important du fait des barrières tant culturelles que linguistiques (compréhension des questions, franchise des réponses) et des multiples intermédiaires (parents, enseignants, expérimentateurs). Aussi, bien que la piste du vocabulaire comme prédicteur de l'apprentissage de la lecture chez les allophones n'ait pas été légitimée avec autant de force que nous nous y attendions, cette recherche indique néanmoins qu'elle est digne d'être explorée. Dans le prolongement de cette étude, un autre angle d'attaque, tel que la comparaison du niveau de lecture avant et après une stimulation lexicale intensive couplée à un entraînement explicite au *set for variability* pourrait ainsi être adopté.

D'un point de vue clinique cette fois, les résultats de cette recherche nous confrontent au retentissement des facteurs non cognitifs sur les apprentissages. Ils nous rappellent la nécessité de ne pas les sous-estimer, même s'ils ne se prêtent pas à une intervention directe. Il n'est pas exclu en effet que le décalage observé en lecture à voix haute entre les deux groupes puisse s'expliquer en partie par un manque de disponibilité aux apprentissages lié à l'anxiété, la loyauté à la langue d'origine, la dépression... Dans la perspective de notre activité d'orthophoniste, cela nous rappelle la nécessité de ne pas nous focaliser sur le symptôme au risque de perdre de vue sa cause, et de penser à inscrire notre travail ciblé sur les aspects cognitifs dans une approche plus globale, qui prenne en compte les aspects socio-économiques avec le concours d'une assistante sociale et psycho-affectifs par le truchement d'une psychologue...

## Bibliographie

Agence Nationale de Lutte Contre l'Illettrisme (ANLCI). Pour en finir avec les idées reçues <http://www.anlci.gouv.fr/Illettrisme/Pour-en-finir-avec-les-idees-recues2/Pour-en-finir-avec-les-idees-recues>

Alegria, J., & Morais, J. (1989). Analyse segmentale et acquisition de la lecture. In L. Rieben & C. A. Perfetti (Eds), *L'apprenti lecteur : recherches empiriques et implications pédagogiques*. Neuchâtel et Paris : Delachaux & Niestlé.

Bara, F., Gentaz, É. & Colé, P. (2004). Les effets des entraînements phonologiques et multisensoriels destinés à favoriser l'apprentissage de la lecture chez les jeunes enfants. *Enfance*, 56, 387-403. <https://doi-org.ressources-electroniques.univ-lille.fr/10.3917/enf.564.0387>

Bignon, M., Mejias, S., Casalis, S. (2022). Apprentissage de la lecture en situation d'allophonie, UNADREO

Bowey, J. (2005). Predicting Individual Differences in Learning to Read. In *The science of reading – A handbook* (M. J. Snowling & C. Hulme, p. 155–172). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470757642.ch9>

Bulletin officiel n°44 du 25 novembre 2021 – Accompagnement des élèves à la pratique de la lecture <https://www.education.gouv.fr/bo/21/Hebdo44/MENE2133905N.htm>

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5-51. [en accès libre : <http://journals.sagepub.com/doi/abs/10.1177/1529100618772271>].

CASNAV de Lille. (2021). Guide pour la scolarisation des EANA : *Inclure les élèves allophones nouvellement arrivés* (Réseau Canopé, 2021) [Guide pour la scolarisation des EANA 2021.pdf](https://www.reseau-canope.fr/ressources/113089) (reseau-canope.fr)

Colé, P. (2012) Apprendre à lire : quelques repères. In Colé, P., Casalis, S., DomÍnguez, A. B., Leybaert, J., Schelstraete, M.-A., & Sprenger-Charolles, L. (Eds), *Lecture et pathologies du langage oral* (Presses universitaires de Grenoble). Presses universitaires de Grenoble.

Demont, E. 2003. Conscience phonologique et mémoire de travail : quelle contribution respective aux difficultés de lecture ? : Etude auprès d'enfants francophones dyslexiques et lecteurs débutants. In Nouri Romdhane, M., Gombert, J. E., & Belajouza, M. (Eds.), *L'apprentissage de la lecture : Perspectives comparatives*. Presses universitaires de Rennes. Doi : 10.4000/books.pur.48450

Demont, É. & Gombert, J. (2004). L'apprentissage de la lecture : évolution des procédures et apprentissage implicite. *Enfance*, 56, 245-257. <https://doi.org/10.3917/enf.563.0245>

Desrochers, A., Carson, R., & Daigle, D. (2012). Une analyse des facteurs de risque dans l'apprentissage de la lecture chez l'enfant. *Enfance en difficulté*, 1, 47-83. <https://doi.org/10.7202/1012123ar>

Desrochers, A., Kirby, J. R., Thompson, G. L. & Fréchette, S. (2009). Le rôle de la conscience phonologique dans l'apprentissage de la lecture. *Revue du Nouvel-Ontario*, (34), 59–82. <https://doi.org/10.7202/038720ar>

Ecalles, J., Labat, H., Thierry, X. & Magnan, A. (2020). Évaluation des compétences en littératie chez les enfants français de 4-5 ans. *Santé Publique*, 32, 9-17. <https://doi-org.ressources-electroniques.univ-lille.fr/10.3917/spub.201.0009>

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250–<https://doi.org/10.1598/RRQ.36.3.2>

Erdoş, C., Genesee, F., Savage, R., et Haigh, C. (2010). Individual differences in second language reading outcomes. *International Journal of Bilingualism*, 15(1), 3-25.

Ferré, S. et Dos Santos, C. (2015). Comment évaluer la phonologie des enfants bilingues ? *Lidil. Revue de linguistique et de didactique des langues*, 51, 11-34

Fluss, J., Bertrand, D., Ziegler, J. & Billard, C. (2009). Troubles d'apprentissage de la lecture : rôle des facteurs cognitifs, comportementaux et socio-économiques. *Développements*, 1, 21-33. <https://doi-org.ressources-electroniques.univ-lille.fr/10.3917/devel.001.0021>

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. [https://doi.org/10.1007/978-3-030-26550-2\\_6](https://doi.org/10.1007/978-3-030-26550-2_6)

Gombert, J. E. et Colé, P. (2000). Activités métalinguistiques, lecture et illettrisme. Dans M. Kail et M. Fayol (Dir.) : *L'acquisition du langage : le langage en développement au-delà de trois ans*. Paris, France : Presses universitaires de France.

Grabe, W. (2009) *Reading in a Second Language : Moving from theory to practice*. Cambridge University Press, Cambridge, England.

Hatcher, P., Hulme, C., and Snowling, M. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry*, 45, 338-358.

Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. V. (2008). Longitudinal predictors of word reading development. *Canadian Psychology / Psychologie canadienne*, 49(2), 103–<https://doi.org/10.1037/0708-5591.49.2.103>

Kirby, J. et Savage, R. (2008). Can the simple view deal with the complexities of reading? *Literacy*, 42(2), 75-82. <http://onlinelibrary.wiley.com/doi/10.1111/j.1741-4369.2008.00487.x/full>

Lecocq, P. (1991), *Apprentissage de la lecture et dyslexie*. Mardaga, Bruxelles

Majerus, S., & Van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, 21(2), 303–310. <https://doi.org/10.1348/026151003765264101>

Melby-Lervåg, M., Lyster, S. A., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin*, 138(2), 322-352. <https://www.ncbi.nlm.nih.gov/pubmed/22250824>

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91 (1), 3-19.

Morais, J., Pierre, R. & Kolinsky, R. (2003). Du lecteur compétent au lecteur débutant : implications des recherches en psycholinguistique cognitive et en neuropsychologie pour l'enseignement de la lecture. *Revue des sciences de l'éducation*, 29(1), 51-74. <https://doi.org/10.7202/009492ar>

Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, 103(3), 296-308. Scopus. <https://doi.org/10.1016/j.jecp.2009.03.004>

Negro, I., Genolo, S., Les prédicteurs en grande section de maternelle de la réussite en lecture en fin de première année d'école élémentaire : l'impact du nom des lettres (292-306), *Bulletin de psychologie*, 2009/3 Numéro 501

Nicolas, S. (2003). 7. L'accès au lexique : la reconnaissance des mots. Dans : , S. Nicolas, *La psychologie cognitive* (pp. 102-120). Paris: Armand Colin.

Nithart, C., Demont, E., & Metz-Lutz, M. 2009. Traitement phonologique en mémoire à court terme chez les enfants dyslexiques et dysphasiques. In Marec-Breton, N., Besse, A., De La Haye, F., Bonneton-Botté, N., & Bonjour, E. (Eds.), *L'apprentissage de la langue écrite : Approche cognitive*. Presses universitaires de Rennes. doi :10.4000/books.pur.60518

Poncelet, M., Van der Linden, M., & Majerus, S. (2001). Role of phonological and visual short-term memory components in acquisition of reading procedures. Paper presented at The Third International Conference On Memory, July 16<sup>th</sup>-20<sup>th</sup> 2001, Valencia (Spain).

Rafoni, J. (2015). Vers un modèle didactique d'apprentissage de la lecture en français langue seconde et maternelle. *Éla. Études de linguistique appliquée*, 179, 315-334. <https://doi.org.ressources-electroniques.univ-lille.fr/10.3917/ela.179.0315>

Ricketts, J., Nation, K. & Bishop, D. V. M. (2007). Vocabulary Is Important for Some, but Not All Reading Skills. *Scientific Studies of Reading*, 11(3), 235-257. <https://doi.org/10.1080/10888430701344306>

Scarborough, H. (1998). Early identification of children at risk of reading disabilities: Phonological awareness and Some other promising predictors.

Sénéchal M., LeFevre, J.-A., Smith-Chant B., Colton, K. (2001). On refining theoretical models of emergent literacy : The role of empirical evidence. *Journal of School Psychology*, 39, pp.439-460

Valdois, S., Ans, B. & Carbonnel, S. (2009). 3. Principes de base pour l'élaboration d'un modèle connexionniste psychologiquement plausible de l'apprentissage de la lecture. *Apprentissage des langues*, 379-404. <https://doi.org/10.4000/books.editions-cnrs.5977>

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology*, 76(6), 1309– 1324. <https://doi.org/10.1037/0022-0663.76.6.1309>

Sprenger-Charolles, L., & Ziegler, J.C. (2019). Apprendre à lire : contrôle, automatismes et autoapprentissage. In A. Bentollila & B. Germain, editor, *L'apprentissage de la lecture*. Nathan, Septembre.

Tiré, M., Vadcar, A., Ragano, S., & Bazile, S. (2015). Pour une approche objective des pratiques d'acculturation à l'écrit en classe de CP : essai de catégorisation et premières analyses. *Reperes*, 52, 77-96. <https://doi.org/10.4000/reperes.943>

Tossonian, M. (2019), *Les règles phonologiques universelles : comment contribuent-elles aux stratégies de lecture chez les adultes normo-lecteurs et illettrés ?* Psychologie, Université Clermont Auvergne (UCA) ; Laboratoire de Psychologie Sociale et Cognitive (LAPSCO), NNT : 2019CLFAL025, <https://hal.science/tel-03879868>

Walley, A., Metsala, J.L., & Garlock, V.M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing: An Interdisciplinary Journal*, 16, 5-20.

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.