

## THÈSE DE DOCTORAT

NNT/NL : 2021AIXM0053/011ED356

Soutenue à Aix-Marseille Université

Le 12 janvier 2021 par

Ludivine JAVOUREY-DREVET

# La simplification de textes comme outil pour améliorer la fluidité et la compréhension de lecture chez les enfants à l'école primaire

Une étude en longitudinal avec des textes littéraires et scientifiques chez des enfants entre 7 et 9 ans

**Discipline :** Sciences de l'éducation

**Spécialité :** Psychologie cognitive

**École doctorale :** Cognition, Langage, Éducation (ED356)

**Laboratoires :**

Laboratoire Apprentissage, Didactique, Évaluation, Formation (UR 4671) AMU

Laboratoire de Psychologie cognitive (UMR 7290) CNRS & AMU

### Composition du jury

Maryse Bianco, MCF, HDR émérite, Université Grenoble Alpes, France	Rapporteure
Brigitte Marin, PR, DR, INSPE, Université Paris-Est Créteil, France	Rapporteure
Edouard Gentaz, PR, Université de Genève, Suisse	Examineur
Liliane Sprenger-Charolles, PR, CNRS, Aix-Marseille Université, France	Présidente
Johannes C. Ziegler, DR, CNRS, Aix-Marseille Université, France	Co-directeur
Jacques Ginestié†, PR, Aix-Marseille Université, France	Directeur

M. Jacques Ginestié n'a pas pu participer à la soutenance suite à son décès fin septembre 2020.

## AFFIDAVIT

Je soussignée, Ludivine Drevet épouse Javourey, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de M. Johannes C. Ziegler et de M. Jacques Ginestié dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Les travaux présentés dans ce manuscrit n'ont pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire dans le cadre d'un autre processus d'examen ou de soutenance.

Fait à Marseille le 07/11/2020

*Signature*

A handwritten signature in black ink, consisting of several loops and a horizontal line extending to the right.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).

# Résumé

La compréhension en lecture est une compétence essentielle à la réussite des apprentissages scolaires et tout au long de la vie. Pourtant, un pourcentage élevé d'élèves rencontre d'importantes difficultés pour comprendre un texte écrit. Pour ces élèves, la simplification de textes peut alors apparaître comme une aide car son objectif est de rendre plus facilement compréhensible un texte donné. La simplification consiste à réduire la complexité d'un texte tout en conservant les informations et le sens original. Dans ce travail de thèse, nous avons étudié l'effet de la simplification sur la fluidité et la compréhension en lecture de textes littéraires et documentaires scientifiques pour des élèves entre sept et neuf ans. Une première étude a été réalisée sur trois ans dans laquelle la simplification a été effectuée sur trois niveaux, lexical, syntaxique et discursif, sur des textes littéraires et documentaires scientifiques. L'étude a montré l'efficacité de la simplification davantage en fluidité de lecture et qu'en compréhension. La simplification était particulièrement bénéfique pour les plus faibles lecteurs qui avaient un faible niveau en phonologie, vitesse de lecture, vocabulaire et orthographe morphologique. Parmi les facteurs textuels qui expliquent les gains de la simplification, nous avons trouvé la fréquence et la longueur des mots, le nombre de mots par phrase et le remplacement des pronoms par leur référent. Dans une deuxième étude, nous avons voulu savoir si un type de simplification – lexical versus syntaxique - était nécessaire ou suffisant pour engendrer des gains de fluidité et de compréhension en lecture. Les résultats ont montré que chaque type de simplification est suffisant pour améliorer la compréhension et la fluidité. Enfin, nous avons voulu savoir si un entraînement uniquement avec des textes simplifiés pendant un certain temps pourrait durablement améliorer la compréhension de n'importe quel texte notamment pour les faibles lecteurs (effet de généralisation ou de transfert). Les résultats n'ont pas montré un effet de transfert spécifique à la lecture des textes simplifiés. En revanche, l'entraînement à la lecture avec les deux types de textes (simplifiés et originaux) s'est révélé bénéfique pour la fluidité par rapport au groupe contrôle sans entraînement avec l'application. En conclusion, cette thèse a montré le potentiel mais aussi les limites de la simplification de textes pour améliorer la fluidité et la compréhension. Elle a aussi permis le développement et la publication de deux ressources numériques, accessibles librement sur internet. Ces ressources proposent l'ensemble des textes originaux et simplifiés ainsi que les questionnaires utilisés pour ce travail de thèse.

**Mots-clés :** simplification de textes, compréhension en lecture, fluence, textes littéraires, documentaires scientifiques, numérique.

# Abstract

Reading comprehension is an essential skill for successful learning in school and throughout life. Yet, a high percentage of students have significant difficulty understanding written text. For these students, the simplification of texts could be beneficial because its objective is to make a given text more easily understandable. Simplification consists in reducing the complexity of a text while preserving the original information and meaning. In this thesis work, we studied the effect of simplification on fluency and comprehension in reading literary and scientific texts for students between the ages of seven and nine. A first longitudinal study was conducted over three years, in which simplification was carried out at three levels: lexical, syntactic and discursive. The study showed that simplification produced significant gains more in reading fluency than in comprehension. Simplification was particularly beneficial for weaker readers who had low levels of phonology, reading speed, vocabulary and morphological spelling. Among the textual factors that explain the gains from simplification, we found word frequency and length, the number of words per sentence, and the replacement of pronouns with their referent. In a second study, we wanted to know whether a type of simplification - lexical versus syntactic - was necessary or sufficient to generate gains in reading fluency and comprehension. The results showed that each type of simplification was sufficient to improve comprehension and fluency. Finally, we wanted to know whether training only with simplified versus original texts can durably improve the comprehension of any text, especially for weak readers (generalization or transfer effect). The results did not show a specific transfer effect when reading the simplified texts. However, training in reading with both types of texts (simplified and original) proved to be beneficial for fluidity compared to the control group without training with the application. In conclusion, this thesis showed the potential but also the limits of simplifying texts to improve fluidity and comprehension. It has also led to the development and publication of two digital resources, freely accessible on the internet. These resources propose all the original and simplified texts as well as the questionnaires used for this thesis work.

Keywords: simplification of texts, reading comprehension, fluency, literary texts, scientific documentaries, numeric.

**À mes enfants Lina et Corentin**

« Si vous avez du pain, et si moi j'ai un euro, si je vous achète le pain, j'aurai le pain et vous aurez l'euro et vous voyez dans cet échange un équilibre, c'est-à-dire : A a un euro, B a un pain. Et dans l'autre cas B a le pain et A a l'euro. Donc, c'est un équilibre parfait. Mais, si vous avez un sonnet de Verlaine, ou le théorème de Pythagore, et que moi je n'ai rien, et si vous me les enseignez, à la fin de cet échange-là, j'aurai le sonnet et le théorème, mais vous les aurez gardés. Dans le premier cas, il y a un équilibre, c'est la marchandise, dans le second il y a un accroissement, c'est la culture. » (Michel Serres - Entretien dans le journal Libération - 28 avril 2009)

## Remerciements

Pour commencer, je tiens à adresser mes remerciements les plus remarquables aux membres du jury, en particulier un immense merci aux rapporteuses, Maryse Bianco et Brigitte Marin pour le temps consacré à l'évaluation du manuscrit, à l'examinateur, Edouard Gentaz, ainsi qu'à Liliane Sprenger-Charolles, présidente du jury.

Mon très cher Jacques, ton grand sourire, ta confiance en l'avenir et en chacun me manque ainsi que ta bienveillance. Nous n'irons pas voir les crocodiles ensemble... mais je vais essayer de faire de mon mieux afin de continuer à « grandir » pour essayer d'apporter ma petite contribution à la vision ouverte et progressive de la recherche que tu avais. Tu m'as guidée et donné confiance comme toi seul savais le faire. J'aurais tellement aimé que tu sois à mes côtés jusqu'à la fin de cette aventure mais il en est ainsi, donc je vais essayer d'être à la hauteur de nos envies. Il me plaît à croire que tu as aimé cette aventure, tout comme moi, en étant pleinement consciente que tu en as vécu de multiples, c'est sans doute pour cette raison qu'elle a été aussi enrichissante grâce à ton intelligence et à ta finesse. Cela a été pour moi un honneur de te rendre fier de notre travail ! Il faut maintenant que je continue l'aventure en ayant perdu un « guide » mais je sais que tu as essayé de me transmettre l'essentiel pour que je puisse voler de mes propres ailes, même si ta présence, au moment de l'envol, n'est pas possible. Je vais donc faire de mon mieux afin de te rendre hommage que ce soit au niveau de la recherche, tout comme au niveau des valeurs que tu nous as transmis.

Naturellement, un grand merci à Johannes Ziegler pour cette formidable aventure. Celle-ci n'aurait pas vu le jour sans toi. Tu m'as fait confiance alors que deux univers se rencontraient mais tu as pris le temps de l'écoute et du partage afin de comprendre. Merci pour ton implication et ton soutien. Tu as toujours été là et j'ai ainsi appris énormément. Je suis pleinement consciente du fait qu'il me reste encore beaucoup à apprendre et à comprendre, c'est ce qui fait tout le charme de ce métier qui me plaît. Cela est une chance d'avoir trouvé un métier qui me permet de m'épanouir au quotidien. Tout cela n'aurait pas été possible sans toi et ta volonté d'avancer avec toute l'exigence et la rigueur que demande la recherche. J'ai un immense respect pour toi. Tu as passé tellement de temps à me former et à m'apprendre ce métier...même si je sais que cela est un travail constant. En espérant ne jamais te décevoir. Tu m'as laissé une immense liberté, qui a été très appréciable, tout en étant présent les soirs, les week-ends... Vivre avec la recherche en soi, au rythme de la recherche est un plaisir immense que j'ai pu ressentir

à tes côtés. Merci pour ces heures de réflexion, de discussion, de relecture, de révision d'articles, de formation tout simplement. La recherche est aussi une aventure humaine et je te remercie d'être là, même si je sais que je vais ensuite m'émanciper ... Tout simplement MERCI !

Je souhaite adresser un inestimable merci à Stéphane Dufau, j'aime ta vision du monde qui comprend le partage et la bienveillance. Nous travaillons ensemble avec complicité car il me semble que nous sommes attachés aux mêmes valeurs et notamment l'envie d'aider les élèves et surtout ceux qui en ont le plus besoin. Merci pour ton aide et ton soutien tout au long du parcours. Le livre interactif n'aurait jamais vu le jour sans tes précieuses compétences, ta patience et tes convictions. J'espère que nous pourrons encore vivre des moments de recherche qui permettront des avancées intéressantes pour les élèves et les connaissances. Pouvoir avoir confiance en une personne dans le cadre professionnel est une chance qui m'a aidée tout au long de l'aventure. J'aimerais que cela ne s'arrête pas...

Je souhaite témoigner ma plus grande reconnaissance à Núria Gala, nous avons appris à nous connaître et à travailler ensemble grâce au projet ALECTOR et ainsi nous avons vécu des moments d'échanges très intéressants au niveau de la recherche et au niveau humain car tu es chaleureuse. Tu m'as étayée dans mes premiers pas vers l'étude de la simplification de textes et en linguistique notamment mais nous avons partagé aussi des moments conviviaux car la recherche est aussi une aventure humaine faite de rencontres et de partage. L'ensemble de l'équipe étant impliquée dans le projet ALECTOR m'a soutenue, tout au long de ce projet. Tu m'as accompagnée comme tu le fais avec chacun de tes étudiants, merci beaucoup !

Je tiens à présenter un immense merci à Thomas François, tu m'as accordé beaucoup de temps afin de me laisser le temps de comprendre, d'apprendre. Ces nombreux moments de réflexion m'ont permis de continuer ma formation, même si je sais qu'elle ne sera jamais aboutie...C'est un processus sans fin et je crois que cela fait partie de son charme. Merci aussi pour ces moments de soutien avec des BD comme les PhD Comics.

Je remercie très chaleureusement, Lisi Beyersmann, pour m'avoir donné cette chance et cette première expérience dans la recherche mais aussi ta gentillesse et tes conseils précieux. Ton efficacité scientifique m'a guidée tout au long de ces années car tu es performante tout en étant chaleureuse et compréhensive.

Je remercie très sincèrement M. Grainger, merci pour la confiance accordée tout au long de ces années alors que mon profil n'était pas « classique » avec un parcours professionnel de 14 ans au sein de l'éducation nationale. Votre regard bienveillant et votre honnêteté intellectuelle permettent de pousser à vouloir donner le meilleur de soi-même et de vivre une recherche la plus efficiente possible.

Je souhaite remercier très amicalement, Colette Pourpe et Frédéric Lombardo du LPC, ils instaurent un accueil des plus agréables et j'ai pu avoir grâce à vous des conditions de travail très appréciables. Je souhaite remercier, Annie Rombi et Anne Scher, sans qui les démarches administratives ne seraient pas possibles. C'est parfois des temps d'écoute et de partage dans

les moments stressants... Merci à tous pour vos implications très importantes dans les laboratoires, ce qui dépasse le cadre purement administratif car la gestion humaine est loin d'être la partie la plus simple à organiser et vous le faites avec beaucoup de bienveillance.

Je souhaite également remercier, l'ensemble des équipes du LPC, Julie Lassault avec qui j'ai évolué depuis le premier jour de notre arrivée ensemble, Edouard Alavoine pour sa bonne humeur et son aide constante, Jean-Patrice Albrand pour son soutien lors de chaque étape. Je repars quand vous voulez... Ma valise est prête ! Jonathan Mirault qui est toujours joignable, constamment présent afin d'aider les autres... Agnès Guerre-Genton pour son écoute tout au long de ce processus si particulier mais aussi Camille Grasso, Marion Fecho, Marco Bressan, Aurélie Calabrèse, sans oublier l'ensemble des chercheurs du laboratoire, Marie Montant, Patrick Lemaire, Xavier Alario ... qui prennent toujours le temps d'écouter les doctorants alors que leur temps est précieux.

Je souhaite également remercier l'équipe du laboratoire ADEF, Pascale Brandt-Pomares, Jérémie Castéra, Marjolaine Chatoney, Hélène Cheneval-Armand, Marie-Christine Félix, Jean-François Hérold, Maria Antonietta Impedovo, Corinne Jégou, Caroline Ladage, Patrice Lesnay, Pascal Terrien... et tout particulièrement Jean Ravestein qui m'a guidée vers le doctorat, pour vos conseils, votre écoute, vos compétences et votre soutien dès le début de ma reprise d'études. Vous m'avez fait confiance et toutes les disciplines continuent à enrichir mon travail. Les doctorants progressent, partagent et vivent la recherche à vos côtés, c'est une chance !

Je tiens à remercier sincèrement, la Communauté d'Agglomération Sud Sainte Baume et plus particulièrement Mme Liliane Giraudet. La CASSB nous a fait confiance et nous a permis d'avoir accès aux tablettes, en prenant soin d'être réactive tout au long du projet. Merci aussi pour l'accompagnement technique de Mme Isabelle Garnier qui a fait un suivi régulier afin que chaque phase expérimentale se déroule au mieux.

Je tiens à remercier sincèrement, l'Inspectrice de la circonscription, Mme Nathalie Greppo-Chaignion qui a permis l'aboutissement du projet et l'accès à l'ensemble des écoles. Merci à Brigitte Casalini, Caroline Schang, Céline Sola et Louis Fangouse, l'ensemble de l'équipe des conseillers pédagogiques a toujours été très réactive et a toujours soutenu notre travail. Rien n'aurait été possible sans leur présence.

Un merci tout particulier aux équipes enseignantes qui ont toujours ouvert leur école, leur classe avec beaucoup de simplicité et montré leur confiance en la recherche et en l'avenir. Merci aussi pour le temps accordé et l'adaptation tout au long de ces années et de cette aventure. Ces nombreux moments d'échanges et de partages ont aidé mes réflexions. J'espère que le travail réalisé continuera à vous être utile.

Je souhaite exprimer mes plus chaleureux remerciements à l'ensemble des élèves qui ont participé à cette recherche en étant pleinement conscients de l'importance de ce travail. Ils ont toujours été très impliqués et ont apprécié l'iBook, nous ont orientés afin d'être au plus proche de leurs attentes. Je n'oublierai jamais les sourires et les visages lorsque j'arrivais dans les

écoles. Nous avons vécu de beaux moments. Je vous ai vu grandir et vous pouvez être fiers de vous. Cela a renforcé ma conviction sur l'intérêt du travail réalisé.

Des remerciements qui n'ont pas de limite, qui n'ont pas de mots pour les exprimer, je les adresse à celui avec qui j'ai la chance de partager ma vie depuis toutes ces années, mon mari, qui a toujours été là pour moi.

« Il y aura des obstacles, il y aura des doutes, il y aura des erreurs, mais si tu travailles fort, il n'y aura aucune limite ». (Auteur inconnu)

# Tête

Commencer un poème par : « Il était une fois ... », on l'aura fait bien des fois, ma foi, sans trop de mauvaise foi, ça ne sera ni la première fois ni la dernière fois, aussi :

Il était une fois, il était une vie, il était une fille, il était sa tête,  
Oui, il était sa tête : tête de mule, tête dure, tête en l'air, tête de piaf car tête dans les nuages mais toujours tête sur les épaules.

Sa tête qui pensait beau mais trop, bien mais rien,

Or rien, ce n'est pas rien, c'est du véritable or de patricien !

Ce n'était pas pour autant qu'elle avait la grosse tête, juste toujours une idée en tête qui, comme une chansonnette, lui revenait à tue-tête,

Une vraie prise de tête, un casse-tête à en avoir le mal de tête !

Imaginez ! Tenir tête à sa tête, il y'en avait vraiment par-dessus la tête !

Quand elle avait une sale tête, cette fille avait parfois en tête de n'en faire qu'à sa tête, d'agir sur un coup de tête ou tête baissée mais après s'être creusé la tête, à tête reposée, tout ceci était si bête !

Il faut dire que sa tête de palourde, vraiment pas lourde, jamais ne la laissait en paix : des idées, des idées, des idées, le jour, la nuit, à midi ou à minuit.

Question conception, elle était toujours en tête, comme en en-tête de la théorie, du thème et de la thèse.

Quoi qu'un peu antithèse, c'était à s'en couper la tête, une fille qui avait toute sa tête : elle gardait la tête froide, enfoncez-le-vous dans la tête.

Non ne lui faites pas la tête, cela vous coûterait les yeux de la tête.

Vous seriez alors comme vêtu de la tête aux pieds, d'une mauvaise idée en tête, et cela n'a juste ni queue ni tête !

Comme elle, redressez donc la tête, même si l'on se paye votre tête, votre port de tête, lui, ne vous devra aucune dette.

A s'en frapper la tête contre les murs, ce poème est pour sûr, à en tourner la tête, un moyen de perdre la tête !

Mais si la tête d'affiche, la tête de classe en est l'héroïne, et non juste une simple tête de linotte ou tête de pitre,

C'est bien qu'elle a raison, et qu'il faut cueillir ce message, comme si l'on butinait une tête d'impression.

C'est alors sans secouer la tête, que vous pourrez vivre sans vous prendre la tête !

## Contexte

Ces travaux de thèse ont été réalisés en parallèle de mes engagements sur le projet de l'ANR franco-allemande « Morphèmes »<sup>1</sup>, coordonnée par Jonathan Grainger. L'objectif de cette recherche sur laquelle j'ai été embauchée en qualité d'ingénieure d'études au sein du Laboratoire de Psychologie Cognitive (LPC) était d'évaluer le rôle des représentations morphologiques lors de l'apprentissage de la lecture. Le but de cette ANR était de préciser les influences relatives aux caractéristiques morphologiques spécifiques à chaque langue avec une approche longitudinale et cross-linguistique, avec les mêmes expériences menées en parallèle en France et en Allemagne. Dans ce cadre, j'ai mesuré les habiletés psycholinguistiques et les connaissances morphologiques en compréhension de langage oral chez des enfants de sept à neuf ans. Durant la même période, le Laboratoire de Psychologie Cognitive était engagé sur une autre ANR « ALECTOR »<sup>2</sup>, coordonnée par Núria Gala. Son objectif était de fournir une aide à la lecture, aux enfants dyslexiques et faibles lecteurs en proposant des textes adaptés grâce à des transformations linguistiques, sans appauvrir les textes écrits, en utilisant le traitement automatique des langues. Dans le cadre de ce projet j'ai effectué manuellement des simplifications linguistiques de textes. Des chercheurs impliqués dans cette ANR ont ensuite analysé ce corpus parallèle afin de déterminer une typologie de transformations pour un développement semi-automatique de certaines règles pour la langue française. Pour remplir ces deux missions je me suis mise en disponibilité de l'Éducation Nationale pendant cette période. Cette situation professionnelle a permis mon inclusion à plein temps dans la recherche et donné l'opportunité de conduire mes travaux de thèse réalisés en quatre ans, en parallèle des emplois occupés. Le projet de recherche n'a pas été directement anticipé et préconçu par ces deux ANR. Toutefois en nous basant sur les moyens financiers et humains de ces deux projets nous avons pu optimiser les actions sur le terrain de recherche. Les expérimentations présentées au sein de ce travail doctoral ont été conçues indépendamment des objectifs de ces deux ANR. Le terrain de recherche, au sein duquel l'ensemble des expérimentations a été conduit, était confiant et ouvert car cette circonscription de l'Éducation Nationale avait déjà été sollicitée dans le cadre de mon mémoire de recherche sur l'année scolaire 2014-2015<sup>3</sup>. Pour autant, les enseignants n'ont pas été les mêmes sur les deux périodes, travail de mémoire et de thèse. Sur ce parcours

---

<sup>1</sup> <https://anr.fr/Projet-ANR-15-FRAL-0003>

<sup>2</sup> <https://anr.fr/Projet-ANR-16-CE28-0005>

<sup>3</sup> « La mise en place des nouveaux rythmes scolaires : une approche anthropologique de la didactique. » sous la direction de Marie-Christine FÉLIX (AMU & ADEF).

de formation universitaire j'ai expérimenté une méthode qualitative, entretien d'explicitation de l'action de Pierre Vermersch pour réaliser le mémoire et j'ai fait appel à des méthodes quantitatives dans le travail doctoral. Dans le cadre des travaux de cette thèse, l'engouement des professeurs pour le dispositif de lecture proposé a été tel que des études ont été ajoutées au projet initial et sont présentées dans ce manuscrit. Ces travaux de thèse sont interdisciplinaires et ont été réalisés sous la direction de M. Jacques Ginestié et de M. Johannes Ziegler.

## Table des matières

<b>Introduction</b> .....	<b>1</b>
<b>Chapitre 1 Évaluations de la compréhension en lecture de textes littéraires et scientifiques</b> .....	<b>4</b>
1.1 Enquêtes internationales .....	4
1.2 Enquêtes nationales .....	6
<b>Chapitre 2 La compréhension : mécanismes, modèles, facteurs, difficultés et interventions</b> .....	<b>9</b>
2.1. Les modèles théoriques de la compréhension .....	9
2.1.1 Modèle Simple de la lecture .....	9
2.1.2 Modèle de construction de structures .....	13
2.1.3 Modèle de Construction-Intégration (C-I) .....	15
2.1.4 La qualité lexicale dans la lecture .....	17
2.2 Les principaux facteurs qui influencent la compréhension écrite .....	20
2.2.1 Décodage et reconnaissance des mots .....	20
2.2.2 Fluidité .....	22
2.2.3 Vocabulaire .....	23
2.2.4 Morphosyntaxe .....	24
2.2.5 Les inférences dans la compréhension .....	25
2.2.6 Le contexte dans la lecture .....	27
2.2.7 Acculturation à l'écrit.....	28
2.3 Les difficultés lors de la compréhension .....	30
2.4 La compréhension en fonction de la nature du texte .....	34
2.5 Approches et outils pédagogiques pour améliorer la compréhension de textes .....	36
2.5.1 Présentation d'interventions destinées à améliorer la compréhension.....	36
2.5.2 Pratiques pédagogiques et outils pour améliorer la compréhension en lecture..	39
<b>Chapitre 3 La simplification de textes</b> .....	<b>44</b>
3.1 Adaptation de textes pour faciliter la lecture .....	44
3.2 Les différentes approches de la simplification .....	46
3.3 Les différents niveaux de simplification de textes .....	48
3.3.1 Simplification lexicale.....	49
3.3.2 Simplification morphosyntaxique .....	52
3.3.3 Simplification discursive.....	54
3.4 Études ayant évalué les effets de la modification de textes.....	57
3.4.1 Adaptation de textes .....	57
3.4.2 Simplification de textes avec ou sans apport d'informations.....	58
<b>Chapitre 4 Développement d'un corpus simplifié de textes scolaires</b> .....	<b>63</b>
4.1 Le corpus développé .....	63
4.1.1 Choix des textes et création des questions .....	63
4.1.2 Pré-tests .....	64
4.1.3 Le corpus .....	65
4.2 Calculer la difficulté d'un texte .....	66
4.2.1 Formules de lisibilité .....	67
4.2.2 Variables textuelles choisies pour analyser la lisibilité.....	69
4.2.3 Analyse de la lisibilité du corpus avec les variables choisies .....	69
4.2.4 Discussion pour donner suite aux analyses textuelles.....	74
<b>Chapitre 5 Objectifs des expérimentations</b> .....	<b>77</b>

<b>Chapitre 6 Effets de la simplification sur la fluidité et la compréhension : suivi de cohorte sur trois années .....</b>	<b>79</b>
6.1 Terrain de recherche et participants.....	79
6.2 Protocole expérimental conduit sur trois ans.....	80
6.3 Tests en individuel.....	83
6.4 Analyses des Questions à Choix Multiples .....	86
6.5 Traitement des données .....	87
6.6 Propriétés statistiques des variables dépendantes.....	87
6.7 Effets de la simplification sur la fluidité .....	88
6.8 Effets de la simplification sur la compréhension.....	90
6.8.1 Compréhension : Questions à Choix Multiples.....	91
6.8.2 Compréhension : Résumé.....	92
6.9 Discussion.....	93
<b>Chapitre 7 Différences inter-individuelles : qui bénéficie de la simplification ?.....</b>	<b>95</b>
7.1 Méthodologie du calcul des gains .....	95
7.2 Lecteurs bénéficiant de la simplification.....	95
7.3 Effets de la simplification suivant le niveau initial en vitesse de lecture de mots familiers et en vocabulaire sur la fluidité et la compréhension .....	97
7.3.1 Effets de la simplification suivant le niveau initial en vitesse de lecture à haute voix de mots familiers sur la fluence et la compréhension.....	98
7.3.2 Effets de la simplification suivant le niveau initial en vocabulaire sur la fluence et la compréhension .....	98
7.4 Discussion.....	101
<b>Chapitre 8 Variables textuelles expliquant les gains en fluidité et en compréhension sur les trois années .....</b>	<b>104</b>
8.1 Variables textuelles qui prédisent les gains en fluidité et en compréhension .....	104
8.2 Discussion.....	109
<b>Chapitre 9 Effets de la simplification : lexicale, syntaxique et lexico-syntaxique chez les lecteurs débutants.....</b>	<b>112</b>
9.1 Population.....	112
9.2 Design expérimental .....	112
9.3 Tests en individuel.....	113
9.4 Résultats en fluidité et en compréhension .....	114
9.4.1 Vitesse de lecture de mots .....	114
9.4.2 Résultats en compréhension : Questions à Choix Multiples .....	116
9.5 Discussion.....	118
<b>Chapitre 10 Étude d'entraînement : la simplification peut-elle durablement influencer les performances de lecture ? .....</b>	<b>120</b>
10.1 Population.....	120
10.2 Design expérimental .....	121
10.3 Tests en individuel pour les trois groupes .....	123
10.4 Analyses des effets de l'entraînement .....	126
10.4.1 Fluence par mot.....	126
10.4.2 Effets de l'entraînement sur la fluence en fonction du niveau initial dans les tests langagiers et cognitifs. ....	128
10.4.3 Compréhension : résultats aux Questions à Choix Multiples.....	130
10.4.4 Effets de l'entraînement sur la compréhension en fonction du niveau initial dans les tests langagiers et cognitifs.....	132
10.5 Discussion.....	132
<b>Chapitre 11 Discussion générale .....</b>	<b>135</b>

<b>Chapitre 12 Ressources disponibles pour donner suite aux études .....</b>	<b>141</b>
12.1 Le corpus ALECTOR.....	141
12.2 Livre interactif Hibou .....	143
<b>Conclusion.....</b>	<b>147</b>
<b>Références bibliographiques .....</b>	<b>148</b>
<b>Liste des sigles et acronymes .....</b>	<b>171</b>
<b>Tableaux.....</b>	<b>172</b>
<b>Figures.....</b>	<b>173</b>
<b>Annexe A .....</b>	<b>176</b>
<b>Annexe B .....</b>	<b>185</b>
<b>Mes publications.....</b>	<b>215</b>

### Introduction

L'activité de lecture, qui intervient dans la construction des apprentissages dans un cadre premièrement scolaire, s'étend par la suite à tous les domaines de connaissance et fait partie intégrante de notre quotidien. Cela fait de la période scolaire un moment déterminant pour l'avenir des lecteurs (Gaussel, 2015).

Bien que l'apprentissage et la compréhension de la langue orale se réalisent de manière implicite, souvent par immersion, nous n'avons pas de prédispositions biologiques pour apprendre la lecture. Cet apprentissage doit être explicite. L'apprentissage de l'écrit implique une « réorganisation » de régions cérébrales dédiées à d'autres fonctions. Ces régions cérébrales sont assez plastiques pour se réorganiser vers l'identification des signes écrits et leur mise en relation avec le langage parlé (Dehaene & Cohen 2007). Un nouveau réseau neuronal se crée, le réseau de la lecture, qui lie les traitements visuels (et/ou moteurs) aux traitements langagiers.

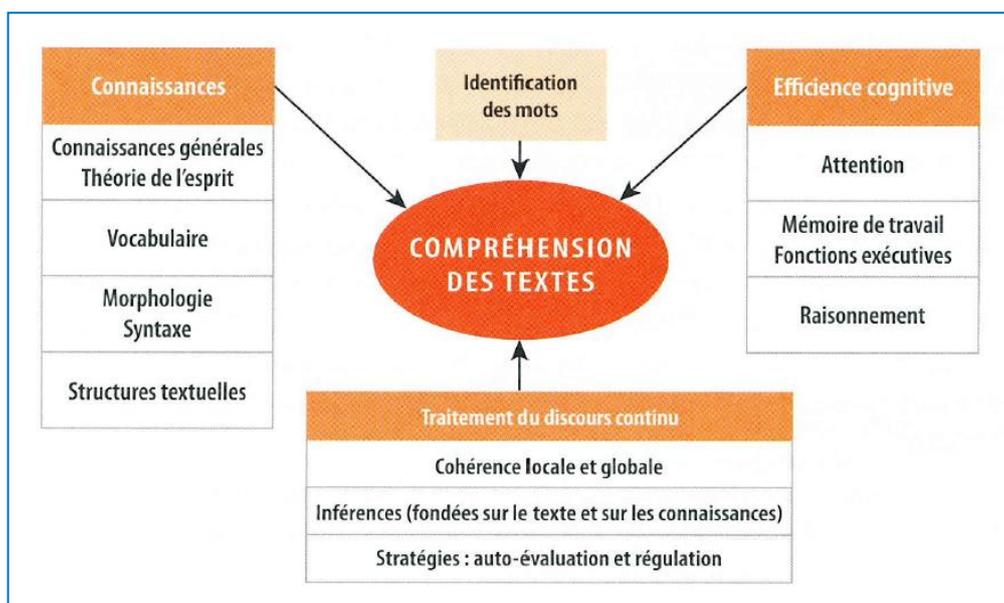
Lors de l'apprentissage de la lecture, l'enfant donne un sens à des signes graphiques au travers de traitements perceptifs et cognitifs. Le mécanisme de décodage joue un rôle fondamental car il permet à l'enfant de convertir des lettres et des chaînes de lettres (graphèmes) en phonèmes lui permettant de retrouver les mots dans son lexique phonologique (Ziegler, Perry & Zorzi, 2014). Par la suite, le décodage, donc la mise en relation entre les graphèmes et les phonèmes, doit être automatisé pour lire de façon aisée les mots qui sont dans le lexique oral de l'enfant afin qu'il puisse extraire le sens d'une phrase et construire une interprétation. La reconnaissance d'un mot écrit ne signifie pas automatiquement compréhension du sens.

« On peut savoir lire un mot sans en connaître la signification (connaissez-vous le sens des mots ciper et phonotactique ? Cela vous empêche-t-il de les lire ?), tout comme on peut très bien connaître la signification d'un mot à l'oral sans pour autant le reconnaître automatiquement, si on ne l'a jamais vu à l'écrit, ce qui est la situation la plus fréquente dans laquelle se retrouvent les apprentis-lecteurs. » (Morais, Pierre & Kolinski, 2003, p.53)

La compréhension s'exerce pour l'oral et pour l'écrit. Ainsi l'élaboration du modèle the “*Simple View of Reading*” (SVR) (Gough & Tunmer, 1986) décrit la compréhension en lecture comme le produit de la reconnaissance des mots et de la compréhension orale. L'identification des mots contraint la compréhension mais ce sont également les composantes cognitives du processus de compréhension qui déterminent l'élaboration réussie d'une représentation mentale à partir du texte (Figure 1). Le traitement du langage, parlé ou écrit, nécessite des opérations

complexes où la mémoire et les processus de contrôle et de manipulation de l'information interviennent. Le modèle SVR continue d'orienter les recherches et a été approfondi grâce à l'évolution des connaissances dans le domaine de la lecture (Bianco & Lima, 2017).

Figure 1 : Structures et mécanismes impliqués dans la compréhension de textes (Bianco & Lima, 2017, p. 21).



Pour construire une représentation mentale du texte, le lecteur construit la cohérence locale et globale du texte en mettant en lien les informations contenues dans le texte et ses connaissances sur le monde (Marin & Legros, 2008). Certains textes peuvent apparaître comme simples à comprendre et d'autres beaucoup moins pour un même lecteur, selon le contenu du texte.

L'objectif ultime de la lecture reste la compréhension. Ne pas accéder à la littéracie<sup>4</sup> est un problème sociétal d'où l'importance d'essayer de lutter contre ce phénomène dès le plus jeune âge. Les termes de « littéracie » ou « lettrure » sont utilisés et définis par l'Organisation de Coopération et de Développement Économiques (OCDE) comme « L'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité

<sup>4</sup> La définition de littératie sur le site Wikipédia<sup>4</sup> précise que le terme provient d'un calque de l'anglais literacy où il désigne initialement la notion d'alphabétisme, dont l'usage explose à partir de 1980. En langue anglaise le concept a été dérivé dans l'expression literacy, en partant du principe qu'il était aussi important de savoir trouver, critiquer et utiliser l'information dans la société de l'information actuelle que de savoir lire et écrire dans la société industrielle. C'est en 1989 que cette expression a été officialisée par la définition qu'en donne l'American Library Association : « Être compétent dans l'usage de l'information signifie que l'on sait reconnaître quand émerge un besoin d'information et que l'on est capable de trouver l'information adéquate, ainsi que de l'évaluer et l'exploiter. ». <https://fr.wikipedia.org/wiki/Litt%C3%A9ratie>

en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités ». Une partie de la population n'accède pas à la littéracie. Lors de la Journée Défense et Citoyenneté (JDC) en 2019, l'évaluation des compétences de lecture chez des jeunes de 16 à 25 ans montre que 22,7% d'entre eux sont des lecteurs « médiocres » et que parmi eux, 6,5% ont de très faibles capacités de lecture et 5,3% ont des difficultés sévères avec un déficit d'automatisme et de vocabulaire qui les rapprochent de la situation d'illettrisme<sup>5</sup> (Chabanon, 2020).

Au regard des enjeux de la lecture et de sa compréhension, nous avons étudié dans quelle mesure la simplification de textes pouvait être un outil pour faciliter la lecture et la compréhension de deux types de textes. La simplification consiste à réduire la complexité d'un texte tout en conservant les informations et le sens original (Siddharthan, 2014). Comme dans les enquêtes internationales, nous avons utilisé des textes littéraires et documentaires scientifiques. Nos recherches ont été menées auprès d'apprentis lecteurs de 7 à 9 ans.

La thèse s'organise en deux parties, une partie théorique et une partie empirique. La partie théorique comporte la présentation des enquêtes nationales et internationales sur la lecture (chapitre 1), la description des modèles de référence et des principaux facteurs qui influencent la compréhension en lecture (chapitre 2), la synthèse des travaux et enjeux de la simplification de textes (chapitre 3) et enfin le développement du corpus simplifié de textes scolaires et son analyse au travers des variables textuelles choisies (chapitre 4). La partie empirique comporte trois grandes études. Le chapitre 5 présente les questions spécifiques de chacune de ces études. Les chapitres 6 à 8 présentent la première étude qui avait pour but d'analyser les effets de la simplification sur trois niveaux avec un suivi de cohorte sur trois ans du CE1 au CM1. Le chapitre 9 présente la deuxième étude, qui avait pour but d'analyser les effets d'une simplification, lexicale, syntaxique et lexico-syntaxique pour des élèves de CE1. Le chapitre 10 présente la troisième étude qui avait pour but d'analyser les effets de trois types d'entraînements de lecture sur trois mois en classe de CE2. Le chapitre 11 présente une discussion générale, la synthèse des études ainsi que les limites et perspectives de ces travaux. Enfin, le chapitre 12 présente le développement de deux ressources numériques issues de ce travail de thèse avant la conclusion.

---

<sup>5</sup> On parle d'illettrisme, selon l'Agence Nationale de Lutte Contre l'Illettrisme (ANLCI) pour des personnes qui, après avoir été scolarisées en France, n'ont pas acquis une maîtrise suffisante de la lecture, de l'écriture, du calcul, des compétences de base, pour être autonomes dans les situations simples de la vie courante. <http://www.anlci.gouv.fr/Illettrisme/De-quoi-parle-t-on/Les-definitions>

## Chapitre 1

### Évaluations de la compréhension en lecture de textes littéraires et scientifiques

Pour mieux situer le niveau des élèves français en lecture et compréhension, dans ce premier chapitre, nous présentons les résultats de différentes enquêtes internationales qui comparent les performances d'élèves de plusieurs pays, puis nous les complétons par des enquêtes nationales.

Ces enquêtes montrent que les performances en compréhension de lecture des élèves français ont baissé sur une période de quinze ans et que 39% d'entre eux se retrouvent dans le premier quartile des plus faibles lecteurs au lieu des 25% attendus en fin d'école primaire. De plus, leurs compétences en analyse et réflexion sur un texte ont davantage baissé que leurs compétences en prélèvement des informations présentes dans un texte.

#### 1.1 Enquêtes internationales

##### *Progress in International Reading Literacy Study (PIRLS)*

L'étude internationale PIRLS mesure les performances en compréhension de l'écrit des élèves en fin de quatrième année de scolarité obligatoire (CM1 pour la France). En 2016, les évaluations internationales PIRLS ont porté sur les performances en lecture et en compréhension dans 50 pays. Dans ces évaluations, les élèves ont lu des textes narratifs et des documentaires scientifiques ce qui a permis une comparaison entre les deux types de textes. Les scores montrent que les élèves français obtiennent de moins bons résultats sur les deux types de textes par rapport au score moyen des élèves de l'Union Européenne. Pour les textes narratifs, le score de la France est de 513, alors que celui de l'Union Européenne est de 542. Pour les textes informatifs, la France obtient un score de 510 alors que celui de l'Union Européenne est de 539 (Mullis, Martin, Foy & Hooper, 2017). Vingt pays disposent de données sur une période de 15 ans (2001 à 2016) ce qui permet une analyse de la stabilité des résultats. En effet, on constate une augmentation de la performance dans 11 pays, dont quatre avec des gains de plus de 40 points (Hong Kong, Russie, Singapour et Slovaquie). Les résultats sont stables pour sept pays (Bulgarie, Angleterre, Allemagne, Lituanie, Nouvelle-Zélande, Suède, Etats-Unis) et les performances ont baissé pour deux pays, la France et les Pays-Bas. Les performances basées sur la compréhension de textes informatifs baissent davantage (- 22 points)

que celles des textes narratifs (- 6 points). Les processus de compréhension les plus complexes « Inférer » baissent davantage (- 21 points) que les plus simples « Prélever » (- 8 points).

Cette baisse de performance ne peut pas être expliquée par une baisse du temps d'enseignement en France. Dans l'enquête PIRLS de 2016 (Colmant & Le Cam, 2017)<sup>6</sup>, les enseignants français affirment effectuer 330 heures pour l'étude de la langue, soit 15% de temps supplémentaire par rapport aux 288 heures inscrites dans les programmes de 2008. Pour les enseignants des autres pays européens, la moyenne annuelle déclarée pour cet enseignement est de 236 heures. Concernant la compréhension en lecture spécifiquement, les enseignants français réservent 165 heures à cette activité alors que dans les autres pays européens ce temps est de 146 heures. Ces données montrent que 37% du temps de classe est réservé à l'enseignement de la langue française et que 19% du temps de classe est réservé spécifiquement à la compréhension en lecture. En moyenne dans les autres pays de l'Union Européenne, 28% du temps de la classe est réservé à l'enseignement de la langue et que 18% du temps de classe est réservé spécifiquement à la compréhension en lecture. De même, les enseignants français disent aussi proposer, à leurs élèves, moins d'activités susceptibles de développer leurs stratégies et leurs compétences en compréhension de l'écrit par rapport à leurs collègues européens (Colmant & Le Cam, 2017).

### ***Programme for International Student Assessment (PISA)***

Tous les trois ans depuis 2000, l'évaluation internationale PISA évalue les compétences des élèves de 15 ans dans trois domaines : la compréhension de l'écrit, la culture mathématique et la culture scientifique. Les résultats en compréhension de l'écrit sont stables depuis 2009 pour la France, avec un score de 493 points en 2018, au-dessus de la moyenne des pays de l'OCDE située à 487 points (Chabanon, Durand de Monestrol & Verlet., 2019). En culture scientifique, avec 493 points, la France se situe aussi au-dessus de la moyenne de l'OCDE qui atteint 489 points (Bernigole, Bret, Chabanon, Roussel & Verlet, 2019). Les performances dans les trois domaines évalués, la compréhension de l'écrit, la culture mathématique et la culture scientifique, sont réparties sur 6 niveaux de compétences, avec 3 degrés d'approche pour le niveau 1, considéré comme le plus bas niveau<sup>7</sup>.

En compréhension de l'écrit, environ 79% des élèves français ont atteint au moins le niveau 2 (niveau moyen), une proportion proche de la moyenne de l'OCDE qui est à 77%. Ces élèves sont capables d'identifier l'idée principale dans un texte de longueur moyenne, de trouver

---

<sup>6</sup>[http://cache.media.education.gouv.fr/file/2017/73/7/depp-ni-2017-24-pirls-cm1-ecrit\\_860737.pdf](http://cache.media.education.gouv.fr/file/2017/73/7/depp-ni-2017-24-pirls-cm1-ecrit_860737.pdf)

<sup>7</sup> [https://www.oecd.org/pisa/publications/PISA2018\\_CN\\_FRA\\_FRE.pdf](https://www.oecd.org/pisa/publications/PISA2018_CN_FRA_FRE.pdf)

des informations fondées sur des critères explicites, et parfois complexes, et de réfléchir au but et à la forme des textes quand on le leur demande explicitement (Givord & Schwabe, 2019). Les élèves français sont plus à l'aise avec les éléments leur demandant de « localiser l'information » dans un texte (496 points en moyenne) qu'avec ceux leur demandant de « comprendre » l'ensemble du texte (490 points en moyenne) ou « analyser et réfléchir » sur un texte (491 points en moyenne).

En culture scientifique, les performances moyennes n'ont pas connu d'évolution notable entre 2006 et 2018 en France. Environ quatre élèves sur cinq ont atteint le niveau 2 (niveau moyen) en sciences (78 %, contre 79 % en moyenne dans les pays de l'OCDE). Au minimum, ces élèves peuvent reconnaître l'explication correcte de phénomènes scientifiques familiers et utiliser ces connaissances pour identifier, dans des cas simples, la validité d'une conclusion sur la base des données fournies (Givord & Schwabe, 2019).

## 1.2 Enquêtes nationales

Plusieurs évaluations nationales sont réalisées en France, du CP à la troisième. Nous avons choisi de présenter quelques résultats pertinents issus des évaluations disciplinaires réalisées sur échantillon (CEDRE) et des évaluations nationales du CE1.

### *Cycle des Évaluations Disciplinaires Réalisées sur Échantillon (CEDRE)*

CEDRE établit les bilans nationaux des acquis des élèves en fin d'école et en fin de collège en évaluant les connaissances et compétences des élèves dans la plupart des disciplines, en référence aux programmes et en tenant compte des pratiques des enseignants. D'après les résultats de CEDRE 2009<sup>8</sup>, 39 % des élèves sont en difficulté à la sortie de l'école primaire. Ces élèves ne sont pas capables d'identifier le thème ou le sujet principal d'un texte, de comprendre les informations implicites d'un texte et de lier deux informations explicites séparées dans le texte. D'après les résultats de CEDRE 2015<sup>9</sup>, 11% des élèves de CM2 appartiennent au groupe des très faibles lecteurs. Parmi eux, 1 % sont en grande difficulté de lecture, ils sont capables de prélever directement une information qui est placée au début du texte et ils savent identifier une famille de mots et le sens d'une expression familière (Andreu et al, 2016). Par ailleurs, d'après les résultats de CEDRE 2018, évaluations des compétences

---

<sup>8</sup> [http://www.cnesco.fr/wp-content/uploads/2017/04/Lecture\\_dossier\\_synthese.pdf](http://www.cnesco.fr/wp-content/uploads/2017/04/Lecture_dossier_synthese.pdf)

<sup>9</sup> [https://www.education.gouv.fr/sites/default/files/imported\\_files/document/DEPP-CEDRE-maitrise-de-la-langue-2015-ecole-rapport-technique\\_915443.pdf](https://www.education.gouv.fr/sites/default/files/imported_files/document/DEPP-CEDRE-maitrise-de-la-langue-2015-ecole-rapport-technique_915443.pdf)

dans le domaine des sciences (Augé, Etève & Ninnin, 2019), 8 élèves sur 10 déclarent aimer apprendre des notions nouvelles pendant les séances de sciences à l'école et pour 85 % des élèves ayant participé à ces évaluations, faire des sciences à l'école c'est se poser des questions sur le monde qui les entoure.

### Évaluations 2019, Repères Niveau CE1

Depuis 2017, le Ministère de l'Éducation Nationale met à la disposition des professeurs des évaluations pour leur permettre de mieux identifier les besoins de leurs élèves et personnaliser encore davantage leurs propositions pédagogiques. Après les passations des évaluations de CE1 en français en septembre 2019 les élèves ont été répartis en 3 groupes (Tableau 1) en fonction de leurs résultats<sup>10</sup> (Andreu et al, 2019) :

- ceux pour lesquels on peut considérer, dès le début de l'année, qu'ils sont en difficulté : un besoin est identifié. Ces élèves nécessitent un accompagnement (Groupe sous le seuil 1) ;
- ceux dont les acquis sont fragiles pour lesquels l'enseignant devra maintenir un niveau de vigilance particulier (Groupe entre les seuils 1 et 2) ;
- ceux pour lesquels il n'y a pas de difficultés identifiées, les pré-requis permettent d'entrer sereinement dans les apprentissages à venir (Groupe au-dessus du seuil 2).

Tableau 1 : Résultats des évaluations Repères, dans le domaine du français, en début de CE1, rentrée 2019.

Domaine évalué	Grpe sous le seuil 1	Grpe entre seuils 1 et 2	Grpe au-dessus du seuil 2
Lire à voix haute un texte	10,1	18,5	71,4
Lire à voix haute des mots	9,4	18,1	72,5
Ecrire des syllabes simples et complexes	6,8	7,7	85,5
Ecrire des mots	9,8	13	77,1
Comprendre un texte lu seul	4,5	10,9	84,6
Comprendre des phrases lues seul	6,7	10,4	82,9
Comprendre des phrases lues par l'enseignant	2,9	14,7	82,4
Comprendre des mots lus par l'enseignant	6,3	17,1	76,6

Note : par le jeu des arrondis, la somme des groupes peut être différente de 100.

Ce Tableau 1 montre que le pourcentage d'élèves sous le seuil 1 en compréhension de phrases lues par l'enseignant est de 2,9% alors que dans la tâche de compréhension de phrases lues seul, le pourcentage d'élèves en dessous du seuil 1 est de 6,7%. Cet écart montre que la capacité de décodage représente encore une part importante dans la tâche de compréhension de phrases

<sup>10</sup> [https://cache.media.education.gouv.fr/file/doctrav2019/05/6/dt2019e03\\_1198056.pdf](https://cache.media.education.gouv.fr/file/doctrav2019/05/6/dt2019e03_1198056.pdf)

pour certains élèves, nous allons y revenir dans la section 2.2.1. Cette hypothèse est renforcée par le pourcentage d'élèves en dessous du seuil 1 dans la tâche de lecture de mots à haute voix. En lien avec ces évaluations, des ressources sont proposées aux professeurs afin d'accompagner leurs élèves.

### **Conclusion**

Suite à l'analyse de ces évaluations, nous constatons que le pourcentage des élèves en difficultés représente 18,4% en CE1, sur les 4 items liés à la compréhension ; il est de 28% en CM1 et monte à 33,4% des élèves scolarisés à l'âge de quinze ans. Le pourcentage d'élèves entre le CE1 et l'âge de 15 ans augmentant considérablement et particulièrement entre le CE1 et le CM1, la simplification de textes nous a semblé être un outil pouvant aider les élèves en lecture et compréhension spécifiquement à l'école élémentaire.

## Chapitre 2

### La compréhension : mécanismes, modèles, facteurs, difficultés et interventions

Selon Content et Peereman (1999), la lecture comporte trois grandes étapes : le traitement perceptif, la reconnaissance des mots écrits et la compréhension. Lors du traitement perceptif, le lecteur extrait des informations visuelles, c'est-à-dire l'identité des lettres et leur position relative. Le traitement perceptif se réalise par la fixation du regard, rythmée par des saccades oculaires permettant au lecteur de parcourir l'ensemble du texte (Rayner, 1998). Si l'étape de reconnaissance des mots écrits est réussie, le lecteur identifie le mot dans son « lexique mental » ce qui permet de « retrouver » le sens du mot. Souvent, le lecteur est amené à déterminer la signification la plus appropriée en tenant compte de la fonction du mot dans la phrase et du sens des mots qui l'entourent (le contexte).

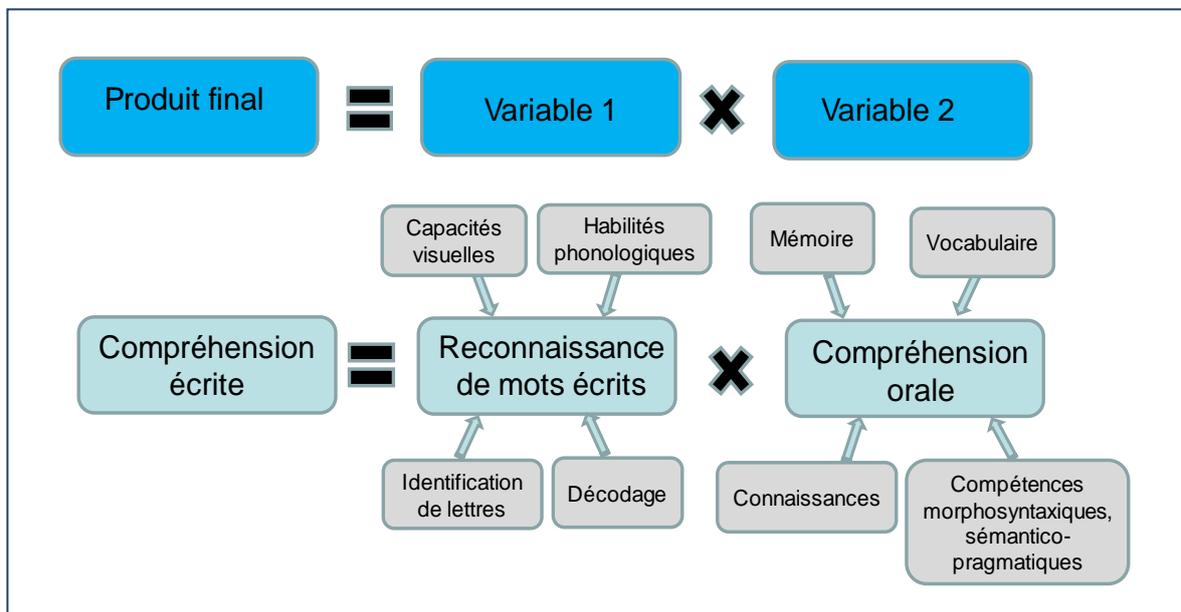
Plusieurs modèles théoriques formalisent les mécanismes de la compréhension de textes. Quatre ont été sélectionnés : le modèle simple de la lecture (Lervåg, Hulme, & Melby-Lervåg, 2018), le modèle de construction de structures (Gernsbacher, 1990), le modèle de Construction-Intégration (Kintsch, 2005) et la qualité lexicale (Perfetti & Stafura, 2014). Après la présentation des modèles, nous décrirons les principaux facteurs qui influencent la compréhension de l'écrit, car c'est sur ces facteurs qu'il faudra mettre en place des simplifications ou adaptations.

#### 2.1. Les modèles théoriques de la compréhension

##### 2.1.1 Modèle Simple de la lecture

Dans le modèle Simple de la lecture (*Simple View of Reading* : SVR), Gough et Tunmer (1986) puis Hoover et Gough (1990) décrivent la compréhension écrite ( $R = \text{reading comprehension}$ ) comme le produit du décodage de l'écrit ( $D = \text{decoding ability}$ ) par la compréhension orale ( $C = \text{comprehension skill}$ ) :  $R = D \times C$  (Figure 2). Ce rapport multiplicatif implique que chaque facteur est nécessaire mais qu'aucun n'est suffisant. En effet, Morais (1993) constate que, chez le lecteur expert qui a automatisé les processus de reconnaissance de mots, la relation est très forte entre la compréhension écrite et la compréhension orale alors que chez le lecteur novice qui n'a pas encore automatisé les procédures de reconnaissance de mots écrits, la relation est très forte entre la compréhension écrite et le décodage.

Figure 2 : Modèle "simple" de la compréhension en lecture (Sprenger-Charolles & Ziegler, 2019, p. 96).



Comme le montre la Figure 2, les deux variables sont elles-mêmes influencées par de nombreux facteurs, comme les capacités visuelles, le décodage, les compétences morphosyntaxiques, le vocabulaire etc. La reconnaissance des mots écrits et la compréhension orale impliquent de nombreuses fonctions cognitives, visuelles, morphologiques, phonologiques, grammaticales, mémoire de travail, vocabulaire, capacités d'inférence et connaissances. Chacun de ces facteurs peut influencer la reconnaissance des mots écrits et la compréhension orale et, par conséquent, la vitesse de lecture et la compréhension du texte (Castles, Rastle & Nation, 2018). Nous reviendrons sur chacun de ces facteurs dans la section suivante. Le modèle SVR a été validé dans de nombreuses études dont certaines sont présentées ci-dessous.

Dans une étude en longitudinal auprès de 206 élèves du CP au CM1, sur quatre ans, avec 55 sujets la dernière année, Hoover et Gough (1990) se sont intéressés à trois hypothèses pour préciser les relations entre les facteurs du modèle simple de la lecture. Les résultats de la première expérience ont montré que si la combinaison linéaire du décodage et de la compréhension orale a largement contribué à expliquer la variation de la compréhension en lecture, l'explication de la variance a été sensiblement améliorée par l'inclusion du produit des deux composantes. Les résultats de la deuxième expérience ont montré que les contributions du

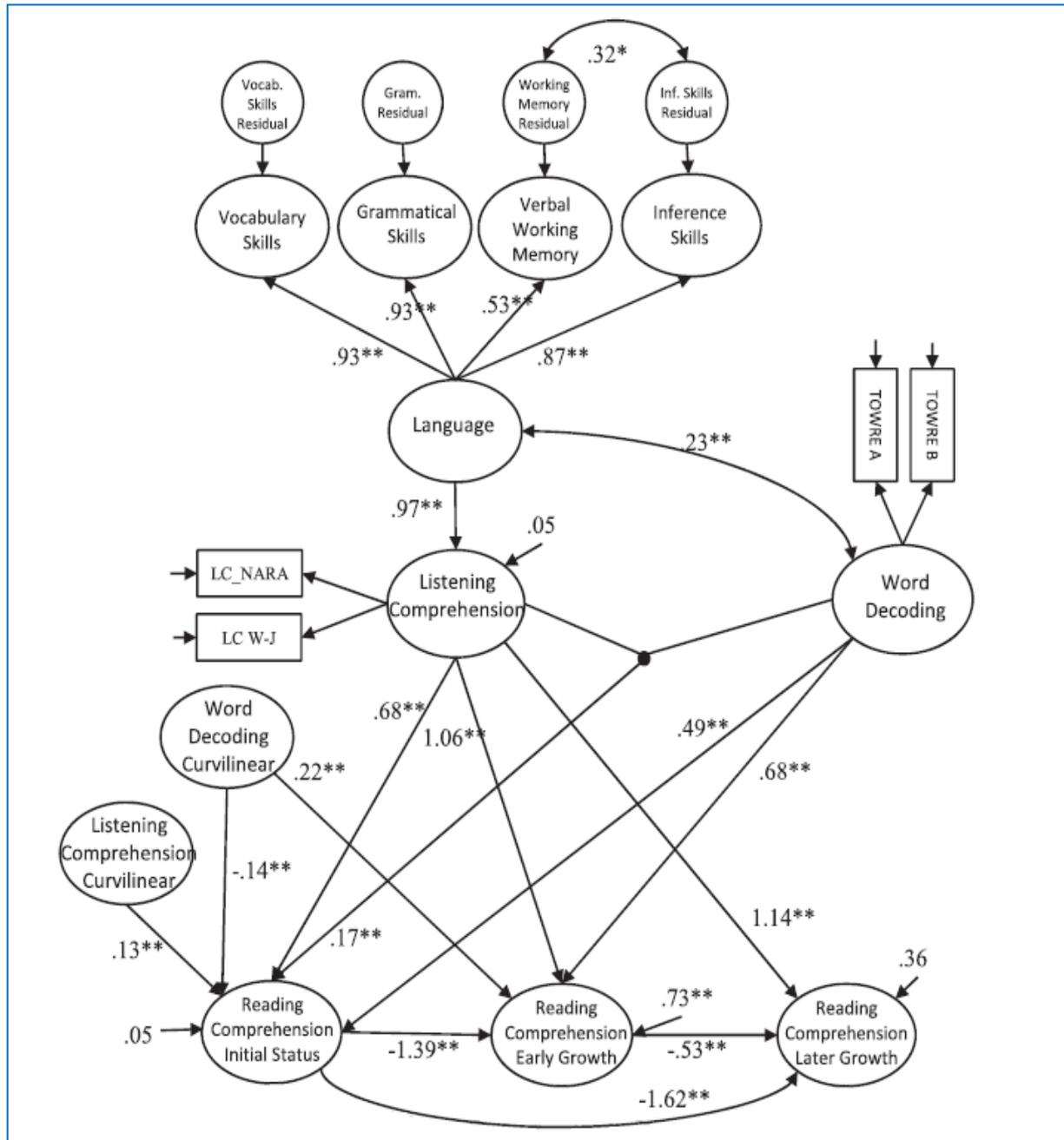
décodage et de la compréhension linguistique sont nécessaires. La compréhension linguistique est définie par la capacité à prélever des informations lexicales et à interpréter des phrases et le discours grâce à des compétences morphosyntaxiques, cette capacité qui existe à l'oral doit aussi s'exercer à l'écrit à partir des informations graphiques. Une faiblesse dans l'une des deux composantes suffit pour devenir un faible 'compreneur'. Les résultats de la troisième expérience ont démontré que la compétence en lecture en tant que produit d'une compétence en décodage et d'une compétence en compréhension linguistique augmente lorsque l'un des deux facteurs progresse.

Le modèle SVR a été également confirmé dans des études francophones. Par exemple, dans le cadre de l'enquête Lire et Écrire au CP (Ifé-ACTé), Goigoux, Cèbe et Pironom (2016) ont analysé les facteurs explicatifs au niveau de la compréhension de textes narratifs à la fin du cours préparatoire en réalisant des analyses de régression sur les performances de 2 395 élèves. Les facteurs expliquant la compréhension écrite étaient le décodage, la compréhension du langage (phrases ou textes entendus) et le vocabulaire.

Cependant, il a été suggéré par Tunmer et Chapman (2012) que l'hypothèse des composantes indépendantes du modèle SVR devrait être quelque peu assouplie, car la compréhension du langage oral (C) semble influencer la compréhension de la lecture (R) non seulement directement, mais aussi indirectement par la reconnaissance des mots (D). La connaissance du vocabulaire contribue directement à la variance de la reconnaissance des mots.

Une étude récente est assez exemplaire quant à l'estimation très sophistiquée et exhaustive des deux composantes indépendantes du SVR (Lervåg, Hulme & Melby-Lervåg, 2018). Cette étude a été réalisée auprès de 198 enfants norvégiens âgés de 7,5 ans qui ont été suivis sur 5 ans. Le modèle est présenté dans la Figure 3. Le facteur langage est estimé par le biais de quatre variables latentes (vocabulaire, grammaire, mémoire de travail verbale et les compétences d'inférence). La mémoire de travail verbale et les compétences d'inférences ont une variance commune partagée dans le facteur de langage. La compréhension orale mesurée avec deux tests classiques est presque isomorphe avec le facteur de langage. Le facteur langage explique 90% de la variance de la compréhension orale. De plus, l'influence du facteur langage sur la croissance de la compréhension en lecture est entièrement médiée par la compréhension orale qui prédit au départ mais aussi plus tard la compréhension en lecture.

Figure 3 : Modèle d'équation structurelle schématisant les effets de la compréhension orale et du décodage de mots sur la croissance des compétences de compréhension en lecture d'après Lervåg, Hulme et Melby-Lervåg (2018, p.1831).

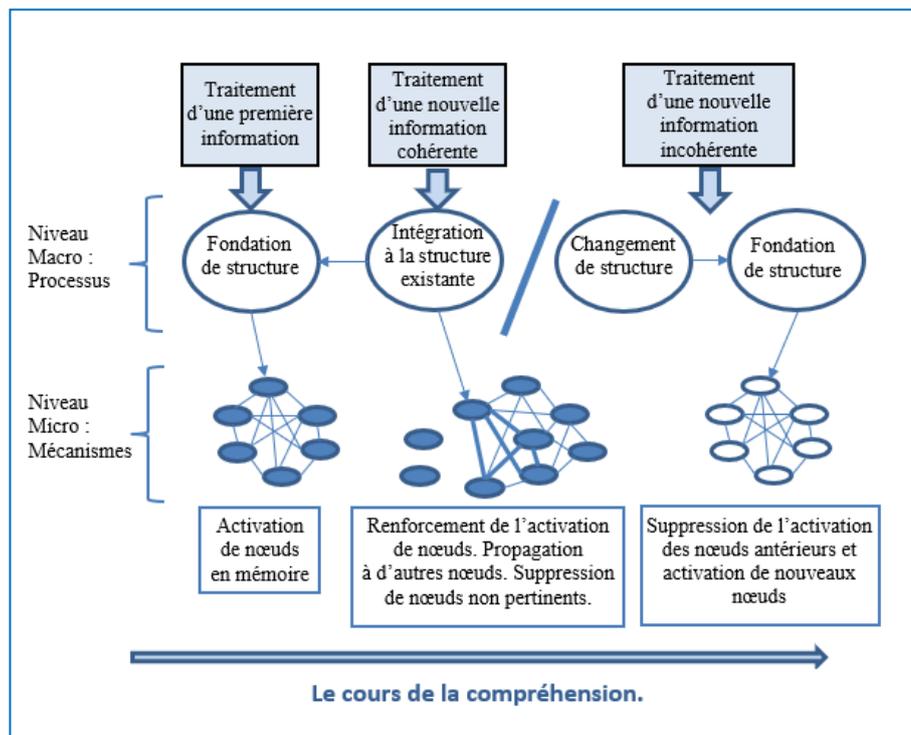


Le décodage est influencé par le langage oral et prédit l'état initial de la compréhension en lecture et la courbe de croissance individuelle (early growth). En résumé, la compréhension orale et le décodage des mots, ainsi que leurs interactions et leurs effets curvilignes, expliquent 96% de la variance dans les trajectoires individuelles de compréhension de lecture.

### 2.1.2 Modèle de construction de structures

Dans le modèle, « *The structure building framework* », Gernsbacher (1990) a présenté la compréhension comme une construction mentale qui s'élabore à partir d'éléments présents dans le texte lu ou entendu (Figure 4). Les premières informations prélevées construisent une représentation mentale qui servira de base à la compréhension. Au fil de la découverte de l'écrit, d'autres informations sont intégrées à cette première représentation qui conduisent à créer une nouvelle structure ou une sous-structure par un processus d'intégration (si les informations sont cohérentes) ou un processus de changement (si les informations ne sont pas cohérentes).

Figure 4 : Modèle de la compréhension de Gernsbacher d'après Blanc et Brouillet (2003).



Ces trois processus (fondation, intégration, changement) sont soutenus par deux mécanismes associés à la mémoire : la suppression ou le renforcement.

« Pour construire la signification d'un texte, il est nécessaire de construire sa cohérence locale « microstructure » et globale « macrostructure » en mettant en rapport différents types d'informations, issues du texte et les connaissances du lecteur stockées en mémoire et activées lors de la lecture. Ce sont ces connaissances qui permettent d'assurer la cohérence locale et globale. » (Marin, & Legros, 2008, p.62)

**La fondation** représente la base d'une représentation mentale. Les 'compreneurs' consacrent plus d'efforts cognitifs au traitement d'informations en position initiale que plus tard dans le discours. Cependant, ceci ne vaut que pour des discours cohérents. Gernsbacher conclut que dans un discours cohérent les 'compreneurs' utilisent les informations initiales pour construire les bases de leur représentation de l'ensemble du discours.

### **L'intégration « Mapping » ou le changement « Shifting »**

Pour construire leur représentation mentale, les 'compreneurs' ajoutent à ces bases les informations ultérieures grâce aux processus de renforcement ou de suppression. Les liens, qui se créent entre les informations du texte et la représentation qui se construit, se renforcent, se modifient, c'est le processus d'appariement « mapping », au regard des connaissances mémorisées, « nœuds de la mémoire », pour garder sa cohérence à la compréhension de l'écrit. Lorsque les informations nouvelles ne correspondent pas aux informations précédentes le lecteur ne peut pas les intégrer à son développement de structure. Dans ce cas, le lecteur opère un changement dans sa structure et développe une nouvelle sous-structure. Le développement de la structure mentale s'appuie sur quatre niveaux de cohérence. La cohérence, entre deux phrases principalement, relève de plusieurs niveaux :

- La cohérence référentielle : si la référence (le sujet, le personnage notamment) reste identique, alors la phrase est lue plus rapidement ;
  - La cohérence temporelle : si le cadre temporel est le même entre deux faits, alors la compréhension est plus rapide ;
  - La cohérence spatiale : si les événements se déroulent ou non dans le même espace ;
  - La cohérence causale : si une logique entre les événements est maintenue dans l'écrit.
- Ces niveaux permettent de construire la cohérence globale du texte. Chacun de ces niveaux participe à la cohérence du discours et aucun n'est suffisant.

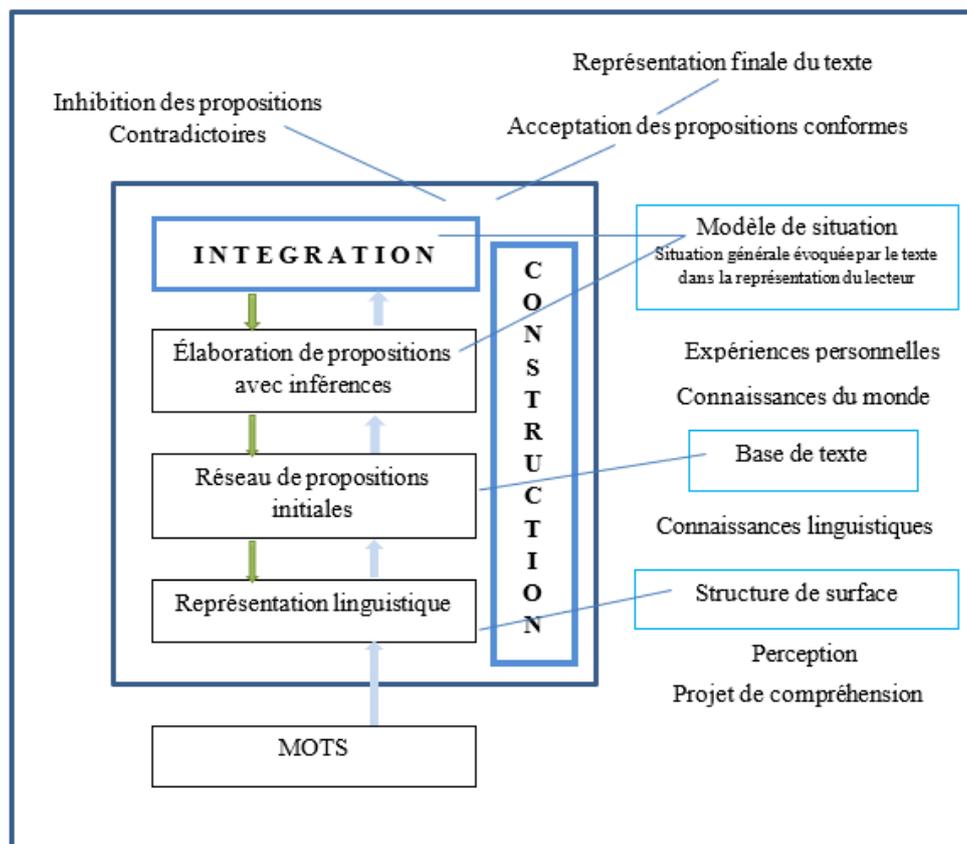
### **Suppression ou renforcement**

Des structures mentales sont représentées dans les nœuds de la mémoire. Ces nœuds de la mémoire ont stocké des traces qui sont automatiquement activées par les stimuli entrants. Selon le cadre de la représentation mentale du 'compreneur', les nœuds activés transmettent des signaux de traitement qui renforcent ou suppriment l'activation d'autres cellules de mémoire. Les représentations mentales des constituants du discours dans la mémoire sont égales mais hiérarchisées. Certaines informations bénéficient d'un meilleur rappel et d'autres sont moins accessibles. Par exemple, le mécanisme de suppression freine l'activation des significations les moins probables.

### 2.1.3 Modèle de Construction-Intégration (C-I)

Dans le modèle C-I, Kintsch (1988) propose que, lorsque les significations des mots sont activées, des propositions sont formées et des inférences et des élaborations sont produites sans tenir compte du contexte du discours dans un premier temps. Ces processus qui aboutissent à la construction du sens, se déroulent dans la mémoire de travail du 'compreneur', par un échange incessant entre la perception et la mémoire à long terme. Selon Kintsch (1998), la lecture n'est pas une résolution active de problème. En revanche, en cas d'échec du mécanisme de satisfaction de contraintes conduisant à l'intégration le processus de résolution de problème s'active. Le modèle C-I présente la « satisfaction de contraintes » comme base de la théorie de la compréhension de textes. La compréhension de textes se base sur deux types de processus, le processus de construction et le processus d'intégration et sur deux structures, le texte et le modèle de situation (Figure 5). Le modèle de situation, intégré au modèle C-I, est défini comme « une représentation cognitive des événements, des actions, des individus et de la situation générale évoquée par le texte » (van Dijk & Kintsch, 1983, p. 11-12).

Figure 5 : Modèle de Construction-Intégration. Schéma réalisé d'après Kintsch (2005).



Dans le modèle C-I, de nombreuses inférences, pertinentes ou non, sont faites en parallèle. Le processus d'intégration les désactive rapidement lorsqu'elles ne sont pas pertinentes.

L'intégration est une satisfaction de la contrainte car les propositions venant du texte et les propositions inférées se renforcent mutuellement lorsqu'il y a cohérence et sont inhibées si elles ne rentrent pas dans le contexte.

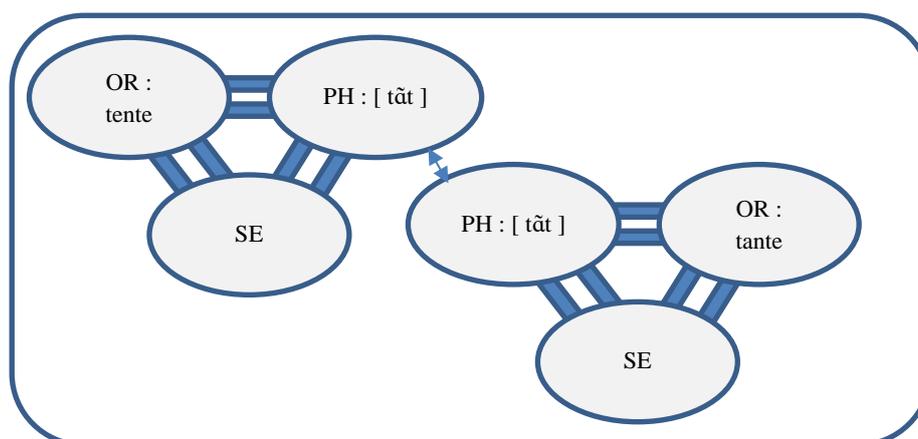
« Le système de représentation dont nous avons besoin doit remplir de multiples fonctions : il doit servir à la représentation mentale des textes mais aussi à d'autres structures de mémoire, telles que les connaissances générales et la signification des mots, et les expériences personnelles (traces mémorielles épisodiques). Parce que les représentations mentales du texte sont en partie dérivées de la connaissance et de l'expérience, il est souhaitable que les deux puissent être décrites dans le même format. » [Traduction libre] (Kintsch 1998, p.34)

Dans le modèle CI, les mots du texte qui activent les représentations linguistiques et les significations des mots servent à construire des propositions sans contrôle descendant, « la base du texte ». Dans la phase de construction émergent des représentations correctes ou non-pertinentes. Ces propositions non-cohérentes et contradictoires seront résolues par un processus d'intégration. Dans les processus ascendants et descendants, les connaissances activées fonctionnent comme des unités de contrôle. En 2005, Kintsch explique que dans le modèle C-I la contrainte qu'exercent les processus ascendants sur les processus descendants guide la compréhension. Le processus ascendant n'est pas plus important que le processus descendant parce que les deux processus interagissent et sont nécessaires pour accéder à la compréhension. Les processus perceptifs impliqués dans la lecture ou la compréhension orale, les analyses syntaxiques et sémantiques, les compétences morphosyntaxiques, les connaissances du monde, ainsi que les expériences et le raisonnement influencent les processus de construction et d'intégration à tous les niveaux de l'élaboration d'une représentation mentale conforme au modèle de situation. Les processus descendants ne dominent jamais, sauf dans les cas pervers, dans lesquels ils guident fortement la compréhension. Le lecteur recherche la signification du texte en confirmant et en infirmant des propositions mais également en tenant compte du contexte de lecture. Le traitement de l'information lue est dirigé par les connaissances du lecteur. Chaque fois que la prédiction échoue, le lecteur fait une nouvelle recherche d'informations donc une nouvelle hypothèse (Stanovich, 1980). Dans le modèle C-I, le rôle du lecteur ne se limite pas à prendre en compte les connaissances préalables, qui contribuent considérablement à l'interprétation du texte, car le projet de compréhension du lecteur et ses objectifs de lecture participent aussi à l'interprétation du texte.

#### 2.1.4 La qualité lexicale dans la lecture

Perfetti & Hart (2002) et Perfetti (2007) ont proposé le modèle de la qualité lexicale. La qualité lexicale décrit les relations entre les représentations phonologiques, orthographiques et sémantiques du mot. Les liens qui unissent ces trois constituants déterminent la qualité de la représentation du mot, l'exactitude et la rapidité de son identification. Perfetti (2007) a défini la qualité lexicale comme une mesure d'une représentation mentale stockée d'un mot spécifié par sa forme et sa signification d'une manière à la fois précise et réutilisable. La précision de la représentation, la connaissance de l'orthographe exacte, permet à un enfant de distinguer un mot écrit de mots d'apparence similaire, permettant ainsi un accès direct à sa signification ; par exemple les mots orthographiquement voisins : lire/dire. De plus, pour un lecteur donné, certains mots fréquemment rencontrés dans son lexique sont de « haute qualité », tandis que d'autres mots moins connus sont de « faible qualité ». Mais au fur et à mesure que l'enfant développe son expérience en lecture, la qualité lexicale augmente régulièrement. L'utilisation fréquente de la représentation du mot dans des contextes multiples est importante car elle permet à un enfant de s'adapter dynamiquement à différentes combinaisons de la forme écrite et à leurs significations (Figure 6). Par exemple, la distinction à faire entre la « porte » et il «porte», homographes, et la distinction entre deux homophones : la tente/la tante montrent que la connaissance des mots a des conséquences sur les processus de signification des mots dans la compréhension.

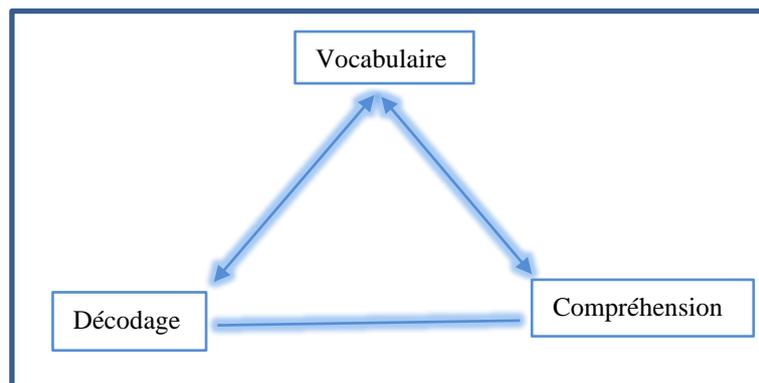
Figure 6 : Entrées lexicales interconnectées par l'entrée phonologique (Perfetti & Hart, 2002, p.194).



Ainsi, la rencontre d'une entrée orthographique par exemple « tante » ou « tente », ayant une forme phonologique [tât] identique pourrait perturber la compréhension d'un apprenti lecteur qui n'aurait pas connaissance de la sémantique singulière de ces deux mots. Perfetti, McKeown

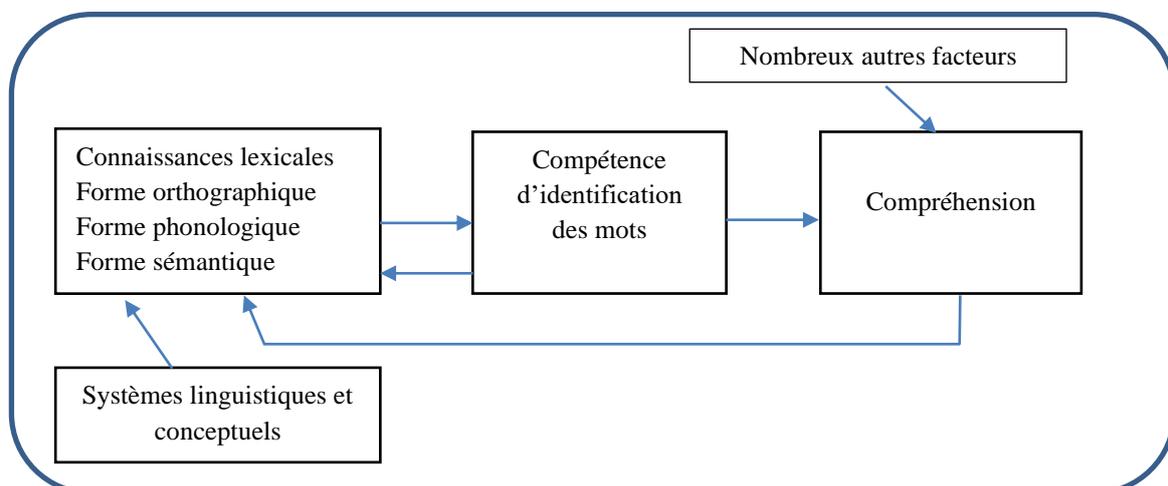
& Kucan (2010) ont proposé un schéma (Figure 7) triangulaire des compétences de lecture Décodage-Vocabulaire-Compréhension (DVC) pour montrer les relations générales entre le décodage, le vocabulaire et la compréhension.

Figure 7 : Triangle DVC. Les capacités de décodage, de vocabulaire et de compréhension se combinent pour produire des compétences générales en lecture (Perfetti, McKeown & Kucan 2010, p. 293).



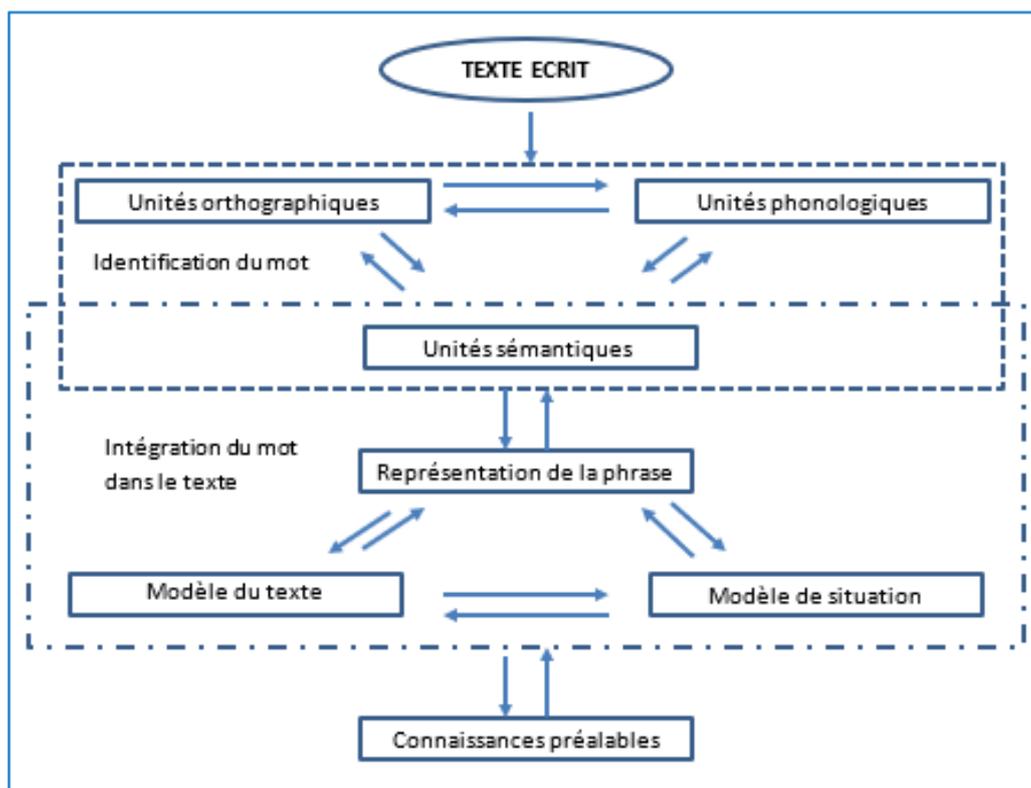
La qualité lexicale, qui englobe la connaissance des formes de mots, les connaissances phonologiques, orthographiques et morphémiques, influence la compréhension de la lecture de façon plus ou moins directe. Or dans le triangle DVC, Perfetti et al (2010) expose l'idée que les distinctions entre les connaissances orthographiques et phonologiques et entre l'identification et le décodage de mots, qui appartiennent à la qualité lexicale, manquent. Afin de tenir compte de la qualité lexicale, Perfetti et al (2010) ont schématisé le flux linéaire (Figure 8), avec les connaissances lexicales (forme et sens des mots), vers les processus d'identification et de compréhension des mots, avec un retour de la compréhension vers la connaissance des mots.

Figure 8 : Schéma simple du flux d'informations, de la connaissance sur la forme et la signification des mots à la compréhension, par l'identification des mots (Perfetti, McKeown & Kucan, 2010, p. 296).



La qualité lexicale varie en fonction de la composante de connaissance des mots et affecte l'identification et la compréhension des mots et elle est à son tour modifiée par les deux : identification des mots (par ex. auto-apprentissage) et compréhension (par ex., acquérir des significations de mots). La qualité lexicale a des conséquences sur la compréhension par les processus qu'elle engage dans la signification des mots. Le vocabulaire a une fonction centrale dans la compréhension de par sa position car il est à l'interface entre l'identification des mots et l'accès à la signification. Une représentation lexicale de qualité est une représentation qui spécifie à la fois la forme du mot, sa phonologie, sa morphologie, son orthographe et ses significations de manière précise et flexible. De plus une bonne représentation lexicale permet de récupérer efficacement et simultanément l'ensemble des informations dès la perception du mot écrit. Ainsi, les personnes ayant une bonne qualité lexicale apprennent de nouveaux mots plus facilement (Perfetti, 2007). Perfetti et Stafura (2014) ont replacé la lecture du mot en contexte au centre du cadre des deux systèmes qui soutiennent la compréhension : le système d'identification des mots et le système de la compréhension des mots dans le texte (Figure 9). Les unités sémantiques activées avec l'identification du mot font partie de ce processus. Au niveau du processus de compréhension du mot dans le texte, c'est un processus d'adaptation qui fait référence aux connaissances préalables et qui peut être corrigé car le choix du sens est influencé par la représentation immédiate du texte par le lecteur.

Figure 9: Compréhension des mots dans le cadre des systèmes de lecture (Perfetti & Stafura, 2014, p. 33).



Les auteurs s'appuient sur le modèle C-I avec la phase de construction, identification des mots (partie supérieure de la Figure 9) et la phase d'intégration, système de compréhension des mots (partie inférieure de la Figure 9). Ce processus est principalement guidé par la mémoire. Les mots du texte récemment lus et les propositions de sens qui en découlent (modèle du texte) restent accessibles en mémoire. Dans le système de compréhension des mots, les liens entre les mots lus et un référent ou évènement se mettent en place sans processus de construction actif. Le sens des mots stockés en mémoire (le lexique) ne donne pas accès directement à la compréhension des mots car une adaptation du sens est nécessaire au regard du contexte (représentation de la situation). Ce processus devient actif lorsque ce qui est activé en mémoire n'est pas pertinent et cohérent avec le modèle de situation présent chez le lecteur. Si la cohérence s'effondre, le lecteur construira de nouvelles structures (Gernsbacher, 1990). La qualité lexicale est importante pour le passage de lecteur novice à lecteur expert car, lorsque la qualité lexicale se construit, les ressources cognitives sont libérées pour la compréhension. Lorsque la qualité lexicale est élevée, les ressources cognitives d'un lecteur peuvent être largement orientées vers la tâche complexe de compréhension car les mots sont reconnus rapidement, automatiquement et avec un effort conscient minimal. En revanche, lorsque la qualité lexicale est faible, certaines des ressources cognitives limitées du lecteur doivent être consacrées à la tâche plus fondamentale de la reconnaissance des mots, ce qui compromet la compréhension. Les données de Verhoeven, Leeuwe, et Vermeer (2011) ont confirmé l'hypothèse de l'importance de la qualité lexicale, car la connaissance des formes et des significations des mots prédit le développement de la compréhension en lecture.

## **2.2 Les principaux facteurs qui influencent la compréhension écrite**

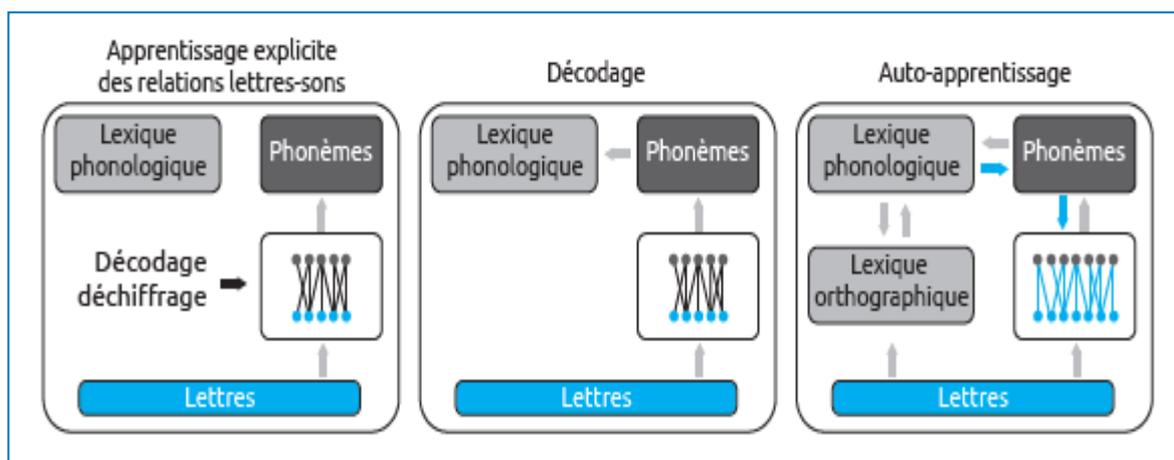
Nous avons vu dans la section précédente que la majorité des modèles convergent sur un nombre de facteurs impliqués dans la compréhension de texte, comme le décodage, le vocabulaire ou les inférences. D'autres facteurs plus spécifiques sont discutés dans la littérature, comme la fluidité, le contexte ou l'acculturation. Dans cette section, nous allons discuter ces facteurs un par un puisqu'ils peuvent tous être impactés par la simplification de textes.

### **2.2.1 Décodage et reconnaissance des mots**

Les systèmes d'écriture sont une invention culturelle qui codent le langage oral. Le code est un ensemble de signes qui permet de transcrire la parole. L'apprentissage de la lecture consiste à maîtriser ce code pour associer l'écrit à la parole grâce à la maîtrise du principe

alphabétique (Sprenger-Charolles & Ziegler, 2019). Lors de l'apprentissage de la lecture, dans un système d'écriture alphabétique, l'enfant doit créer la correspondance entre les unités de base de l'écrit, les graphèmes, et les unités de base de l'oral, les phonèmes. La mise en relation de chaque graphème avec le phonème correspondant demande de bonnes capacités d'analyse phonémique car l'enfant doit être capable de découper les mots oraux en phonèmes (Gentaz & Sprenger-Charolles, 2014). Lorsque les enfants progressent dans l'apprentissage de la lecture, leur forte dépendance au décodage alphabétique diminue progressivement (Wagner & Torgesen, 1987 ; Harm & Seidenberg, 2004). La phase alphabétique est fondamentale car elle permet la mise en route de l'auto-apprentissage (Share, 1995). En effet, lorsque l'élève rencontre un mot nouveau, il peut lui attribuer, par utilisation du décodage, une phonologie plausible (Figure 10). Si celle-ci correspond à une forme orale connue (qui a été enregistrée dans le lexique phonologique), l'enfant pourra mémoriser le mot écrit (stockage dans le lexique orthographique) en association avec la forme orale et le sens (Ziegler et al, 2014). Share insiste sur le fait que, lors du décodage, en traduisant les graphèmes en phonèmes, l'enfant se concentre sur les lettres du mot et leur séquence. L'acte de décodage est donc aussi l'occasion d'acquérir des connaissances orthographiques. En effet, suite à l'enseignement explicite des relations grapho-phonémiques, l'enfant sachant décoder peut lire de façon autonome un texte. Le sens qu'il donne à ce qu'il lit, dans le contexte de la phrase ou du texte, va permettre à l'enfant d'auto-corriger ses éventuelles erreurs. Cette phase d'auto-apprentissage renforce l'automatisation de la lecture.

Figure 10 : Modèle d'apprentissage de la lecture basé sur le décodage et l'auto-apprentissage des représentations orthographiques (Sprenger-Charolles & Ziegler, 2019, p. 101).



Une fois l'enseignement du code alphabétique réalisé, l'enfant est en mesure d'appliquer le décodage appris pour tous les mots, même ceux qu'il n'aura jamais rencontrés, mais dont il

identifie la forme au sein de son lexique phonologique et le sens grâce à l'acquisition du langage oral. Chaque fois qu'un mot est correctement décodé, c'est-à-dire que la forme phonologique correspondante est activée, le mécanisme de décodage est renforcé (Figure 10). Selon le processus d'auto-apprentissage, chaque rencontre d'un mot inconnu avec un décodage réussi permet d'acquérir une orthographe spécifique du mot : c'est le fondement de la reconnaissance orthographique des mots (Fayol & Jaffré, 1999 ; Écalte & Magnan, 2002). Si le lecteur trouve un mot décodé dans son lexique phonologique il va renforcer les connexions entre les graphèmes et les phonèmes à l'origine du décodage. Le mécanisme de décodage se perfectionne grâce à l'auto-apprentissage. Grâce à ce mécanisme d'auto-apprentissage, l'enfant au bout de nombreux essais réussis peut lire correctement la quasi-totalité des mots (Sprenger-Charolles & Ziegler, 2019).

### 2.2.2 Fluidité

La fluidité, c'est-à-dire le taux de mots lus correctement et la prosodie de la lecture d'un texte, a longtemps été considérée comme l'une des nombreuses composantes d'une lecture efficace (National Reading Panel Reports of the Subgroups, 2000). Pour définir la fluidité nous nous sommes référés à Piluiski et Chard (2005) :

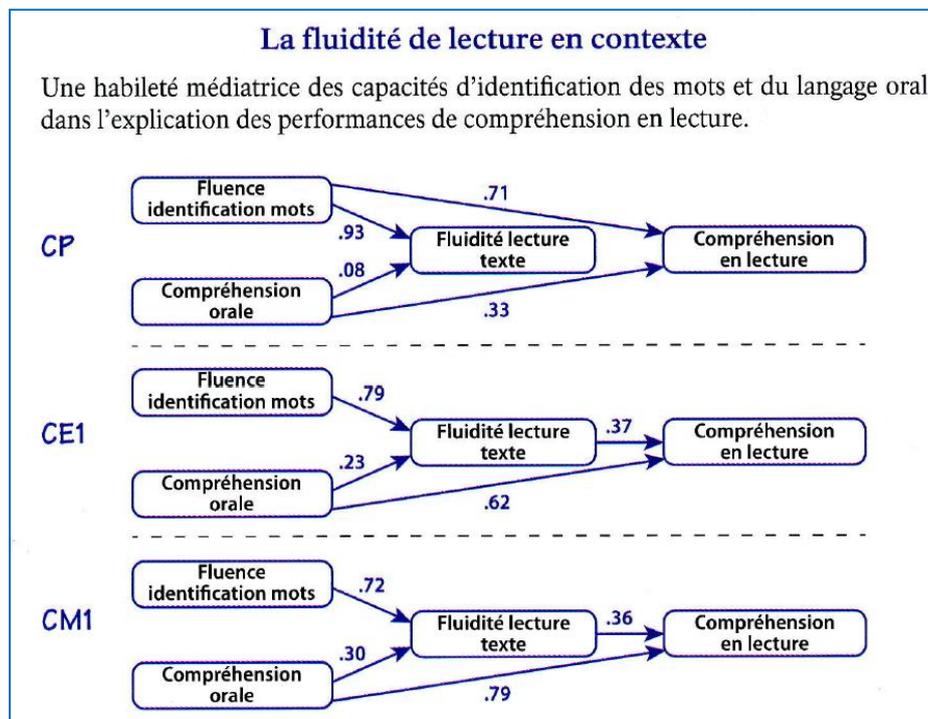
« La fluidité de la lecture désigne la rapidité, l'efficacité, la précision, les compétences de reconnaissance des mots qui permettent au lecteur de construire le sens d'un texte. La fluidité se manifeste également par une lecture orale précise, rapide et expressive ; elle est appliquée pendant la lecture silencieuse rendant possible la compréhension. » [Traduction libre] (Piluiski & Chard, 2005, p. 511)

Adlof, Catts et Little (2006) ont mené une étude auprès de 604 enfants de la deuxième à la huitième année (CE1, CM1 et 4<sup>ème</sup>) dont l'objectif était de savoir si une composante de fluidité distincte devait être ajoutée au modèle SVR. Les auteurs évaluent la fluidité pour les trois niveaux scolaires en se basant sur la maîtrise de la lecture (précision et temps). La modélisation d'équations structurelles a été utilisée pour déterminer si la fluidité expliquait une part de variance unique de la compréhension en lecture après avoir contrôlé la précision de la reconnaissance des mots et la compréhension orale. Les résultats ont montré que la fluidité n'était pas une variable unique dans la compréhension de la lecture mais que la variance de la compréhension dépendait de la reconnaissance des mots et de la compréhension orale. En CE1, la reconnaissance des mots représente 14,3% de la compréhension en lecture contre 2,9% en

CM1. De plus la compréhension orale explique 19,4% de la compréhension en lecture en CE1 contre 45,2% en CM1. Peu d'individus avaient des problèmes de fluidité indépendamment de la précision de la reconnaissance des mots ou de la compréhension orale.

Dans une étude longitudinale réalisée par Kim et Wagner (2015), auprès de 316 enfants anglophones entre le CP et le CM1, l'évolution de la fluidité (précision et rapidité en lecture de textes à l'oral) et les liens entretenus avec les autres facteurs de la compréhension en lecture ont été analysés. La fluence est un facteur qui est corrélé aux compétences de lecture de textes au-delà de la seule compétence de la lecture de mots isolés (Figure 11).

Figure 11 : « La fluidité de lecture orale est liée, mais distincte, de la fluence de décodage et de lecture de mots isolés. » (Bianco & Lima, 2017, p. 27).



D'après l'étude de Kim et Wagner (2015), au départ, chez les jeunes élèves, la fluidité de la lecture de textes était fortement liée à la fluence de lecture de mots isolés ( $r = 0,96$  en 1<sup>ère</sup> année), mais au cours du développement, la lecture de textes paraît davantage influencée par la compréhension orale et de plus en plus dissociée de la fluence de lecture des mots. Cette constatation est conforme au modèle SVR.

### 2.2.3 Vocabulaire

Le niveau de vocabulaire est un facteur prédictif de la qualité de compréhension en lecture (Ouellette & Beers, 2010). Dans l'étude de Ouellette et Shaw (2014) auprès de 96 élèves

en 6<sup>ème</sup> année (collège), le but était d'analyser la nature des liens entre le vocabulaire oral et la compréhension écrite. L'étendue du vocabulaire explique une part significative mais pas unique de variance en compréhension de lecture, au-delà de la reconnaissance de mots et des compétences phonologiques. Le vocabulaire comprend à la fois l'étendue du lexique oral, c'est-à-dire le nombre d'entrées lexicales intégrées par le sujet, mais aussi sa profondeur, correspondant au degré de connaissances sémantiques, autrement dit la qualité de l'organisation lexico-sémantique : définitions, associations, catégorisation. Une connaissance approfondie du vocabulaire englobe non seulement le nombre de mots connus individuellement, mais aussi leur capacité à être reconnus et leur utilisation dans un contexte donné, ce qui est essentiel étant donné que la majorité des mots sont polysémiques, c'est-à-dire qu'ils ont plusieurs significations ou « sens » dans une plus ou moins large mesure (Rodd, 2020). Au-delà de la simple connaissance des mots, la compréhension du texte exige la connaissance d'expressions idiomatiques qui se présentent fréquemment dans les textes, par exemple le sens de « se regarder le nombril ; avoir la tête dans les nuages ; tomber dans les pommes » ne peut pas être déduit du sens des mots pris individuellement. Face à un mot inconnu, deux processus d'acquisition peuvent être déployés : une analyse externe du mot en ayant recours à une analyse du contexte, et une analyse interne du mot, c'est-à-dire une analyse des composantes morphémiques du mot (Hattouti, Gil & Laval, 2016).

#### **2.2.4 Morphosyntaxe**

Comme le vocabulaire, la connaissance de la morphologie (la forme des mots) et de la syntaxe (combinaisons et relations entre les mots dans les phrases) fait partie du répertoire de la langue parlée de l'enfant. La morphosyntaxe porte à la fois sur les formes des mots et sur l'organisation des mots et des groupes de mots dans un énoncé ou une phrase.

Le lecteur effectue un travail de compréhension morphosyntaxique pour extraire le sens de la phrase lue (Hoover & Gough, 1990 ; Gombert, Martinot & Nocus, 1996 ; Simard, Foucambert & Labelle, 2014). Un enfant doit non seulement connaître le sens des mots mais aussi savoir comment les mots d'une phrase fonctionnent ensemble, afin de construire le sens. Le principe syntaxique est tout aussi important que le principe alphabétique, il le complète et fait partie intégrante de l'apprentissage de la lecture (Bentolila, 2019). En effet, plusieurs études ont montré que la compréhension syntaxique et la conscience de la morphologie dans le langage parlé sont associées à la compréhension en lecture (Lervåg, Hulme & Melby-Lervåg, 2018 ; Muter, Hulme, Snowling & Stevenson, 2004). Les enfants ayant de moins bons résultats que leurs pairs sur des mesures de conscience morphologique ont tendance à avoir une mauvaise

compréhension de la lecture. Colé, Royer, Leuwers et Casalis (2004) ont présenté des données issues d'une recherche longitudinale conduite du CP au CE1 en incluant un groupe contrôle de lecteurs de CE2, qui montrent que dès le CP, le niveau de lecture atteint par les apprentis-lecteurs est associé au niveau de développement des connaissances morphologiques dérivationnelles implicites<sup>11</sup>. Colé et ses collaborateurs (2004) ont précisé qu'à partir du CE1 et au CE2, on observe une influence des connaissances morphologiques explicites de ces lecteurs sur leur niveau de lecture, répliquant pour partie des résultats déjà observés. Colé, Marec-Breton, Royer et Gombert (2003) ont suggéré que, dans un premier temps, les moins bons lecteurs utiliseraient implicitement la structure morphologique des mots pour les lire en profitant du partage du sens entre les mots d'une même famille morphologique (morphologie dérivationnelle). Les connaissances morphologiques explicites sur les mots ne seraient exploitées que progressivement, les lecteurs considérant alors conjointement la proximité formelle et la parenté sémantique des mots de la même famille morphologique.

### 2.2.5 Les inférences dans la compréhension

Van den Broek (1997) a expliqué le rôle critique de l'inférence, qui chez le lecteur met en lien une partie du texte avec d'autres parties du texte et des connaissances du monde. Les différents cadres théoriques mettent l'accent sur plusieurs aspects de la compréhension en lecture et partagent la notion centrale d'une compréhension en lecture qui appelle la construction d'une représentation mentale cohérente du texte dans la mémoire du lecteur. Les mécanismes de production d'inférences sont utilisés par le lecteur pour apporter de l'information non explicite dans le texte afin d'en obtenir une représentation (Martins & Le Bouédec, 1998). Cette représentation rapproche des informations textuelles et des connaissances du monde, reliées au travers de relations sémantiques (par exemple, des relations causales, référentielles et spatiales).

Le traitement de l'inférence est expliqué comme un traitement automatique par McKoon et Ratcliff (1992) qui ont proposé une vision qualifiée *d'hypothèse minimaliste*. Selon cette hypothèse, les seules inférences qui sont encodées automatiquement lors de la lecture sont celles issues d'informations facilement disponibles en mémoire, tirées d'énoncés explicites dans le texte ou de connaissances générales, nécessaires pour établir la cohérence locale du texte (par exemple, références pronominales, relations causales). Les inférences minimalistes créées

---

<sup>11</sup> Même si la distinction implicite/explicite est parfois délicate à effectuer théoriquement, elle se révèle nécessaire au plan empirique. En effet, les études corrélationnelles ont montré que tâches nécessitant une manipulation explicite des phonèmes sont très directement prédictives de la maîtrise du principe alphabétique, ce qui n'est pas le cas des tâches impliquant l'utilisation plus implicite des connaissances phonologiques (Muter et al., 1997 ; Duncan et al., 2000)

automatiquement fournissent la représentation de base du texte. Les auteurs ont mis en évidence que durant la compréhension du texte, le lecteur avait accès différemment aux informations locales et globales. Ils précisent que les inférences relatives à la cohérence globale ne sont pas générées au cours de la lecture. Pour Graesser, Singer, et Trabasso (1994), ce type d'inférences ne permet pas d'intégrer l'ensemble des informations du texte en une représentation cohérente. Les auteurs mettent en avant des lacunes de l'hypothèse minimaliste. Pour eux, le traitement de l'inférence est expliqué comme une construction du lecteur qui se crée une représentation cohérente du texte faisant des liens entre les informations du texte et ses connaissances du monde. Ils montrent une activité délibérée du lecteur, activité d'interprétation qui repose sur un travail d'élaboration d'hypothèses, d'explications de liens entre les éléments pour générer la cohérence globale. Ils proposent une théorie constructionniste qui tient compte des inférences fondées sur les connaissances qui sont construites lorsque les lecteurs comprennent le texte. Les processus stratégiques et de recherche en mémoire sont contrôlés par le lecteur et demandent des ressources cognitives supplémentaires. À partir des éléments du texte, une recherche permet de récupérer des éléments en mémoire à long terme, les informations dont le lecteur a besoin pour construire la cohérence globale du texte.

Pour construire une représentation mentale, trois niveaux de traitement de l'information peuvent être considérés : le mot, la phrase et le texte. Ainsi, dans cet extrait d'un texte utilisé dans notre expérimentation de simplification de textes, début du livre *Crin-Blanc* (Guillot, 2008) : « *L'eau a encore baissé, pensa Folco. Le jeune garçon sentit sous ses pieds sa vieille barque vermoulue racler le fond de vase.* ». La représentation mentale qui se crée en associant deux informations de la phrase au plan local, la vase raclée et la sensation ressentie par le jeune garçon, se complète d'une inférence fondée sur les connaissances. Les connaissances générales sur le niveau de l'eau permettant de circuler avec une barque vont renforcer la représentation mentale construite en prenant en compte les éléments du texte. Le lien causal entre les deux phrases, la baisse de l'eau qui a pour conséquence le raclement de la barque sur la vase crée la cohérence causale (Voir chapitre 2.1.2).

De plus, les relations sémantiques sont reconnues par le lecteur grâce à des processus inférentiels passifs et stratégiques (Kintsch, 1988). Les processus d'inférence passifs, se créent automatiquement à partir de la représentation linguistique du texte alors que les processus stratégiques exigent des ressources de mémoire attentionnelle et de travail de la part des lecteurs car ces inférences sont liées aux connaissances du lecteur et au contexte. Pour maintenir la cohérence au travers des processus inférentiels, le lecteur s'appuie donc sur ses connaissances préalables lorsque des éléments du texte semblent s'opposer. Il se base sur des informations

inscrites dans le texte pour relier les informations explicites et implicites et maintenir une cohérence entre tous les éléments présents : connaissances, informations explicites, éléments implicites du texte. Le recours à ces informations évoquées implicitement dans le texte définit l'inférence. Lors de la production d'inférences, des processus cognitifs de haut niveau se mobilisent autour de la capacité d'organiser et de réfléchir sur l'information dans les limites de la capacité de mémoire de travail du lecteur (Cain, Oakhill, Barnes & Bryant, 2001).

### 2.2.6 Le contexte dans la lecture

Le contexte est un facteur qui influence la compréhension et qui est présent dans les cadres théoriques décrits ci-dessus. Le contexte ne peut pas toujours aider un enfant qui ne décode pas correctement. Ainsi dans les phrases : « Le professeur est venu **lire** des poèmes aux enfants » / « Le professeur est venu **dire** des poèmes aux enfants », le rôle du décodage est central. Si l'enfant décode, grâce à l'auto-apprentissage, il peut enrichir son vocabulaire avec le contexte de sa lecture. Le décodage devient de plus en plus "lexicalisé" au cours du développement de la lecture dans le processus d'auto-apprentissage (Share, 1995). Les correspondances régulières lettre-son évoluent avec les contraintes lexicales imposées par le nombre croissant de connaissances orthographiques. Le lexique écrit se développe et l'enfant peut réaliser des correspondances au-delà du niveau simple graphème-phonème en tenant compte de contraintes contextuelles, positionnelles et morphémiques. Le résultat de ce processus de « lexicalisation » est une lecture dont la reconnaissance des relations entre l'écrit et le son évolue à un degré qui le rend indiscernable d'un mécanisme lexical qui ne maintient aucune correspondance orthographe-son sous les règles lexicales pour exemple : monsieur, aquarium. La connaissance des correspondances entre l'orthographe et la phonologie devient de plus en plus sensible au contexte, répondant au processus de « lexicalisation », (Share, 1995).

Le contexte n'est pas suffisant mais nécessaire, il aide en cas d'ambiguïté sur le décodage. Si la connaissance des lettres et de leur séquence est nécessaire pour différencier les mots, rien / nier / rein par exemple, des connaissances linguistiques et des connaissances du monde sont aussi nécessaires pour retrouver dans son lexique mental le mot et le référencer à une situation. La prise en charge des informations contextuelles complète le premier niveau du décodage chez des lecteurs plus faibles en leur apportant des indices supplémentaires pour comprendre. L'étroite interaction qui existe entre la reconnaissance des mots écrits et le contexte a été nommée « modèle interactif compensatoire » (Stanovich, 1980). La capacité des jeunes enfants (7-8 ans) à s'appuyer sur le contexte, à l'écrit, pour déduire la signification de nouveaux mots de vocabulaire dépend de leur niveau de compétences linguistiques mais aussi

de la proximité du nouveau mot avec le contexte permettant de le définir. La performance de la déduction de la signification des nouveaux mots est meilleure quand le nouveau mot est proche de son contexte de définition dans le texte (Cain, Oakhill & Elbro, 2003).

Le contexte est à associer à la qualité lexicale (Perfetti, 2007), pour exemple, le mot 'manche' peut être associé à : une manche de veste / une manche à air / un manche de pelle / traverser la Manche / faire la manche / se débrouiller comme un manche (familier) / jeter le manche après la cognée (expression datant du 14<sup>ème</sup> -15<sup>ème</sup> siècle, registre soutenu). Le contexte est aussi en relation avec les connaissances du monde du lecteur.

### 2.2.7 Acculturation à l'écrit

Les stratégies potentiellement plus efficaces pour augmenter les apprentissages des élèves en lecture, suggérées dans la recherche dirigée par Goigoux (2016)<sup>12</sup> dans l'étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages, se répartissent en deux grandes catégories : maximiser la valeur de la lecture et faciliter le choix des supports. La variable « acculturation à l'écrit » a été construite à partir de trois variables : les écrits (exposition à l'écrit, production de textes) ; les albums (nombre d'albums et autres ouvrages) et les espaces et les usages (coin-lecture, fréquentation d'une bibliothèque, projet culturel et abonnement à une revue). Selon le nombre considéré de ces variables (1, 2 ou 3), les classes « peu acculturantes » représentent 34% et « moyennement acculturantes » 37%. Les classes « très acculturantes » (29%) favorisent « la progression des élèves de niveau faible et intermédiaire dans leurs différentes performances de lecture-écriture. » (Goigoux, 2016, p. 368)

De même, au moyen d'une série de questionnaires, Ahr et Butlen (2012), ont analysé les relations qui pouvaient être établies entre les performances scolaires d'adolescents de 15 ans et des critères comme le nombre de livres au domicile ou le goût pour la lecture. Il s'avère que les performances sont corrélées à la présence de livres au domicile. En effet, un quart des élèves ayant obtenu les plus faibles scores déclarent avoir moins de 10 livres chez eux. Or ce cas de figure ne se retrouve jamais chez les élèves ayant obtenu les meilleurs résultats. De plus, il apparaît que la majorité des adolescents les moins performants ne fréquente que rarement les lieux publics liés à la lecture comme la bibliothèque ou le Centre de Documentation et d'Information. Mais surtout, une absence de plaisir à la lecture est décrite par 45% des élèves les moins performants, alors que les plus performants lisent, par plaisir, plus de 30 minutes par

---

<sup>12</sup> <http://ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire>

jour. Au regard de cette étude, on comprend alors que l'exposition régulière à l'écrit est liée à une meilleure compréhension écrite, même si le sujet est en difficultés.

« Les élèves qui ont un large accès aux livres chez eux ont tendance à s'intéresser davantage à la lecture et à diversifier leurs pratiques, leur manière de lire est plus performante. Ceux qui ont un accès aux livres plus limité tendent à moins s'investir dans les pratiques de lecture et ont des résultats plus faibles. » (Ahr & Butlen, 2012, p. 222)

Dans leur analyse, Ahr et Butlen (2012) ont recueilli plusieurs témoignages d'adolescents en difficultés vis-à-vis de la lecture. Pour eux, ne pas comprendre le récit provoque un réel déplaisir, déplaisir souvent attribué par extension à la lecture en elle-même.

Nous retrouvons les mêmes éléments dans l'analyse réalisée par Bart et Daunay, (2015) avec les données de PISA 2009. Le désir de lire est intégralement lié à la capacité de lecture elle-même : les enfants sont plus motivés à lire et s'y engagent davantage lorsqu'ils sont bons dans ce domaine (Mol & Bus, 2011). Les auteurs ont réalisé une synthèse de recherche, méta-analyse de 99 études (N= 7669) qui portaient sur la lecture pendant les loisirs auprès d'enfants en maternelle jusqu'au collège et à l'université. Les résultats montrent que les enfants qui sont plus compétents en compréhension, en lecture et en orthographe lisent davantage parce qu'ils accèdent à une plus grande exposition à l'écrit et que leurs compétences se sont améliorées chaque année de leur scolarité. Par exemple, à la maternelle, l'exposition à l'écrit expliquait 12% de la variance des compétences linguistiques orales, à l'école primaire 13%, au collège 19%, au lycée 30% et à l'université 34%.

Anderson, Wilson et Fielding (1988) ont étudié les habitudes de lecture extrascolaire, à la fois de livres et d'autres types de textes, d'un groupe d'enfants américains de 5<sup>ème</sup> année (âgés de 10 et 11 ans). Sur la base déclarée du temps passé par les enfants à lire quotidiennement, Anderson et ses collaborateurs ont estimé le nombre de mots auxquels les enfants auraient été exposés pendant plus d'un an. Les auteurs ont estimé que ceux du 10<sup>ème</sup> centile de temps passé à lire sont exposés à environ 60 000 mots par an ; ceux du 50<sup>ème</sup> centile, 900 000 mots ; et ceux du 90<sup>ème</sup> centile, plus de 4 millions de mots. Ces différences d'exposition ont des effets cumulatifs sur la capacité de lecture au fil du temps, « les riches s'enrichissant et les pauvres s'appauvrissant », ce que l'on appelle l'effet Matthew (Stanovich, 1986).

### 2.3 Les difficultés lors de la compréhension

Nous avons vu dans la section précédente que deux composantes indépendantes expliquent une grande partie de la compréhension de textes, à savoir la reconnaissance de mots et le langage oral (voir Figure 2, sous-section 2.1.1). En effet, si la reconnaissance de mots écrits est déficitaire, la compréhension de l'écrit sera déficitaire. C'est le cas des élèves dyslexiques, par exemple, pour qui la reconnaissance de mots écrits ne s'automatise pas malgré une compréhension orale normale (Snowling, 2000 ; Demonet, 2004 ; Valdois, 2010). Chez ces enfants, la lecture reste souvent lente, laborieuse et sérielle (Ziegler, Perry, Ma-Wyatt, Ladner & Schulte-Körne, 2003). À l'inverse, si la compréhension orale est déficitaire malgré une reconnaissance de mots écrits efficace, la compréhension de texte sera également déficitaire. C'est le cas des enfants chez qui le langage oral est faible. Ces enfants parviennent éventuellement à décoder correctement mais ils ne comprennent pas ce qu'ils lisent.

Nous avons également vu dans la section précédente que chaque composante est influencée par différents facteurs (voir Figure 2, sous-section 2.1.1). En effet, de nombreuses études ont montré que des déficits au niveau de l'un ou de plusieurs de ces facteurs provoquent des déficits de fluidité et de compréhension de textes. Par exemple, les déficits visuels ou visuo-attentionnels empêchent la mise en place d'un traitement rapide et parallèle des mots écrits (Bosse, Tainturier & Valdois, 2007). Des déficits phonologiques, notamment une conscience phonémique déficitaire, sont omniprésents chez les enfants dyslexiques et les faibles lecteurs (Ziegler & Goswami, 2005 ; Goswami, 2011 ; Casini, Pech-Georgel & Ziegler, 2017). Ces déficits empêchent le décodage qui, comme nous l'avons vu précédemment, constitue une étape initiale fondamentale pour l'apprentissage de la lecture, notamment car il permet l'auto-apprentissage (Ziegler, Perry & Zorzi, 2014 ; Ziegler, Perry & Zorzi, 2020).

Des déficits dans la compréhension entraînent des difficultés pour les enfants et, notamment, pour trouver la signification d'un mot nouveau au regard du contexte ; par conséquent ils sont fragilisés dans la construction d'inférences. Le vocabulaire n'est pas le seul élément qui permet d'influencer les capacités de compréhension car les systèmes de mémoire utilisés pour le stockage et le traitement simultané des informations contribuent à la compréhension. Des déficits dans le vocabulaire entraînent des difficultés pour le lecteur car le vocabulaire soutient l'inférence de deux manières : par la connaissance des significations des mots nécessaires pour générer des inférences et dans le processus de mémoire. Ces enfants, qui ne peuvent pas générer une série d'inférences, ne font donc pas les déductions nécessaires pour construire une représentation cohérente de la signification d'un texte (Cain, Oakhill & Lemmon,

2004 ; Currie & Cain, 2015). Des déficits dans les fonctions exécutives peuvent exister chez les faibles lecteurs, indépendamment de leurs déficits de traitement phonologique. Or, nous avons vu dans la présentation des modèles de compréhension que la mémoire de travail joue un rôle essentiel dans l'intégration des informations pendant la compréhension de la lecture (Swanson, Howard & Sáez, 2007).

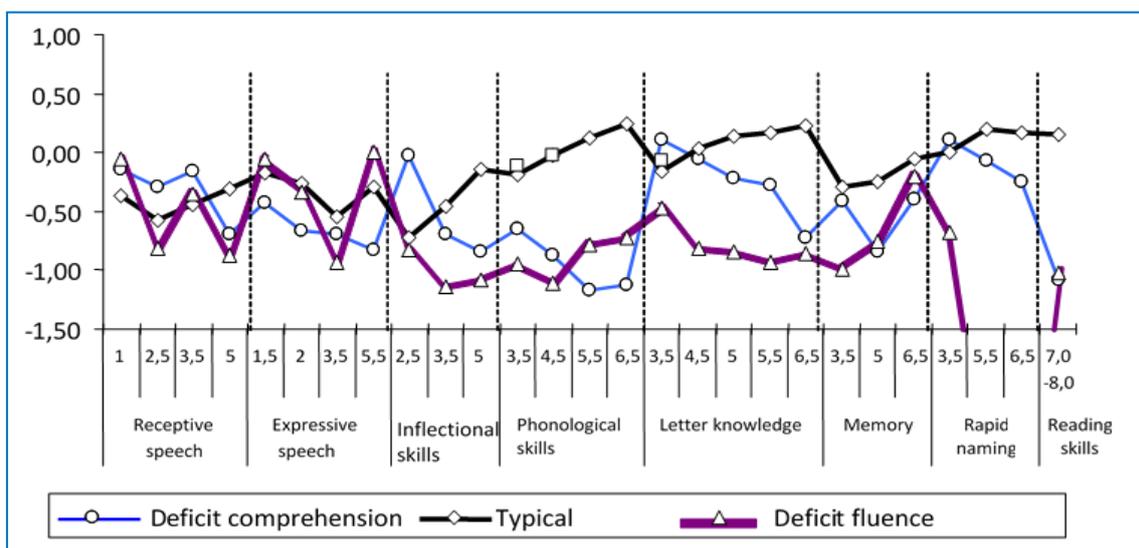
La compréhension orale a été étudiée dans une étude longitudinale, du CE1 au CM1, grâce à des tests individuels en compréhension orale, compétence complexe impliquant de nombreux processus cognitifs et linguistiques. La contribution de la compréhension orale à la compréhension de la lecture augmente au fil des années et, par conséquent, les malentendus dans la compréhension écrite augmentent également pour les mauvais 'compreneurs' en langue orale. Les influences du langage oral sur la compréhension écrite concernent le vocabulaire, les connaissances de base et l'inférence, en particulier (Hogan, Adlof & Alonzo, 2014).

En cas de déficit de la parole, d'un retard de langage, d'un trouble articulo-phonatoire ou bien d'une dysphasie, la compréhension peut être altérée au-delà de la reconnaissance des mots écrits. Des difficultés sur le traitement des inférences, de la cohérence locale (relation entre les phrases), des difficultés d'utilisation du contexte pour déduire le sens d'un mot, et un autocontrôle peu efficace de la compréhension ont été relevés (Macchi, Casalis & Schelstraete, 2016). Les enfants atteints d'un trouble du développement du langage grave ne présentent pas simplement un retard homogène dans leur capacité à reconnaître les mots écrits (Zourou, Écalle, Magnan & Sanchez, 2010) mais semblent avoir plus souvent recours à l'orthographe lors de la lecture plutôt qu'à la voie phonologique (Macchi et al., 2019). Les enfants ayant des niveaux faibles de connaissances générales ou du monde, antérieures à la lecture d'un texte sont désavantagés lors de la lecture car ces connaissances peuvent notamment les aider à faire des déductions pour comprendre (McNamara, Floyd, Best & Louwerse, 2003 ; Best, Ozuru, Floyd & McNamara, 2006).

Une étude longitudinale de Lyytinen, Erskine, Tolvanen, Torppa, Poikkeus & Lyytinen, (2006) a suivi 200 enfants finlandais de 12 mois à 8 ans. L'objectif était de comprendre pourquoi certains enfants ont des difficultés dans l'acquisition de compétences en lecture. Pour moitié de l'effectif, les enfants étaient considérés comme étant à risque de dyslexie au regard de leurs antécédents familiaux, pour l'autre moitié de l'effectif, groupe témoin, les enfants étaient considérés sans risque. Des groupes sont ressortis après les analyses de leurs différences de développement, par le biais d'une modélisation mixte avec sept domaines de compétences (Figure 12) : compétences langagières en réception et en expression, compétences morphologiques flexionnelles, capacité de mémoire, connaissances des lettres, conscience

phonologique, vitesse de dénomination d'éléments en série (lettres, couleurs, ...). À l'âge de sept ans la compétence en lecture a été évaluée en précision et en vitesse de lecture de mots et non-mots. L'année suivante, la fluidité a été évaluée dans des phrases et des textes. Les enfants ayant un développement classique (groupe type), dont 55% issus du groupe témoin, maîtrisaient tous les domaines liés à la lecture, tels que la phonologie et la connaissance des lettres. Ils ont montré une progression continue de leurs performances tout au long de leur développement. Ces enfants étaient ceux qui avaient le plus de compétences acquises avant l'entrée à l'école. Les enfants ayant un déficit en fluidité (précision et rapidité de lecture) dont la majorité faisait partie du groupe à risques avaient de grandes difficultés en vitesse de dénomination. Leur score dans cette compétence est inférieur à moins deux d'écart-type, signifiant une grande difficulté (tracé non apparent sur la figure présentée). Cependant, comme le montre la figure 12, le score composite général de lecture pour ce sous-groupe est faible. Cette faiblesse en vitesse de lecture a persisté sur les deux premières années scolaires. L'enseignement de la lecture a permis la précision de lecture mais pas la maîtrise (fluidité) pour ces enfants. Ils ont obtenu une meilleure performance relative en mémoire verbale à court terme. Ces enfants étaient en difficultés dans tous les domaines à l'exception de l'expression orale. Les enfants ayant un déficit de compréhension en lecture (n= 35 ; 24 à risque vs 11 du groupe témoin) partageaient une caractéristique différente des autres groupes. Leurs performances diminuaient continuellement dans la plupart des domaines de compétences sauf en vitesse de dénomination. Ce déficit est plus fréquent lorsqu'existe un risque familial de dyslexie. Environ 20% des enfants de ce groupe montraient des problèmes en lecture ou en orthographe. À la fin de la première année scolaire et plus tard encore ils ont rencontré des problèmes dans le développement de la fluidité de lecture.

Figure 12 : Moyenne des valeurs en score Z. Performances dans les sept domaines entre 1 an et 6,5 ans et dans les compétences en langue écrite pour les deux premières années de l'école d'après Lytinen et al. (2006).



Dans cette étude, nous pouvons constater que les difficultés en longitudinal sur les compétences langagières en réception et en expression, la morphosyntaxe, la phonologie et la connaissance de lettres prédisent l'apprentissage et la compréhension en lecture. Lorsque les élèves ont des difficultés dans ces composantes de la lecture à l'âge de 3, 4, 5 ans déjà, les difficultés lors de l'apprentissage apparaissent immédiatement.

Dans une étude francophone, Gentaz, Sprenger-Charolles et Theurel (2015) ont testé l'importance de quelques-uns des facteurs prédicteurs de la compréhension en lecture dans un milieu à faible statut socio-économique. L'effectif de 392 élèves de CP a été divisé en 3 groupes d'après leur niveau de décodage. Les résultats ont montré que la compréhension orale intervient davantage chez les bons et moyens décodeurs que chez les décodeurs en difficultés. Le pourcentage d'enfants avec de faibles capacités de compréhension de l'écrit est plus élevé chez les faibles décodeurs (55%) que chez les moyens (7%) et les bons décodeurs (0%). De faibles capacités de décodage sont fortement corrélées à un faible niveau de compréhension. Gentaz et ses collaborateurs (2015) ont montré que 58% de la variance dans la compréhension de la lecture ont été expliqués par les quatre principaux prédicteurs suivants : les compétences de décodage se révélant être le meilleur prédicteur, la compréhension orale étant le deuxième, le vocabulaire étant le troisième prédicteur et la conscience phonémique le dernier. Le QI non verbal et la conscience syllabique n'ajoutent aucune part significative à la variance de la compréhension de la lecture. Seulement 1,5% des enfants n'ayant aucun déficit en décodage et/ou en compréhension orale ont obtenu des scores faibles en compréhension de l'écrit.

Enfin, quelques études chez les lecteurs adultes se sont intéressées aux facteurs qui expliquent une mauvaise compréhension, laquelle ne peut pas être expliquée par un mauvais décodage. Gernsbacher (1990) a évalué auprès de 270 étudiants du premier cycle universitaire quelles étaient les compétences nécessaires pour construire une représentation ou « structure mentale cohérente » du texte. Elle a analysé les processus de compréhension de faibles lecteurs. Si l'on s'en tient à la compréhension d'une phrase, deux moments principaux existent : l'activation de significations préexistantes, notamment lexicales, et leur assemblage, contrôlé par des règles de compétences syntaxiques et sémantiques. Au niveau des différences individuelles dans la construction de structures, les faibles 'compreneurs' sont moins aptes à supprimer les informations non pertinentes, ils ont un accès plus limité à des informations récemment comprises car leurs capacités de mémoire sont moins importantes. Leurs compétences de compréhension sont donc moins efficaces. De plus, ces résultats ont été renforcés grâce à plusieurs expériences montrant que les faibles 'compreneurs' avaient plus de difficultés à supprimer les significations inappropriées des mots ambigus et des homophones.

Les faibles 'compreneurs' sont capables de rejeter les informations contextuellement inappropriées mais principalement lorsque le contexte est prévisible (Gernsbacher et al. 1991).

## 2.4 La compréhension en fonction de la nature du texte

La nature du texte influence la façon dont le lecteur va intégrer les informations directement explicites du texte et activer ses connaissances antérieures pour l'appréhender (Wolfe, 2005). Chaque type de texte, selon qu'il soit plus narratif ou plus descriptif, peut déclencher une stratégie plutôt qu'une autre. Les expériences menées par Wolfe (2005) ont montré que l'organisation des éléments du texte et les associations sémantiques liées à ces éléments influencent la mémorisation. Les éléments du texte sont mieux mémorisés (et donc mieux compris) quand ils sont centraux dans l'organisation du texte et quand ils sont sémantiquement liés au contexte de la lecture et aux sujets étudiés (Wolfe, 2005). Ainsi, les textes documentaires informatifs, seraient plus difficiles à comprendre et à mémoriser que les textes narratifs, diverses hypothèses sont susceptibles d'expliquer cette différence. Ce qui distingue le texte narratif et le texte informatif est la complexité conceptuelle et les relations entre les mots dans un texte. Hiebert et Cervetti (2012) ont rapporté que les mots scientifiques sont souvent monosémiques, plus longs, plus complexes sur le plan conceptuel et plus liés thématiquement entre eux et au sujet du texte que les mots présents dans des textes narratifs.

« Les textes scientifiques sont souvent caractérisés par une plus grande concision et des termes monosémiques. Le lexique est spécialisé et difficile à mémoriser. Le contexte ne suffit pas à éclairer le sens des mots. Les indices contextuels n'ont aucun effet sur la construction de la cohérence de l'information et il est subordonné au contrôle d'un vocabulaire, d'un langage spécialisé et de connaissances précises sur le terrain évoquées par le texte. » (Marin, Crinon, Legros & Avel, 2007, p 120)

Duke (2000) a souligné la « rareté » des textes informatifs utilisés dans les classes élémentaires. En moyenne, en classe, les activités d'étude de la langue écrite sur les textes informatifs occupent un temps scolaire de 3,6 minutes par jour. La place du texte informatif est très rare dans la classe, que ce soit sur les murs, dans la bibliothèque, dans les activités. Jeong, Gaffney et Choi, (2010), s'intéressant à l'évolution de l'utilisation du texte informatif en classe, ont comparé la proportion de textes informatifs présents en classe par rapport aux autres textes, notamment narratifs. Dans leur étude, la proportion de textes informatifs n'avait pas augmenté

de manière significative en CE1, CE2, CM1, notamment en bibliothèque et dans l'affichage en classe. Baker et ses collaborateurs (2017) ont montré les avantages d'accéder aux textes informatifs, avec l'augmentation du nombre de livres informatifs dans les classes et l'enseignement sur leur utilisation. L'objectif de leur étude était de déterminer si une intervention en classe pourrait engager et motiver de jeunes élèves, entre le CE1 et le CM1, sur la lecture de textes informatifs. Les auteurs ont constaté un effet sur les préférences de lecture des élèves, qui ont choisi les textes informatifs parmi les deux livres favoris de l'année. Ils ont expliqué ce choix par un désir d'informations et ont déclaré lire plus de livres en autonomie à l'extérieur de l'école. Non seulement les textes informatifs motivent les lecteurs qui construisent et élargissent leurs connaissances préalables (Guthrie, Hoa, Wigfield, Tonks, Humenick & Littles, 2007) mais ces textes exposent les enfants à des structures et à des caractéristiques organisationnelles qu'ils ne rencontreraient pas dans les textes narratifs (Pappas, 1993).

Wright, Reeves, Williams, Morrison-Love, Patrick, Ginestié, Mammes et Graube (2018) ont montré que l'apprentissage des STEM (Science, Technology, Engineering, and Mathematics) peut être satisfaisant pour les élèves car cet enseignement est pour eux l'occasion de développer leurs capacités conceptuelles et leur réflexion. Dans cet article, Wright et ses collègues (2018) ont comparé le développement de ces disciplines (STEM) dans cinq pays (Allemagne, Australie, Ecosse, Etats-Unis, France) et ont signalé que ces disciplines (STEM) apportaient des connaissances et compétences importantes dans la vie des jeunes. Wright et ses collaborateurs (2018) ont suggéré d'enseigner les principes de la technologie et de l'ingénierie comme des cours de base qui aideront à contextualiser d'autres matières.

Pour autant, si l'accès et l'exposition aux textes informatifs progressent, la compréhension de textes narratifs reste plus élevée que celle de textes informatifs (Best, Floyd & McNamara, 2008 ; McNamara, Ozuru & Floyd, 2017). Selon les textes officiels du MEN<sup>13</sup>, les textes informatifs sont inclus dans le programme scolaire parce qu'ils contribuent à l'acquisition de connaissances, aident à la réussite ultérieure en lecture et motivent les lecteurs. Pour renforcer l'exposition des élèves à ce type d'écrits, les sollicitations auprès des professeurs sont régulières, par exemple, que ce soit l'invitation annuelle à participer à la Fête de la Science<sup>14</sup> ou l'accompagnement scientifique donné par la fondation de la Main à la Pâte (Charpak, 1996), entrepris en 1995 par l'Académie des Sciences<sup>15</sup>.

---

<sup>13</sup> [eduscol.education.fr/ressources-2016](http://eduscol.education.fr/ressources-2016) - Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche

<sup>14</sup> <https://www.fetedelascience.fr/>

<sup>15</sup> <https://www.fondation-lamap.org/>

Kendeou, Van Den Broek, Helder et Karlsson (2014) ont proposé des adaptations du matériel de lecture ; ils ont recommandé d'adapter les textes en fonction de leur objectif. Il est important de différencier les textes utilisés pour enseigner les compétences en lecture et les textes utilisés pour appréhender le contenu en termes de connaissances, pour construire progressivement une représentation cohérente des connaissances. Si le but est d'enseigner des compétences de niveau supérieur telles qu'une maîtrise de l'inférence, alors le niveau de difficulté des textes est à augmenter progressivement en fonction des exigences imposées aux processus d'inférence. Ainsi, les connexions implicites dans le texte seront diversifiées tout en augmentant progressivement la distance entre les omissions conceptuelles dans le texte et leurs référents, méthodiquement et régulièrement. Si le but est d'enseigner le contenu (par exemple, la science, l'histoire), les textes doivent demander le moins d'efforts possible au niveau des processus cognitifs pour limiter les difficultés liées aux inférences notamment. Pour cela il est possible de valoriser les informations importantes à connecter en les plaçant à proximité dans le texte et en rendant les connexions implicites plus précises et explicites (McNamara, Ozuru & Floyd, 2017). Le contenu doit être présenté de manière logique afin d'aider le lecteur à se créer une représentation mentale cohérente du texte avec moins d'efforts en s'appuyant sur les connecteurs (par exemple : tel que, comme, donc ...). En effet, les parties de texte, étant fortement liées, facilitent en général la mémorisation du contenu, même pour des lecteurs en difficulté (Cain, Oakhill, Barnes & Bryant, 2001).

## **2.5 Approches et outils pédagogiques pour améliorer la compréhension de textes**

Dans cette section, nous allons nous focaliser sur les approches interventionnelles qui ont été testées pour améliorer la pratique d'enseignement de la compréhension de textes et sur certains outils existants.

### **2.5.1 Présentation d'interventions destinées à améliorer la compréhension**

Comme nous l'avons vu dans la section 2.3 relevant les difficultés des élèves, les enfants qui commencent leur scolarité avec de bonnes compétences linguistiques pourront développer de bonnes capacités en compréhension de lecture et réussir leurs apprentissages. En revanche, les difficultés linguistiques d'enfants qui ont de faibles compétences en langue orale dans leurs premières années, provenant parfois aussi de facteurs liés au statut économique et social faible de leur environnement, peuvent faire apparaître la langue d'enseignement comme langue seconde. Ces facteurs sont considérés comme des facteurs de risque d'échec de l'acquisition de

compétences linguistiques et limiteront l'accès à la littéracie (Rogde, Hagen, Melby-Lervåg & Lervåg, 2019). De plus, dans une méta-analyse, Hulme, Snowling, West, Lervåg & Melby-Lervåg (2020) ont mis en avant des études (Bowyer-Crane, Snowling, Duff, Fieldsend, Carroll, Miles & Hulme, 2008 ; Hagen, Melby-Lervåg & Lervåg, 2017) qui ont démontré qu'il est possible d'améliorer, faiblement mais significativement, le langage oral des enfants par des interventions linguistiques ciblées. Les effets d'interventions actives et régulières peuvent améliorer ensuite la compréhension en lecture.

Kim, Hemphill, Troyer, Thomson, Jones, LaRusso & Donovan (2017) ont examiné l'efficacité d'une programmation de lectures complémentaires pour les élèves faibles lecteurs dans le Nord-Est des États-Unis. Le programme de lecture complémentaire, *Strategic Adolescent Reading Intervention* (STARI) comprend l'enseignement de la fluidité, du vocabulaire et de la compréhension, et l'échange entre pairs pour promouvoir l'engagement dans la lecture et la compréhension. Les 483 élèves, de la 6<sup>ème</sup> à la 8<sup>ème</sup> année, ont été répartis dans deux conditions expérimentales : condition de contrôle et condition de lecture stratégique. Les élèves STARI ont surpassé les étudiants du groupe contrôle sur les mesures de reconnaissance des mots ( $d = 0,20$ ), l'efficacité de la compréhension de la lecture ( $d = 0,21$ ) et la conscience morphologique ( $d = 0,18$ ). L'engagement en lecture sous sa forme comportementale a été évalué par la participation et l'implication des élèves dans le programme STARI. En mesurant combien d'activités du programme STARI ont été effectuées pendant l'année scolaire, les auteurs soutiennent l'hypothèse d'un engagement effectif qui renforce le développement de la lecture des adolescents ayant des difficultés.

Les avantages de l'enseignement des stratégies de compréhension semblent émerger après relativement peu de temps d'enseignement. L'étude de Connor, Morrison et Petrella (2004) a examiné l'effet de l'enseignement explicite de la compréhension de lecture sur la progression des capacités de compréhension d'élèves de CE2 (N=73). Cette étude longitudinale d'une durée de 6 mois s'est intéressée à l'enseignement explicite et implicite de la compréhension au regard des compétences initiales des élèves. Le niveau initial de l'élève expliquait environ 33% de la variance des scores de compréhension de la lecture à la fin de l'année scolaire. Les résultats ont révélé que les enfants ayant des capacités de compréhension moyennes à inférieures en lecture, au pré-test, étaient ceux qui avaient le plus progressé en compréhension de la lecture grâce à un temps plus important réservé aux enseignements explicites. En revanche, les enfants ayant des capacités de compréhension de lecture supérieures à la moyenne au départ ont eu besoin de moins de temps d'enseignement explicite de la part de l'enseignant. L'enseignement explicite de la compréhension, géré par l'enseignant, a un effet

positif sur le développement de la compréhension en particulier chez les enfants dont les compétences en compréhension de la lecture sont en dessous du niveau scolaire moyen. Les pratiques pédagogiques explicites relèvent d'activités de langage oral et écrit, d'échanges interactifs entre pairs et avec l'enseignant lors de corrections collectives et lors de discussions générées par l'enseignant sur les stratégies de compréhension, c'est-à-dire, faire des hypothèses, déduire, résumer et analyser le contenu du texte. Tous les aspects de la langue orale ont été pris en compte notamment les aspects morphosyntaxiques.

Une seule recherche française à notre connaissance (Bianco et al. 2010, 2012), ayant expérimenté une intervention d'enseignement avec deux programmes, dès la maternelle auprès d'une grande cohorte, a été réalisée. Les enfants ont participé à des programmes de formation spécifiques centrés soit sur les compétences en code, formation à la conscience phonologique, soit sur les compétences en compréhension. Celle-ci comprenait plus spécifiquement les recherches d'incohérences, le travail sur les inférences nécessaires et logiques du texte, le travail sur le raisonnement déductif pour construire un modèle de situation et comprendre la structure d'une histoire. Le but était de faire prendre conscience aux élèves que des difficultés de compréhension existent et que le lecteur se questionne (Bianco et al. 2002 ; 2006). La formation à la conscience phonologique a aidé particulièrement les élèves peu performants à comprendre le principe alphabétique et a amélioré la conscience phonologique. La formation précoce en compréhension a développé les compétences de compréhension orale, et a aussi eu un effet bénéfique pour la compréhension ultérieure en lecture. La taille de son effet le soulignait ( $\beta = 0,32$ ) alors que l'effet a été évalué neuf mois après la fin de tous les programmes de formation. De plus, les auteurs ont suggéré que le travail de compréhension orale développe des capacités linguistiques qui sont impliquées avec les compétences en code dans l'acquisition de la lecture. Les compétences linguistiques des enfants avant l'apprentissage de la lecture (capacités sémantiques, syntaxiques et discursives) les aident à construire les mécanismes d'identification des mots et la compréhension à la suite du décodage.

Les résultats faisant consensus sont rarement applicables et faciles à intégrer dans les classes car la mise en application des connaissances fondamentales demande la mobilisation de chercheurs et d'acteurs de terrain (Gentaz, Sprenger-Charolles, Colé, Theurel & Gurgan, 2013). L'objectif de cette étude était d'évaluer les effets de la mise en place d'un programme d'entraînement axé sur le décodage et la compréhension, afin d'améliorer les performances en lecture d'enfants scolarisés en Réseau d'Éducation Prioritaire (REP). Ce programme comportait quatre types d'entraînement centrés sur : la phonologie, le décodage, la compréhension et la fluence de lecture. Ce travail de recherche a été réalisé à grande échelle, auprès de 2375 élèves

de CP. Les élèves ayant participé au programme n'ont pas progressé de façon plus importante que les élèves du groupe témoin. Cependant, les auteurs signalent que l'implication des enseignants dans le programme a été bénéfique au niveau du comportement individuel et collectif des enfants pour l'ambiance de la classe et le travail en équipe entre professionnels.

### **2.5.2 Pratiques pédagogiques et outils pour améliorer la compréhension en lecture**

Enseigner les stratégies a du sens si elles ne sont pas considérées comme le développement des compétences de lecture à poursuivre mais comme des 'astuces' qui, une fois expliquées et découvertes, sont disponibles pour les enfants, qui peuvent alors les utiliser dans d'autres situations (Willingham, 2006). Dans cette optique, l'enseignement explicite d'une stratégie aide les enfants à comprendre le but de la lecture plus rapidement qu'ils ne le feraient autrement, en découvrant seuls ces stratégies. Willingham (2006) a également expliqué que des effets plus cohérents sont observés lorsque l'enseignement des stratégies de compréhension est appliqué, à partir de la quatrième année d'apprentissage de la lecture.

Bianco et Lima (2017) ont proposé six stratégies pour favoriser l'enseignement de la compréhension de l'écrit et accompagner les stratégies développées par les enfants :

- Activer ou construire les connaissances encyclopédiques du lecteur, activer les connaissances du monde pour les intégrer dans des structures déjà présentes en mémoire, clarifier le vocabulaire ;
- Produire des inférences, grâce au modelage des stratégies de production d'inférences (Lima, Bianco, Guerin, Nanot & Reale-bruyat, 2016) ;
- Se questionner et répondre régulièrement à des questions pour contrôler la compréhension et enclencher un processus de remédiation de la compréhension si nécessaire ;
- Résumer le texte déjà lu pour réactiver les connaissances, exprimer le thème général, faire le point sur l'interprétation du texte (Lima, Sylvestre & Bianco, 2006) ;
- « Visualiser mentalement ou représenter visuellement, à travers le dessin ou le schéma, par exemple, le contenu du texte pour augmenter la précision de cette représentation. » (Bianco & Lima, 2017, p.115) ;
- Réguler sa compréhension, prendre conscience des difficultés de compréhension par exemple avec l'outil d'échelle d'évaluation de la compréhension après un paragraphe :
  - « - j'ai bien compris, je serai capable d'expliquer le paragraphe à mon voisin,
  - j'ai bien compris mais je ne pourrai pas bien expliquer le paragraphe,
  - je n'ai pas compris certaines parties ou l'ensemble du paragraphe. »

Dès l'entrée à l'école, il est essentiel de repérer les élèves à risque avec de possibles fragilités au niveau des habiletés langagières à l'oral (Hulme et al., 2020). Comme vu dans l'étude de Lyytinen et ses collègues (2006) chez les enfants, dès l'âge de 4 ans, les difficultés en compréhension orale engendrent plus tard des difficultés en fluence et compréhension en lecture. De plus, l'inexpérience de certains élèves, via le manque d'exposition à l'écrit pourrait engendrer des difficultés. Le rôle que jouent les enseignants dès la maternelle dans l'approche de l'écrit est essentiel. Un programme d'enseignement de l'orthographe a été proposé par Morin & Montésinos-Gelet (2007), avec des effets significatifs, afin d'aider des élèves à risque à approcher l'écrit.

L'enseignement du lexique est un moyen permettant l'amélioration de la compréhension en lecture. Les connaissances sur le lexique sont mises en relation avec les connaissances sur la syntaxe et la morphologie de la langue. Le sens est construit par une exposition répétée en contexte, notamment lors de lectures, et par l'étude structurée du lexique (Tremblay & Polguère, 2014). D'après Cellier (2011), pour un meilleur apprentissage du vocabulaire, idéalement les outils utilisés en classe doivent rendre compte de la structure du lexique (un réseau d'unités interreliées) et doivent faciliter la mémorisation et la réactivation des mots du vocabulaire de l'apprenant (en particulier grâce aux relations entre les mots). L'enseignant doit organiser, faire des liens, catégoriser, restructurer sans cesse ce qui a été acquis au niveau du vocabulaire car l'acquisition est un processus dynamique (Anctil, 2015).

Des outils d'entraînement à la compréhension en lecture tels que « Lectorino et Lectorinette » (Cèbe & Goigoux, 2009 ; Goigoux & Cèbe, 2013), le « Moniteur de lecture a.r.t.h.u.r » (Rémond, Richaudeau & Bentolila, 1991) proposent un travail sur les stratégies de compréhension et sur le lexique avec des tâches demandées aux élèves de productions orales et écrites pour multiplier et réutiliser le vocabulaire appris dans différents contextes. Les jeux proposés dans le cadre de l'apprentissage, sont aussi des occasions de répétition du lexique appris ou des moments de créativité, de développement des habiletés langagières sous forme ludique (Schmoll, 2016). Il existe des lotos de mots, des mots croisés, des devinettes, des charades, etc. Ces jeux de société ont un intérêt lexical pour accroître les compétences en langue, par exemple : le jeu du prince des mots tordus, Imagidés, Scrabble Flash, La Course aux mots, Dixit etc. Dès la maternelle, les enseignants entraînent les enfants à la compréhension en utilisant des outils tels que la méthode Narramus (GS et CP) pour stimuler et structurer le langage oral et développer la compréhension de la langue orale au travers de messages et textes entendus (Roux-Baron, Cèbe & Goigoux, 2017).

Des dispositifs pédagogiques numériques existent également. C'est le cas du *rallye-lecture*<sup>16</sup> qui est proposé sur une plateforme en ligne depuis septembre 2014. C'est la promotion de la lecture qui est visée au travers de concours en quatre étapes : sélection de livres, lectures, questionnaires, consultations des résultats. Un autre dispositif, le *Réseau des Observatoires Locaux de la Lecture (ROLL)*<sup>17</sup>, initialement commencé en format papier pour certains niveaux de l'élémentaire, est maintenant disponible en format numérique. Son objectif reste le même depuis sa création : développer une démarche pour apprendre à comprendre les textes en associant pédagogie de la compréhension et différenciation. Ce dispositif suit trois étapes sur deux types de textes (narratif et explicatif) : évaluation des compétences des élèves, propositions de remédiation adaptées en fonction des besoins des élèves et ateliers de compréhension choisis par l'enseignant. Un autre exemple « *La Machine à comprendre* »<sup>18</sup> (MAC) propose des ateliers de compréhension de textes afin d'accompagner et d'assurer la compréhension tout au long des ouvrages que choisit le professeur. Un dernier exemple, *TACIT*<sup>19</sup> est un outil de pédagogie différenciée en lecture, capable de proposer des méthodologies adaptées au niveau de compétences de chaque élève, afin de le faire progresser de manière individuelle, grâce à des tests et des exercices adaptés (Quaireau, Lavandier, De La Haye, Noël, Le Bohec & Nogues, 2016). L'outil informatique permet le contrôle de l'activité, des progrès, de l'avancée sur la lecture et le suivi individuel, par exemple, pour le professeur. Mais Laisney, Brandt-Pomares & Ginestier (2011) ont montré que si l'outil informatique informe le professeur de l'accomplissement de la tâche par l'élève, notamment de difficultés rencontrées, cet outil contraint aussi le professeur à suivre l'activité de l'élève. Les auteurs précisent que l'activité de l'enseignant est alors « calée » sur celle des élèves.

L'efficacité d'un entraînement informatisé comme aide à l'amélioration des difficultés de compréhension en lecture de textes a été déjà étudiée par Johnson-Glenberg (2005) auprès de 20 élèves de collège. L'auteur a observé des gains de compréhension chez les faibles 'compreneurs' grâce au développement de stratégies, notamment verbales, dans un environnement numérique. D'après Stetter et Hughes (2010), qui ont réalisé une analyse de la littérature sur les effets de l'enseignement assisté par ordinateur pour améliorer la compréhension de lecteurs en difficultés, les outils informatiques offrent aux élèves ayant des troubles d'apprentissage et des difficultés de lecture un soutien numérique pour la lecture. Les

---

<sup>16</sup> <https://rallye-lecture.fr/>

<sup>17</sup> <https://www.roll-descartes.fr/charte>

<sup>18</sup> <https://fodem-descartes.fr/la-machine-a-comprendre/>

<sup>19</sup> <https://tacit.univ-rennes2.fr/presentation/accueil>

auteurs ont constaté que les effets de l'enseignement avec un outil numérique ne sont pas systématiquement positifs.

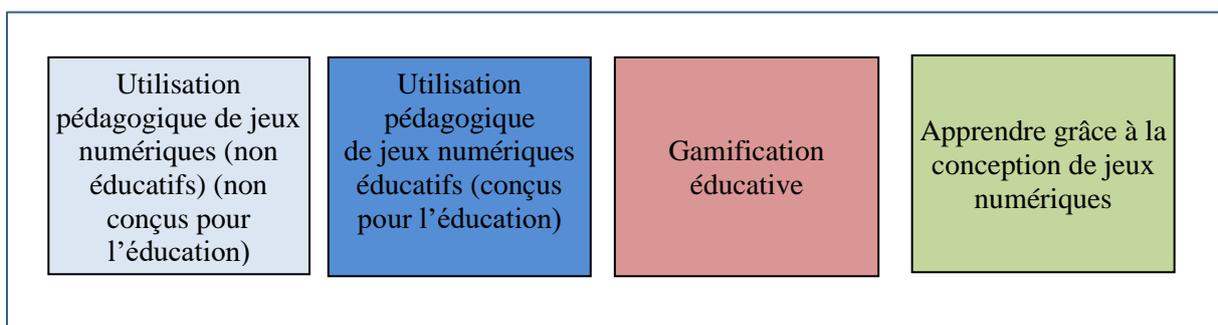
Il apparaît qu'un important travail reste encore à réaliser quant à l'amélioration des conditions d'insertion d'un tel entraînement dans le milieu scolaire et à son acceptabilité. Pourtant les exemples d'insertion du numérique pour la lecture se multiplient pour tous les niveaux de la scolarité. L'outil *LIRALEC*<sup>20</sup>, développé par l'Université de Poitiers propose aux collégiens d'exercer leurs compétences en lecture sur des textes documentaires.

Aujourd'hui les différentes applications développées conduisent l'outil numérique à être de plus en plus souvent présent dans les situations éducatives. L'utilisation du numérique permet de proposer une différenciation dans la classe entre les élèves mais propose aussi d'articuler deux phases importantes du temps d'éducation : l'apprentissage et l'enseignement (Mioduser, Tur-Kaspa, & Leitner, 2000, Lassault & Ziegler, 2018).

« Le système éducatif doit évoluer et, si l'on regarde les évolutions des quarante dernières années, il doit évoluer vers un système d'éducation pour tous tout au long de la scolarité de base, suivi d'un accompagnement d'orientation et d'insertion professionnelle. Cette évolution est en effet largement soutenue par le développement des TIC (Technologies de l'Information et de la Communication) et la généralisation de leur utilisation à l'école. » [Traduction libre] (Wright, et al. (2018), p 45)

Les approches pédagogiques incluant les jeux numériques dans l'éducation ont augmenté. Le terme de jeux numériques pour l'apprentissage (*Digital Game Learning*, DGL), comprend une diversité d'utilisations pédagogiques (Figure 13). L'objectif pédagogique est basé sur la technologie du numérique qui permet l'activité de jeu (Romero, 2020).

Figure 13 : Utilisation pédagogique des jeux numériques pour l'apprentissage (Romero, 2020, p. 22).



<sup>20</sup> <http://cerca.labo.univ-poitiers.fr/liralec/>

Quatre types d'utilisation des jeux numériques pour l'apprentissage existent (Romero, 2020) :

- Les jeux numériques qui n'ont pas été conçus pour l'apprentissage, sans aucune intention pédagogique mais qui sont détournés pour une utilisation pédagogique par exemple *Animal Crossing*<sup>21</sup> et *Simcity*<sup>22</sup> pour apprendre à gérer un budget ;
- L'utilisation pédagogique de jeux éducatifs sérieux qui ont été créés dans un but éducatif, par exemple *Graphogame*<sup>23</sup> pour la lecture, *Kalulu*<sup>24</sup> pour la lecture et les mathématiques ;
- La gamification dans un contexte éducatif, qui utilise la mécanique du jeu, comme, la compétition, le hasard présents dans des contextes non éducatifs par exemple le *jeu de l'oie*<sup>25</sup>, pour s'exercer et apprendre la langue ;
- L'objectif pédagogique est intégré dans le processus de conception interdisciplinaire d'un jeu en tant que processus de mise en forme des connaissances, par exemple l'application *Scratch* pour apprendre à programmer.

## Conclusion

Nous avons vu dans ce chapitre les principaux modèles de référence et des facteurs interdépendants expliquant une grande partie de la compréhension de texte. La simplification de textes peut représenter une aide à l'élaboration des premières propositions de représentation mentale du texte chez le lecteur, relations locales qui constituent la micro-structure du texte. En revanche, la simplification n'intervient pas directement dans l'organisation et la hiérarchisation de ces propositions au niveau global, dans la macro-structure du texte. De plus, la simplification de textes réalisée peut libérer sans doute de la mémoire de travail sollicitée jusqu'au modèle de situation. En exposant les faibles lecteurs à un texte plus accessible, la simplification de textes est un moyen de lutter contre l'effet Matthew (Stanovich, 1986) et peut-être de moins creuser les écarts entre lecteurs. La simplification de textes peut contribuer à percevoir des éléments, connaissances, émotions culturelles et plaisir de lire auxquels certains lecteurs ne pourraient pas avoir accès en lisant des textes originaux. Dans le chapitre suivant, nous présentons les procédés de simplification et d'adaptation de textes en lien avec des facteurs linguistiques du texte, à savoir le lexique, la morphosyntaxe et le discours.

---

<sup>21</sup> <https://animal-crossing.com/fr/>

<sup>22</sup> <https://www.ea.com/fr-fr/games/simcity>

<sup>23</sup> <http://grapholearn.fr/>

<sup>24</sup> <https://moncerveaualecole.com/kalulu-pour-apprendre-a-lire-et-a-calculer/>

<sup>25</sup> <http://jeudeloie.free.fr/plateau.htm>

## Chapitre 3

### La simplification de textes

La simplification consiste à réduire la complexité d'un texte, tout en conservant sa signification d'origine (Siddharthan, 2014). Son objectif est de rendre accessible à un plus large public le contenu du texte et les informations qui en découlent. Elle vise toujours un public cible en tenant compte des caractéristiques de cette population pour définir les modalités d'adaptation du texte.

« La simplification est utilisée dans une variété de contextes, et diverses populations de lecteurs cibles peuvent bénéficier de la simplification du texte. Par exemple, il existe des preuves que la simplification syntaxique facilite la compréhension pour les aphasiques et les personnes sourdes, tandis que les dyslexiques bénéficient de la simplification lexicale. Ce soutien à la simplification du texte manuelle informe et motive généralement la recherche sur la simplification automatique du texte. » [Traduction libre] (Siddharthan, A., 2014, p 268)

Il existe de nombreuses approches concernant la simplification de textes. Dans ce chapitre, nous allons d'abord présenter l'adaptation de textes et les différentes approches de la simplification, puis les niveaux linguistiques sur lesquels s'opèrent les transformations des textes (lexical, morphosyntaxique, discursif). Enfin, nous présenterons plusieurs études ayant évalué les effets de la simplification sur des publics divers.

#### 3.1 Adaptation de textes pour faciliter la lecture

Concernant les aides à la lecture et à la compréhension qui s'adressent principalement à des enfants dyslexiques et à des adultes en difficultés, il faut distinguer deux situations d'adaptation du texte : l'adaptation au niveau typographique pour rendre la lecture plus accessible ou la création de textes pour un public spécifique.

Une première démarche d'adaptation consiste à proposer des aménagements visuels afin de faciliter les traitements perceptifs et de rendre les textes plus accessibles à un public en difficultés, par exemple aux enfants dyslexiques (voir Zorzi et al., 2012). C'est le cas des supports numériques proposés, par exemple, dans la collection « *Dyscool* » chez *Nathan Editions* qui cherche à faciliter le déchiffrement et la compréhension grâce à différentes

adaptations du texte. L'éditeur propose des adaptations sur plusieurs points : longueur des lignes, nature et taille de la police, coloration des phonèmes et syllabes, place de l'illustration etc... L'enfant peut également régler le contraste, utiliser une fenêtre de lecture ou un marque-ligne. Enfin, la compréhension est aidée grâce à un dictionnaire contextuel et la lecture audio. Un autre exemple est celui de la collection « *J'aime lire Dys* »<sup>26</sup> chez *Bayard Editions*. L'éditeur propose en version numérique la présentation du texte avec des modifications possibles telles que la police, la mise en page ou le suivi audio. Un dernier exemple d'aménagement visuel du format de présentation avec des mises en pages aérées et un alignement à gauche, une police spéciale pour la dyslexie, est celui des livres de la collection « *Castelmore Dys* »<sup>27</sup> qui existent en formats papier et numérique.

Une deuxième démarche, pour répondre aux besoins de lecteurs en difficultés, est de créer un texte répondant à un cahier des charges précis. Les prescriptions linguistiques et textuelles sont explicitées aux auteurs pour un public-cible. Les textes sont créés en tenant compte des difficultés des lecteurs et ils ne font pas référence à un texte original considéré comme complexe ou difficile (Gala, François, Javourey-Drevet & Ziegler, 2018). C'est le cas des livres de la collection *Colibri*,<sup>28</sup> *Belin Education*, qui propose des livres pour des enfants à besoins spécifiques, concrètement pour des lecteurs dyslexiques, avec quatre niveaux progressifs en fonction de la fréquence de correspondance graphème-phonème en français et de leur complexité croissante (par exemple les graphèmes « on » et « t » correspondent au premier niveau, les graphèmes « ss » et « re » au deuxième, les graphèmes « ée » et « s » au troisième et les graphèmes « pr » et « pl » au quatrième niveau). De plus, dans cette collection, les histoires sont accompagnées d'activités pédagogiques afin de préparer la lecture mais aussi de vérifier la compréhension : échauffement pour s'entraîner à lire le graphème et préparer la lecture ; quiz et questionnaire de lecture pour vérifier la compréhension du texte ; présentation des personnages principaux pour bien se repérer dans l'histoire, dictionnaire donnant le sens des mots susceptibles d'être inconnus du lecteur. L'auteur Joel Verbauwhede<sup>29</sup> (2017) propose aussi des textes créés en tenant compte des difficultés des lecteurs sans référence à un texte original considéré comme complexe ou difficile. Ces textes sont publiés aux éditions « *Mondes parallèles* » (2020) en versions papier et ebook, pour tous les lecteurs, jeunes et adultes. Leur mise en pages est spécialement adaptée pour faciliter la lecture des personnes en difficultés.

---

<sup>26</sup> <https://www.jaimelirestore.com/livres/6-8-ans/special-dyslexie.html>

<sup>27</sup> <https://www.castelmore.fr/dys>

<sup>28</sup> <https://www.materieldys.com/16-livres-dys-colibri>

<sup>29</sup> <http://editionsmondesparalleles.free.fr/>

Une troisième démarche consiste à proposer des textes résumés. L'outil d'adaptation RESOOMER<sup>30</sup>, par exemple, permet de résumer des textes. Pour autant, l'objectif du résumé étant de conserver les idées et le contenu du texte sans interprétation, sans modification des idées, l'utilisation de l'outil se limite à des textes argumentatifs et documentaires. Un autre outil, en langue suédoise, est conçu pour résumer des textes, COGSUM<sup>31</sup>. Suite à l'analyse des textes obtenus par cet analyseur, Smith & Jönsson (2011) mettent en évidence un impact positif sur la lisibilité des textes de différents genres (textes de journaux, textes de magazines scientifiques, textes administratifs relatifs à l'assurance sociale) et avec des résumés de différentes longueurs.

### 3.2 Les différentes approches de la simplification

La démarche de simplification de textes vise à transformer un texte original en un texte simplifié afin de le rendre plus accessible à un public-cible considéré comme ayant des difficultés de lecture et/ou compréhension, par exemple les lecteurs aphasiques (Carroll, Minnen, Canning, Devlin & Tait, 1998), des adultes illettrés (Aluísio & Gasperin, 2010), ou des apprenants en langue seconde (Keshavarz, Atai & Ahmadi, 2007). La simplification de textes se réalise suivant quatre procédures majeures : déplacements, substitution, expansion ou réduction des contenus (au niveau du mot, des constituants ou de la phrase). Un exemple de ce type de démarche est celle mise en œuvre aux États-Unis avec le *Plain Language*<sup>32</sup> (ou *Plain English*), dans les années 1970, pour répondre aux besoins d'une partie de la population ne comprenant pas les documents administratifs (Gala et al., 2018). Les consignes de simplification du *Plain Language* ont été développées au milieu des années 1990. Elles portent sur le type de mots ou de constructions syntaxiques à éviter (les mots du jargon administratif, les phrases négatives, la voix passive, etc.). Elles incluent également des consignes stylistiques, par exemple favoriser les paragraphes courts et des éléments de mise en page visuels (par exemple, titre, liste ...) <sup>33</sup>. Un autre exemple est celui du corpus *Newsela*<sup>34</sup> (corpus en anglais de 1130 textes) où quatre niveaux de simplification ont été produits manuellement par les éditeurs : des articles de presse ont été réécrits pour des élèves et forment un corpus parallèle. Pour les enfants, un autre exemple est celui des adaptations de textes encyclopédiques avec une

---

<sup>30</sup> <https://resoomer.com/>

<sup>31</sup> [https://www.researchgate.net/figure/The-COGSUM-interface\\_fig3\\_215528352](https://www.researchgate.net/figure/The-COGSUM-interface_fig3_215528352)

<sup>32</sup> <https://www.plainlanguage.gov/>

<sup>33</sup> <https://plainlanguage.gov/about/definitions/>

<sup>34</sup> <https://newsela.com/>

rédaction facilitée. Par exemple, sur le modèle de *Wikipédia*, *Vikidia*<sup>35</sup> propose des articles avec des contenus simplifiés afin de rendre les informations plus accessibles aux enfants de 7-11 ans. Il existe une version de *Vikidia* en français mais aussi en espagnol, en italien, en russe, en anglais, en catalan, en allemand mais aussi en basque, en sicilien, en grec et en arménien. Enfin, l'éditeur Nathan propose, en support papier dans la collection *Dyscool*<sup>36</sup>, des simplifications de textes :

- « - Les phrases complexes sont réécrites : en phrases courtes et simples (Sujet + Verbe + Complément), chronologiques et sans incise.
- Les mots à graphie complexe ou longs sont remplacés, les mots et expressions complexes sont expliqués, les figures de style sont rendues explicites ou remplacées.
- Les personnages sont explicitement nommés. »

Les caractéristiques communes aux exemples cités ci-dessus sont de s'adresser principalement aux enfants, d'être édités en format papier ou présentés en version numérique, d'être souvent payants. Les adaptations consistent à alléger le contenu, à supprimer des informations non fondamentales, à réexpliquer parfois certains termes (exemple *Vikidia*), à remplacer des mots considérés complexes et rares.

À côté des initiatives comme celles que nous venons de mentionner, la simplification automatique de textes (SAT) a commencé à se développer dans le domaine du traitement automatique des langues à la fin des années 1990, initialement considérée comme une étape de prétraitement destinée à faciliter d'autres tâches automatiques comme par exemple l'analyse syntaxique ou l'extraction d'informations (Siddharthan, 2014).

Peu après, la SAT a été envisagée comme un moyen de simplification de la lecture de documents pour divers groupes cibles (Saggion, 2017). Ainsi, de nombreux projets de recherche ont vu le jour, avec pour cible des publics spécifiques dans plusieurs domaines, par exemple la dyslexie (Rello, Baeza-Yates, Bott & Saggion, 2013a), l'autisme (Barbu, Martín-Valdivia, Martínez-Cámara & Ureña-López, 2015 ; Yaneva & Evans, 2015), l'illettrisme (Candido, Maziero, Specia, Gasperin, Pardo & Aluísio, 2009). Des travaux existent aussi dans le domaine de la terminologie médicale (Cardon, 2018) afin de rendre les textes de spécialité plus accessibles à un public néophyte. Ces travaux existent pour plusieurs langues, comme l'anglais (Woodsend & Lapata, 2011), l'espagnol (Saggion, Gómez-Martínez, Etayo, Anula & Bourg, 2011), le portugais (Aluísio, Specia, Pardo, Maziero, Caseli & Fortes, 2008 ; Wilkens, Dalla

---

<sup>35</sup> <https://fr.wikidia.org/wiki/Vikidia:Accueil>

<sup>36</sup> <http://dyscool.nathan.fr/>

Vecchia, Boito, Padró & Villavicencio, 2014), le japonais (Inui, Fujita, Takahashi, Iida & Iwakura, 2003), le néerlandais (Ruiter, Beijer, Cucchiarini, Krahmer, Rietveld & Strik, 2012 ; Bulté, Sevens & Vandeghinste, 2018), l'italien (Dell'Orletta, Montemagni & Venturi, 2011 ; Barlacchi & Tonelli, 2013), le suédois (Rybing, Smith & Silvervarg, 2010) et le français (Seretan, 2012 ; Brouwers, Bernhard, Ligozat & François, 2014).

La SAT comprend trois tâches principales : la simplification lexicale, syntaxique et discursive. À ce jour, des systèmes sont en cours d'élaboration mais un système qui simplifie automatiquement un texte n'existe pas encore en français et les résultats sont très inégaux selon les langues.

### 3.3 Les différents niveaux de simplification de textes

Les simplifications d'un texte peuvent être réalisées à différents niveaux linguistiques. Dans ce qui suit, des exemples des trois niveaux de simplification sont présentés en bleu. La version originale (ORIG.) est suivie de la version simplifiée (SIMP.). Pour chaque niveau, un exemple est extrait de textes littéraires (LIT) et un exemple vient des textes scientifiques (SCI).

- **Simplification lexicale**, elle consiste généralement à remplacer des mots ou des segments de phrases par un équivalent plus simple.

- Ex. LIT : Bientôt revint le temps de *la cueillette* / Bientôt revint le temps de *la récolte*.

- Ex. SCI : La *production* de savon *s'est développée* grâce à la culture *très répandue* de l'olivier [...] / La *fabrication* de savon *a évolué* grâce à la culture *importante* de l'olivier [...]

- **Simplification morphosyntaxique**, elle consiste à remplacer ou déplacer des mots ou des segments de phrases afin de rendre les textes plus abordables.

- Ex. LIT : Et puis *docile, caressante, se laissant traire* sans bouger, *sans mettre* son pied dans l'écuelle, un amour de petite chèvre. / Elle était *sage, aimait les caresses, se laissait traire* sans bouger. Elle *ne mettait pas* son pied dans son assiette. C'était un amour de petite chèvre.

- Ex. SCI : *La tradition orale veut* qu'en nettoyant avec des cendres sa marmite recouverte de graisse, une femme aurait obtenu une matière qui serait à l'origine du savon. / *On raconte* qu'une femme, nettoyant avec des cendres sa marmite pleine de graisse aurait trouvé une matière devenue savon.

- **Simplification discursive**, elle consiste à remplacer les pronoms par leurs référents, à supprimer les expressions référentielles (déterminants possessifs ou pronoms), à ajouter des connecteurs logiques pour expliciter la logique du raisonnement et la cohésion textuelle.

- Ex. LIT : Il *la* regardait, sur le sol, comme il aurait regardé une bête dangereuse. / Il regardait *la baguette*, sur le sol, comme il aurait regardé une bête dangereuse.

- Ex. SCI : Les fleurs donnent des fruits appelés cabosses. *Elles* sont récoltées à la main, deux fois par an. / Les fleurs donnent des fruits appelés cabosses. *Ces cabosses* sont récoltées à la main, 2 fois par an.

Une même phrase peut présenter différents phénomènes linguistiques susceptibles d'être simplifiés. Ainsi, une transformation lexicale peut aussi entraîner d'autres changements (Gala et al., 2018) comme dans les phrases : « Alors, *exaspéré*, le meunier acheta un jour un gros chat *tigré* pour chasser toutes ces souris. Le meunier avait *si* mauvais caractère *qu'*il le frappait souvent à coups de pied » (ORIG.) / « Alors, *mécontent*, le meunier acheta un jour un gros chat pour chasser toutes ces souris. Le meunier avait mauvais caractère. Il frappait le chat à coups de pied » (SIMP.).

Pour identifier certaines modifications faites lors de la simplification du texte original, il est possible de s'appuyer sur un outil : MEDITE (Sorbonne Université)<sup>37</sup>. L'outil présente en comparant les deux formes, originale vs simplifiée, les remplacements (bleu), les insertions (vert), les suppressions (rouge) mais n'apporte pas de précision sur les catégories des mots, les parties du discours. Par la suite, nous présentons les différents niveaux de simplification dans le détail.

### 3.3.1 Simplification lexicale

La simplification lexicale peut soutenir les lecteurs ayant une mauvaise maîtrise du décodage, une reconnaissance de mots déficitaire, une lecture lente ou encore des déficits en vocabulaire, alors que, selon les quatre modèles présentés dans le chapitre 2, les facteurs de la compréhension en lecture sont interdépendants.

On distingue deux types de simplifications lexicales (Siddharthan, 2014). D'une part, les simplifications qui visent à expliquer les unités lexicales complexes au moyen d'une définition ou d'une paraphrase explicative (par exemple dans Vikidia) ; d'autre part, les simplifications consistant à remplacer les unités lexicales complexes par un synonyme plus

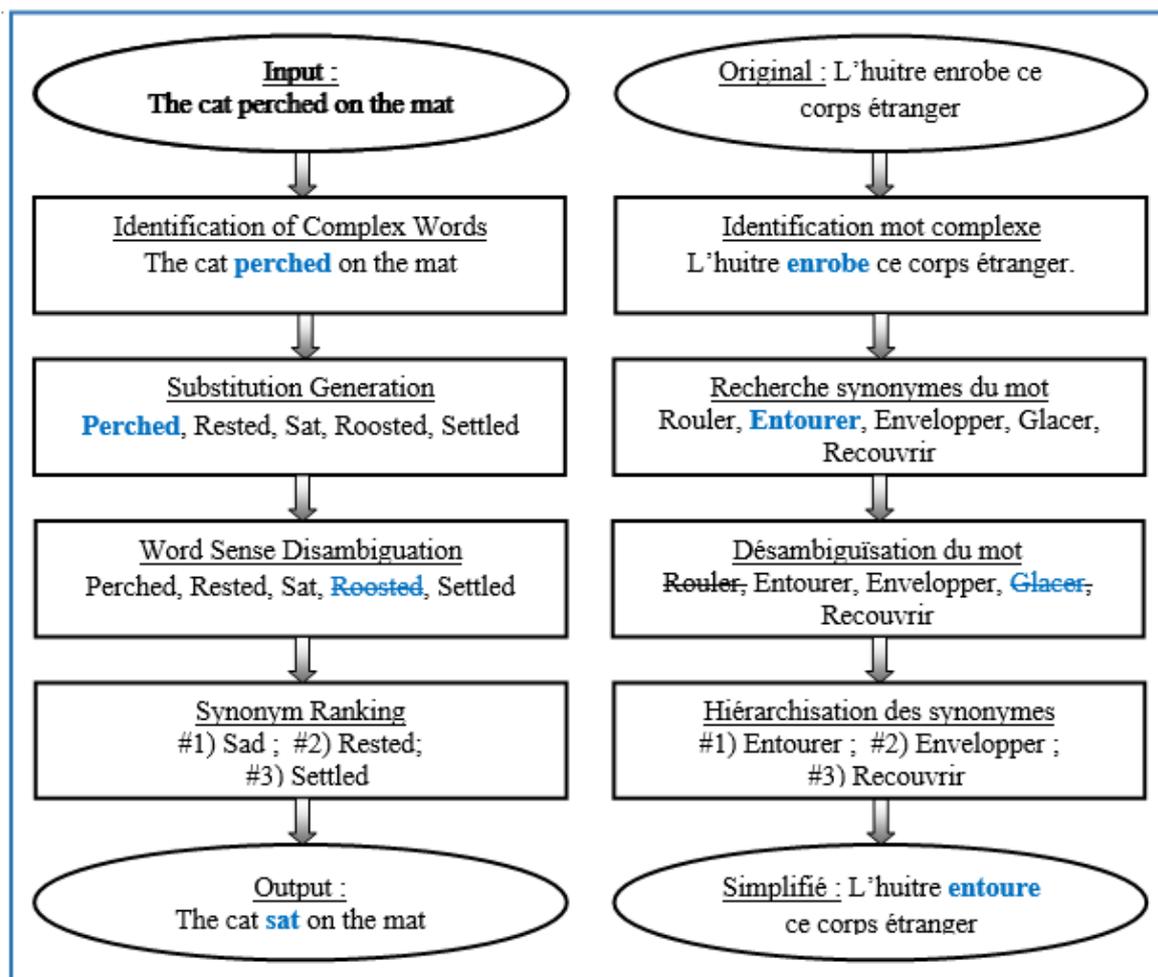
---

<sup>37</sup> <http://obvil.lip6.fr/medite/>

simple. Dans ce cas, plus fréquent, la difficulté principale est d'identifier le mot dans son contexte et de le remplacer par un substitut alternatif approprié (qui doit avoir le même sens). Il peut s'agir de suppression d'adverbes, de substitutions d'unités lexicales par des synonymes plus fréquents (généralement plus courts) choisis sur la base des informations présentes dans certaines ressources. Par exemple, pour le français, il est possible de se référer à Lexique 3 (New, Pallier, Ferrand & Matos, 2001) ou Manulex (Lété, Sprenger-Charolles & Colé, 2004) ou Resyf (Billami, François & Gala, 2018).

Selon le modèle de Shardlow (2014), la simplification lexicale suit différentes étapes (Figure 14) : (1) identifier les mots complexes, (2) générer une liste de synonymes appropriés en fonction du contexte et (3) identifier le synonyme le plus simple de la liste à utiliser comme substitut du mot complexe.

Figure 14 : Exemple de simplification lexicale d'après le modèle de Shardlow (2014).



Ces étapes ont été reprises et développées ensuite par Paetzold et Specia (2017) :

1. L'identification des mots complexes consiste à décider, pour une phrase donnée, quels sont les mots qui risquent de ne pas être compris par la population cible.

2. La génération de synonymes ou autres substituts possibles consiste à trouver des mots ou des expressions pouvant remplacer le mot complexe cible.
3. La sélection du substitut consiste à décider quels substituts seraient les meilleurs candidats pour remplacer le mot complexe sans altérer la signification de la phrase et son contexte.
4. La hiérarchisation des substituts consiste à ordonner les candidats restants selon leur simplicité, le plus simple étant celui choisi pour remplacer le mot complexe.

Dans cette procédure, la première étape est cruciale. En effet, la notion de complexité lexicale est liée à de multiples caractéristiques dont certaines difficilement quantifiables : le niveau d'abstraction associé au concept qu'un mot décrit, la capacité d'un mot donné à susciter des images mentales, la fréquence d'exposition à un mot, l'âge auquel le sujet a acquis le mot (Jauhar & Specia, 2012). Il existe à l'heure actuelle plusieurs méthodes de simplification semi-automatique du lexique, notamment en espagnol et en anglais. En espagnol *Context-Aware Synonym Simplification Algorithm* (CASSA) (Baeza-Yates, Rello & Dembowski, 2015) ou LexSiS (Bott, Rello, Drndarević & Saggion, 2012) sont deux exemples remarquables. Pour l'anglais nous citerons LEXenstein<sup>38</sup> qui fournit des approches de substitution et de classement (Paetzold & Specia 2015). Deux paramètres interviennent classiquement dans la simplification lexicale et ont été abondamment décrits dans la littérature liée à la lisibilité (cf. François 2011 par exemple) : la fréquence et la longueur des mots.

La fréquence d'un mot quantifie le nombre d'occurrences d'apparition de ce mot dans la parole ou l'écrit. La fréquence a ainsi un effet sur la lisibilité, car plus un mot est fréquent, plus on a de chances de l'avoir déjà rencontré à l'oral comme à l'écrit et donc, de correctement le décoder et l'interpréter. En effet, l'effet de fréquence, c'est-à-dire le fait de lire et reconnaître un mot fréquent plus rapidement et plus correctement qu'un mot moins fréquent, est sans doute l'effet le plus robuste et omniprésent en psycholinguistique. Une étude de Rello, Baeza-Yates, Dempere-Marco et Saggion (2013b) sur la simplification de textes auprès de 46 adultes espagnols, dont 23 personnes dyslexiques et 23 personnes dans le groupe témoin, a montré que l'utilisation de mots plus fréquents et plus courts a permis aux participants dyslexiques de lire plus vite. Parmi le groupe-contrôle aucun effet significatif en vitesse de lecture n'a été constaté.

D'autre part, la longueur d'un mot, qui s'apprécie par le nombre de lettres qu'il contient, est une variable qui a un impact dans la lisibilité d'un texte (elle est, par ailleurs, fortement corrélée à la fréquence). Pour les personnes dyslexiques, par exemple, il est bien établi qu'elles

---

<sup>38</sup> <http://ghpaetzold.github.io/LEXenstein/>

sont très sensibles à la longueur des mots puisque leur lecture est encore sérielle : le temps de lecture augmente alors de façon linéaire avec chaque lettre supplémentaire (Zoccolotti, De Luca, Di Pace, Gasperini, Judica & Spinelli, 2005).

### 3.3.2 Simplification morphosyntaxique

La simplification morphosyntaxique peut soutenir les lecteurs ayant des difficultés à élaborer la base du texte ou ayant peu de compétences syntaxiques et sémantiques, alors que, selon les quatre modèles présentés dans le chapitre 2, les facteurs de la compréhension en lecture sont interdépendants. Elle comprend plusieurs procédures :

- Remplacer les mots complexes ou peu fréquents par des mots les plus fréquents de la même famille morphologique.
- Modifier les morphèmes grammaticaux, notamment changement des flexions verbales, comme le remplacement du passé simple par le présent.
- Réorganiser des structures syntaxiques complexes, notamment les propositions négatives, les tournures impersonnelles, le discours indirect et les propositions subordonnées pour faciliter la compréhension (Brouwers et al, 2014).
- Diviser une phrase longue en plusieurs phrases indépendantes visant à supprimer des propositions relatives, coordonnées, subordonnées, prépositionnelles ou adjectivales.
- Regrouper plusieurs syntagmes en une seule phrase concise.
- Transformer la voix passive en phrase active.

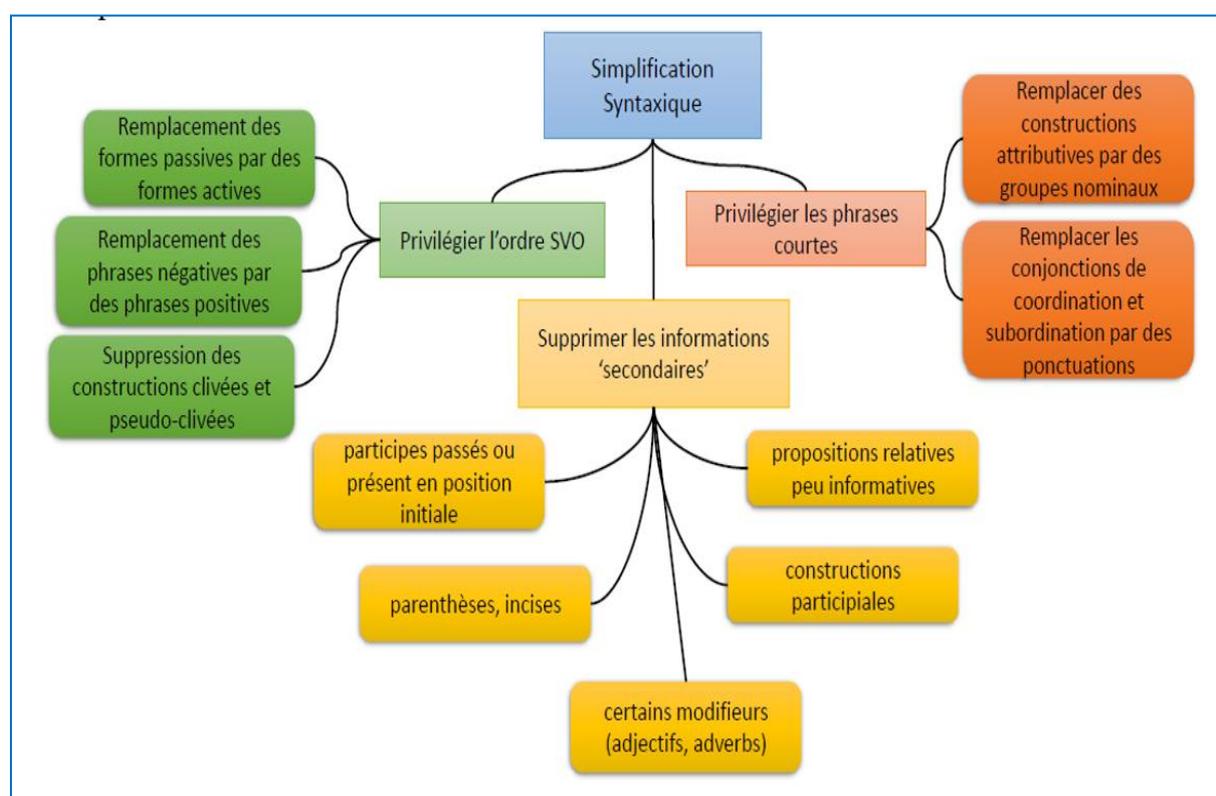
Brunato, Dell'Orletta, Venturi et Montemagni (2014), pour l'italien, ont appliqué des méthodes de SAT regroupant plusieurs catégories d'opérations de simplification. Ils ont travaillé sur de courts romans pour enfants avec un corpus parallèle, c'est-à-dire l'alignement des phrases en parallèle, originales *versus* simplifiées, de 32 textes authentiques et de leur équivalent simplifié manuellement. Les auteurs, suite à l'analyse du corpus, proposent des règles : découpage (au niveau des conjonctions ou des relatives), regroupement en une seule phrase, changement d'ordre (syntaxique), insertions (sujet, verbe, autres constituants), suppressions et transformations (lexicales, morphosyntaxiques et syntaxiques).

Le travail de Seretan (2012), en français, s'est concentré sur l'application de méthodes de SAT pour la réduction de la complexité syntaxique. Elle a choisi des articles de journaux comme corpus pour son étude du système de simplification. Un ensemble de règles a été défini, applicables manuellement, pouvant être complété par la détection semi-automatique de constructions syntaxiques nécessitant une simplification. Les transformations syntaxiques

modifient la structure d'une phrase complexe, c'est-à-dire le fractionnement des propositions, l'extraction des propositions et le désenclavement, tout en laissant intactes (la plupart) des structures des propositions.

Gala, Todirascu, Bernhard, Wilkens et Meyer (2020a) ont modélisé, suite à l'analyse de deux corpus parallèles les transformations syntaxiques qui aident les enfants faibles lecteurs et dyslexiques (Figure 15).

Figure 15 : Typologie des variations syntaxiques en vue d'une simplification de textes (Gala et al., 2020a)



La Figure 15 (Gala et al., 2020a) présente les règles de simplification syntaxiques. Chacun des trois objectifs de la simplification syntaxique, présentés dans des rectangles colorés, est lié aux opérations effectuées manuellement dans les corpus analysés. Dans les rectangles aux angles arrondis il est possible d'identifier les moyens d'atteindre les objectifs de simplification visés : privilégier l'ordre Sujet-Verbe-Objet, (vert) ; supprimer les informations « secondaires » (jaune) et enfin privilégier les phrases courtes (orange). Gala et al., (2020a) ont analysé notre corpus expérimental. Parmi les transformations syntaxiques, la coupure de phrase est la plus fréquente tant dans les corpus littéraires que scientifiques. Il est à noter que les signes de ponctuation sont plus importants dans les textes simplifiés (14,40% vs 13,58%). En ce qui concerne les catégories du discours une augmentation des noms communs dans le corpus simplifié s'explique notamment par la substitution des pronoms par leur référent. Le nombre de

verbes (formes personnelles) augmente également au détriment des verbes à l'infinitif (formes non personnelles). Le nombre de pronoms, de conjonctions, d'adjectifs, d'adverbes, diminue. La première catégorie du discours, les pronoms, diminue en raison du remplacement par les référents, alors que les autres catégories (conjonctions, adjectifs, adverbes) diminuent par suppression si cette dernière n'entraîne pas une perte de cohérence logique. Dans le corpus qui a été analysé, les trois niveaux de transformations ont été observés : lexical, morphosyntaxique et discursif.

L'une des problématiques de la simplification syntaxique est la section de phrases qui peut entraîner la suppression de conjonctions et d'une part d'explicitation apportée par ces conjonctions. En effet le remplacement du rôle sémantique des conjonctions par une paraphrase augmente la longueur des phrases, leur suppression nécessite donc d'effectuer davantage d'inférences qui peuvent être erronées et entraîner une perte de sens. C'est pourquoi la suppression de conjonctions, notamment si elles peuvent permettre un lien causal (car, or, donc) n'est pas systématique. Le découpage des phrases n'entraîne pas nécessairement la suppression de la conjonction. Mason et Just (2004) ont constaté que le manque de cohésion entre des phrases consécutives augmentait le temps de lecture des sujets qui doivent effectuer de nombreuses inférences, réduisant ainsi les ressources mobilisées pour les processus de compréhension.

### 3.3.3 Simplification discursive

La simplification discursive peut soutenir les lecteurs ayant des difficultés à traiter les inférences ou peu acculturés à l'écrit ou peu ouverts sur les connaissances du monde, alors que, selon les quatre modèles présentés dans le chapitre 2, les facteurs de la compréhension en lecture sont interdépendants. La simplification discursive consiste à reformuler les chaînes de référence afin de mettre en évidence l'élément déjà évoqué : par exemple le remplacement de pronoms personnels par leur référent. Les pronoms peuvent être ambigus en induisant des inférences et notamment anaphoriques<sup>39</sup>. Lorsque des référents multiples sont présents dans le texte, la tâche de résolution des pronoms est encore plus complexe (Martins & Le Bouédec, 1998). Selon Sprenger-Charolles (2017), la compréhension de ces formes, appelées anaphores, dépend de trois principaux facteurs dont le plus important est la saillance du référent. Ainsi, le personnage principal d'un récit est plus saillant que le, ou les, personnages secondaires. Les reprises pronominales, ou nominales, d'un nom seront donc préférentiellement rattachées à ce

---

<sup>39</sup> Processus par lequel un segment du discours (dit *anaphorique*) renvoie à un autre segment (dit *antécédent*) apparu dans le même contexte. <https://www.larousse.fr/dictionnaires/francais/anaphore/3258>

personnage. Un second facteur est la distance entre la première dénomination et sa reprise, plus elle est grande, moins il est aisé d'établir un lien entre les deux. Le troisième facteur est le parallélisme des fonctions : dans l'énoncé « **Luc** est venu avec Pierre. **Il** a 22 ans », le rattachement du pronom sujet « il » se fait préférentiellement avec le nom qui est également le sujet de l'énoncé précédent, à savoir « Luc ». Selon Schneedecker (1997) les chaînes de référence sont composées d'au moins trois expressions référentielles (appelées aussi *maillons*), qui renvoient à la même entité de discours. Les chaînes de référence représentent des obstacles à la compréhension.

Par exemple, voici un extrait de « *L'œil du loup* », Daniel Pennac<sup>40</sup>, texte littéraire utilisé dans notre expérimentation :

« Son enclos vide, car la louve est morte la semaine dernière. Son enclos triste, avec son unique rocher gris et son arbre mort. Puis le loup fait demi-tour, et voilà de nouveau ce garçon, avec sa respiration régulière, qui fait de la vapeur blanche dans l'air froid. « Il se lassera avant moi », pense le loup en continuant de marcher. Et il ajoute : « Je suis plus patient que lui. » et il ajoute encore : « Je suis le loup. » »

Dans ce passage le garçon est évoqué par le biais de quatre mentions (ce garçon, qui, il, lui) qui créent une chaîne de références. Le nombre de références multiplie les obstacles inférentiels par rapport à la présence d'un unique référent. De plus, dans ce passage, le pronom « il » est multi-référentiel (le garçon et le loup).

Un autre exemple dans un documentaire scientifique, utilisé également dans nos expérimentations : « *Comment les huitres fabriquent-elles des perles ?*<sup>41</sup> »

« Il vous est sûrement déjà arrivé de croiser le chemin de coquillages à l'aspect repoussant et pourtant très appréciés par de nombreuses personnes : les huitres. Aujourd'hui nous allons vous faire découvrir qu'au-delà de leur image, ils peuvent se révéler être de véritables orfèvres. Réparties sur de nombreuses mers du globe, les huitres font partie de la famille des mollusques bivalves, animaux possédant une coquille pouvant s'ouvrir en deux parties et vivant accrochés sur différents supports comme les rochers ou encore enterrés sous divers sédiments (sable, boue, etc...). »

---

<sup>40</sup> Source : <http://educalire.fr/extraits-l-oeil-du-loup.php>

<sup>41</sup> Source : <http://kidiscience.cafe-sciences.org/articles/comment-les-huitres-fabriquent-elles-des-perles/>

Dans ce passage, on note que les huitres sont citées sous six mentions (coquillages, huitres, ils, véritables orfèvres, mollusques bivalves, animaux). Cette chaîne de référence multiplie d'autant les difficultés de compréhension du passage qu'elle mobilise des catégories différentes de mots et des connaissances du monde, antérieures à la lecture : expression imagée, synonyme et classement scientifique des espèces. Les liens référentiels s'appuient sur la connaissance directe d'un univers de référence (Corblin, 1995). Parfois, on retrouve des références associatives, établies entre un nom et une partie constituante (*animaux, coquillages, mollusques bivalves, huitres*). Les chaînes de référence font partie des liens de cohérence qui participent à la construction globale du sens que le lecteur imagine (Charolles, 2006).

Todirascu, François, Bernhard, Gala, Ligozat & Khobzi (2017) ont analysé un corpus de 84 textes littéraires, extraits de manuels de français langue étrangère (FLE), niveaux A2 à C1 (Niveaux du Cadre Européen Commun de Référence pour les Langues ou CECRL)<sup>42</sup> et un corpus de 44 textes informatifs extraits des sites Wikipédia et Vikidia. Le rapport entre le nombre de maillons inclus dans les chaînes de références et la longueur du texte est un paramètre important à considérer ; ce rapport fixe la densité des maillons. En effet, le nombre de maillons divisé par le nombre de mots du texte permet de juger de la difficulté du texte sur le plan syntaxique. D'autre part les auteurs ont réalisé un premier constat : la nature des maillons ou la composition de chaînes seraient attachées davantage au type de texte (informatif vs narratif) qu'à la difficulté. Les auteurs précisent que la longueur moyenne des chaînes approche 5,98 maillons dans les textes narratifs et 3,88 maillons dans les textes informatifs. Le nombre des entités est plus important dans les textes informatifs (2,82) que dans les textes narratifs (1,75). Les expressions sont moins nombreuses dans les textes informatifs même si elles sont plus longues que dans les textes narratifs. De plus, les chaînes de références sont plus longues dans le corpus *Wikipédia*, destiné aux adultes que dans celui de *Vikidia*, destiné aux enfants. Concernant les textes littéraires, la longueur des chaînes augmente avec la difficulté des textes (A2 vers C1). Trouver l'antécédent ou le référent du pronom nécessite souvent un niveau supérieur de connaissances linguistiques ou une production accrue d'inférences.

Enfin, Lima et Bianco (1999) ont mis en évidence, grâce à une expérience auprès d'élèves du CE2 et CM2, que l'interprétation des pronoms-sujets (il) est plus simple que celles des pronoms-objets indirects (lui). Plus la langue est maîtrisée, plus l'interprétation est facilitée. Le pronom-objet est souvent mal compris alors qu'il est fréquemment rencontré dans les textes

---

<sup>42</sup> <https://www.cambridgeenglish.org/fr/exams-and-tests/cefr/>

écrits. Ce point est susceptible de représenter un obstacle important dans la compréhension de texte. Voici quelques exemples tirés de l'article de Lima et Bianco, (1999, p. 87) :

Exemple 4 : « Gérard appelle Jérôme au téléphone. Il lui parle en anglais. »

Exemple 5 : « Jean appelle Georges au téléphone. Il lui laisse un message sur son répondeur. »

Exemple 6 : « Au guichet de l'hôtel, Pierre questionne Paul. Il lui indique la direction à suivre. »

### **3.4 Études ayant évalué les effets de la modification de textes**

Une première approche pour faciliter l'accès à la compréhension des lecteurs en difficulté est la révision du texte via l'ajout d'informations, la modification du texte et l'explicitation causale. Nous présentons par la suite deux études fondées sur une révision manuelle de la cohérence du texte (sous-section 3.4.1) et cinq études relatives à la simplification de textes (sous-section 3.4.2).

#### **3.4.1 Adaptation de textes**

Linderholm, Everson, Van Den Broek, Mischinski, Crittenden et Samuels (2000) ont étudié comment la compréhension de textes pouvait être améliorée en révisant et modifiant la structure de cohérence pour influencer la compréhension de lecteurs plus ou moins qualifiés. Les aménagements de la cohérence consistaient à : (a) organiser les événements textuels dans l'ordre temporel, (b) rendre explicites les éléments implicites et (c) améliorer des ruptures de cohérence causées par une explication inadaptée, une causalité multiple ou des relations inférentielles distantes. Des étudiants de premier cycle universitaire venant de domaines de formation différents ont lu des versions révisées et originales de récits classés par les auteurs comme faciles et difficiles. Les résultats ont montré que tous les lecteurs, qu'ils soient experts ou moins qualifiés ont bénéficié des modifications, mais uniquement pour les textes considérés comme difficiles.

Beck, McKeown, Sinatra & Loxterman (1991) ont aussi modifié des textes extraits de manuels scolaires en sciences humaines du niveau de cinquième année primaire (CM2, N=85). Leur travail a porté sur quatre extraits d'un livre d'histoire traitant d'une période qui a conduit à la révolution américaine. Les auteurs ont « révisé » les informations de manière à exposer le raisonnement qui relie une cause à un événement et un événement à une conséquence. Les versions originales des quatre textes et les versions « révisées » pour l'étude ont été lues par des élèves de classe moyenne d'une même école, au cours de séquences individuelles en lecture silencieuse. Les élèves devaient ensuite produire un rappel des textes et répondre à des

questions ouvertes de compréhension. Les élèves qui ont lu les versions « révisées » ont rappelé davantage d'informations et ont répondu à davantage de questions que les sujets qui ont lu les versions originales.

### **3.4.2 Simplification de textes avec ou sans apport d'informations**

D'autres études ont utilisé l'apport d'informations et la simplification comme nous l'avons définie en début de ce chapitre. Nous avons examiné trois études réalisées auprès d'enfants entre 8 et 12 ans, en langue française et deux études auprès d'adultes, l'une en anglais, langue étrangère et l'autre en langue espagnole.

#### **➤ Études auprès d'enfants en langue française**

La simplification a également été considérée dans une perspective pédagogique pour des élèves de CM2 dans une étude réalisée par Marin, Crinon, Legros et Avel (2007) en situation classe portant sur différentes aides lors de la lecture de textes scientifiques. Les 156 élèves d'environ 10 ans ont été répartis en 4 groupes. La compréhension du contenu du même texte, sous des formes différentes, a été proposée aux 4 groupes. Le groupe 1 (G1) qui est le groupe témoin a lu le texte original. Le groupe 2 (G2) est celui qui a bénéficié d'ajouts d'informations explicitant le vocabulaire difficile du texte ; le groupe 3 (G3) a aussi bénéficié d'informations supplémentaires (ajouts d'informations facilitant l'activité inférentielle et permettant la construction de la cohérence de la signification globale du texte, relations causales et inférences explicitées) et enfin le groupe 4 (G4) a lu le texte simplifié sur le plan lexical et/ou syntaxique. Tous ont réalisé les mêmes tâches : lecture du texte sur 15 minutes, questionnaire sur la scolarité, utilisé comme distracteur de la mémoire et du contenu du texte, enfin rappel, sur 15 min. des informations du texte conservées en mémoire dans un écrit qui ne devait pas nécessairement être un résumé.

Les élèves ayant bénéficié d'informations sur le contenu du texte (G2), et ceux qui ont lu le texte simplifié (G4) ont donné des informations exactes plus longues que ceux du groupe témoin. Les élèves (G3) ayant bénéficié d'informations renvoyant aux connaissances du monde évoquées par le texte, ont produit plus d'informations que ceux des autres groupes. Les notes donnant des informations supplémentaires ont permis de créer des liens de causalité entre les informations contenues dans le texte. Ces notes ont permis des rappels plus complets et plus articulés causalement que dans les autres conditions. L'explication des informations non explicites dans le texte ont aidé le lecteur à construire l'image mentale du texte. Ainsi, lire et acquérir des connaissances semblent des activités liées. Apprendre à lire et à comprendre ne

saurait constituer un apprentissage purement technique, détaché des contenus de connaissances véhiculés par les textes.

Marin et ses collègues (2007), au sein de l'équipe Cognition et Didactique du Texte, ont réalisé une recherche « *Hypertexte et aide à la lecture* » (Crinon, Legros, Marin & Ave, 2007). Les concepteurs ont créé et évalué le prototype d'un logiciel d'aide à la lecture de textes documentaires. Cet outil, *Coditex*<sup>43</sup>, comporte une base de textes ouverte et des aides hypertextuelles. L'équipe a proposé quelques textes avec plusieurs aides possibles apportées aux élèves : le même texte simplifié mais uniquement au niveau syntaxique, des explications supplémentaires, la signification de certains mots et le résumé des textes.

Par ailleurs, une étude pilote de simplification de textes sur les trois niveaux : lexical, syntaxique et discursif a montré des gains significatifs de la vitesse et de la qualité de la lecture ainsi que de la compréhension de textes sur un corpus de dix textes simplifiés, auprès de dix enfants dyslexiques francophones de 8 à 12 ans (Brunel & Combes, 2015). Les élèves ont lu à haute voix des textes standardisés, originaux et simplifiés manuellement, issus de *International Reading Speed Texts (IReST)*<sup>44</sup> (textes courts narratifs), puis ils ont répondu à des questions de compréhension. Gala et Ziegler (2016) ont montré que la simplification a été bénéfique en termes de vitesse de lecture par mot ( $F=4,105$ ,  $p = ,003$ ) et d'erreurs de lecture car celles-ci ont diminué sans perte de compréhension. Trois types d'erreurs ont été répertoriés : erreurs de lecture à proprement parler (mot répété, saut de ligne, ligne lue deux fois), erreurs au niveau lexical (mots incorrectement lus pour des raisons lexicales, le mot lu est un pseudo-mot ou un autre mot) et erreurs pour des raisons grammaticales avec le mot lu qui appartient à la même catégorie du discours mais présente une variation au niveau des morphèmes grammaticaux (nombre, genre, temps, mode ou personne). Les élèves ont commis moins d'erreurs dans les textes simplifiés que dans les textes originaux (5,5% contre 7,7%) avec des erreurs lexicales et grammaticales qui se sont produites aussi souvent l'une que l'autre. La complexité lexicale, pour les lecteurs dyslexiques en français, porte principalement sur les verbes et les noms plus longs que sept caractères ou présentant des lettres ayant des équivalents similaires (inversion de lettres 'miroir' p/q ou d/b) ou possédant des syllabes complexes (par exemple, CVCC) ou une relation graphème-phonème irrégulière (par exemple, « c » ou « g » dont la prononciation dépend du contexte). Afin d'éviter les erreurs de lecture, lors des remplacements lexicaux, des mots plus courts et plus fréquents et des mots ayant une régularité dans la relation graphème-phonème doivent être privilégiés.

---

<sup>43</sup> <http://coditexte.creteil.iufm.fr/index.htm>

<sup>44</sup> IReST, International Reading Speed Texts

À la suite de cette première étude, il y en a eu une deuxième, réalisée par Nandiegou et Reboul (2018) auprès de 21 enfants dyslexiques âgés de 9 à 12 ans et scolarisés en milieu ordinaire en cycle 3, qui montre les bénéfices de la simplification lexicale pour des enfants dyslexiques. Cette étude a été menée en utilisant un corpus, simplifié uniquement au niveau lexical, de cinq textes narratifs et cinq textes documentaires scientifiques. Chaque enfant, qui bénéficiait d'une prise en charge en orthophonie, a lu des textes originaux et des textes simplifiés, à haute voix avec un enregistrement afin d'évaluer la qualité de lecture. La compréhension a été évaluée avec cinq questions et trois propositions de réponses pour chacune. L'ensemble des analyses de variance (ANOVA) a été réalisé en intra-sujet. Des gains significatifs ont été relevés sur la vitesse. En outre, cet effet est plus important pour les textes documentaires. Concernant la qualité de la lecture le pourcentage moyen de mots mal lus dans les textes simplifiés a été significativement inférieur au pourcentage moyen de mots mal lus dans les textes originaux. Les gains observés au niveau de la compréhension des textes n'ont pas été significatifs. Dans des analyses complémentaires (Voir Annexe A) en considérant les textes originaux *vs* les textes simplifiés, Gala, Tack, Javourey-Drevet, François et Ziegler (2020b) ont montré que les mots modifiés lexicalement et les suppressions de certains mots ont permis une qualité de lecture plus performante avec moins d'erreurs lors de la lecture à haute voix. L'effet sur la diminution de la probabilité de mauvaise lecture du mot dans la version simplifiée est important (W de Kendall = 0,527). Par conséquent, nous constatons que les mots ciblés pour la simplification ont été non seulement bien choisis, mais aussi remplacés par des alternatives nettement plus faciles à lire, ce qui a permis d'améliorer la lisibilité des textes pour les 21 lecteurs dyslexiques.

➤ **Études auprès d'adultes, en anglais, langue étrangère et en espagnol**

Une autre étude pour l'anglais en tant que langue étrangère a été menée par Safari et Mohaghegh Montazeri (2017) auprès d'étudiants iraniens. Les auteurs ont comparé les performances en compréhension écrite de 3 groupes d'adultes comparables selon le *Test of English as a Foreign Language* (TOEFL 2005) en niveau de lecture. Chaque groupe a lu quatre textes et a bénéficié d'une version différente d'un corpus de texte : une version simplifiée lexicalement, une version simplifiée syntaxiquement et une version simplifiée lexicalement et syntaxiquement. C'est le groupe ayant bénéficié des simplifications lexicales et syntaxiques combinées qui obtient les meilleures performances de manière significative en comparaison des deux autres groupes. De plus, il n'existe pas de différence de performances entre le groupe ayant bénéficié de la simplification lexicale et celui ayant bénéficié de la simplification

syntactique. Ce serait donc une utilisation combinée des simplifications syntaxiques et lexicales qui serait la plus efficace pour la compréhension auprès de sujets adultes.

Enfin, dans l'étude de Rello, Baeza-Yates, Bott et Saggion (2013a) qui a été réalisée auprès de 96 adultes hispanophones, 47 personnes dyslexiques et 49 personnes constituant le groupe témoin, quatre conditions de lecture pour les participants ont été expérimentées. Chacun des participants n'est passé que dans une condition : groupe lisant les textes originaux, groupe bénéficiant d'une substitution lexicale, groupe bénéficiant de propositions de plusieurs synonymes, groupe bénéficiant d'une simplification lexicale manuelle. Les substitutions lexicales et les propositions de plusieurs synonymes ont été réalisées avec le logiciel LexSIS (Bott et al., 2012). Chaque participant a lu deux articles scientifiques et a répondu à un questionnaire à choix multiples (QCM) avec trois propositions de réponses. Les analyses ont été réalisées en intergroupes. Les deux groupes, dyslexiques ou groupe témoin, lisent plus rapidement avec le texte simplifié manuellement. Néanmoins, aucun effet significatif sur le temps de lecture n'a été montré dans le groupe des dyslexiques ou dans le groupe témoin pour aucune des conditions (texte original, texte avec substitutions lexicales, texte avec simplification lexicale manuelle). Au niveau de la compréhension, les participants atteints de dyslexie ont un pourcentage de bonnes réponses moins important que les participants du groupe témoin dans les quatre conditions énoncées au départ. Cependant, la différence de pourcentage de bonnes réponses entre les groupes n'était pas statistiquement significative. Aucun effet significatif de la simplification des textes sur le score de compréhension dans le groupe des dyslexiques ou dans le groupe témoin n'a été montré. Les auteurs ont suggéré qu'une des raisons possibles de ces résultats était que la compréhension du texte dépend de segments de textes plus longs, c'est-à-dire que la compréhension ne dépend pas de mots isolés mais des relations entre les mots. L'une des principales stratégies pour comprendre est de prendre en compte le contexte du mot. C'est sur la partie qualitative de cette étude que certains effets de la simplification ont été significatifs. Les facilités de lecture, de compréhension et de mémorisation du texte ont été évaluées par les participants sur une échelle de Likert à cinq points. Les participants atteints de dyslexie ont trouvé que les textes avec les propositions de synonymes étaient plus faciles à lire que les textes avec une simplification manuelle ( $p = ,034$ ) et que les textes originaux ( $p = ,015$ ). Les participants atteints de dyslexie ont trouvé que les textes avec les propositions de synonymes étaient plus faciles à comprendre que les textes originaux ( $p = ,001$ ) et que les textes avec les substitutions lexicales ( $p = ,013$ ). Pour le groupe témoin, les participants ont trouvé que les textes avec substitution lexicale étaient significativement plus difficiles à comprendre que les textes avec les trois autres conditions.

Aucun effet significatif des simplifications lexicales sur le temps de lecture et la compréhension n'a été relevé mais selon l'avis des participants dyslexiques les textes avec propositions de synonymes sont plus lisibles et compréhensibles.

### **Conclusion**

Nous avons présenté dans ce chapitre la simplification de textes et les différents niveaux linguistiques sur lesquels elle s'effectue : lexical, morphosyntaxique et discursif. Les travaux de recherche que nous avons présentés, relatifs à des adaptations ou simplifications, étaient adressés à différentes cohortes, d'enfants et d'adultes. Dans toutes les études, la simplification de textes permet d'obtenir un gain sur la fluidité de lecture davantage que sur la compréhension, même si des gains en compréhension ont été aussi relevés. Dans le chapitre suivant, nous présentons le développement du corpus simplifié de textes scolaires que nous avons utilisé dans notre travail expérimental et une comparaison des versions originales et simplifiées au travers d'une analyse des variables textuelles choisies.

## Chapitre 4

### Développement d'un corpus simplifié de textes scolaires

Dans le cadre de ce travail de thèse, nous avons créé un corpus de textes parallèles, versions originales vs simplifiées, afin de tester les effets de la simplification à différents niveaux linguistiques. Ce corpus a été pré-testé avant la phase expérimentale ; puis nous avons estimé la difficulté des textes grâce à des variables textuelles de lisibilité. Les méthodes d'analyse de la lisibilité se sont considérablement développées ces dernières années avec les avancées de la technologie et plusieurs critères sont nécessaires pour déterminer les textes appropriés pour des apprentis-lecteur (Benjamin, 2012). Nous avons réalisé des analyses textuelles pour évaluer le niveau de difficulté des textes utilisés en tenant compte de trois types de variables : lexicales, morphologiques et syntaxiques.

#### 4.1 Le corpus développé

Afin de réaliser l'ensemble des expériences présentées dans les chapitres 6 à 10 de ce manuscrit, nous avons simplifié des textes originaux. Des heuristiques ont été choisies afin d'élaborer un protocole de simplification et de création des questions.

##### 4.1.1 Choix des textes et création des questions

Les textes sont des extraits prélevés sur des sites web avec des contes<sup>45</sup> ou documentaires<sup>46</sup>, des morceaux choisis de revues (*BTJ*<sup>47</sup>, *WAPITI*<sup>48</sup> par exemple), et des ouvrages (*Chichois de la rue des Mauvestis*<sup>49</sup>, Ciravégna (1995)). Ces textes sont à visée scientifique ou de type 'littérature jeunesse', destinés aux lecteurs de cette tranche d'âge. Le texte original donné à lire aux élèves ne faisait pas référence à des images ou autres supports d'aide à la compréhension souvent présents dans les ouvrages jeunesse (schémas, dessins, photos ou graphiques ...). Les sujets des différents textes proposés n'étaient pas liés afin qu'un texte mal compris par l'élève ne le pénalise pas lors de la lecture du texte suivant. Pour évaluer

---

<sup>45</sup> Conte burkinabé : [http://www.contesafricains.com/article.php3?id\\_article=28&Valider=Afficher+le+conte](http://www.contesafricains.com/article.php3?id_article=28&Valider=Afficher+le+conte)  
à découvrir sur <http://www.contesafricains.com>

<sup>46</sup> <http://kidiscience.cafe-sciences.org/articles/comment-les-huitres-fabriquent-elles-des-perles/>

<sup>47</sup> <https://www.icem-vente-en-ligne.org/btj>

<sup>48</sup> <https://www.wapiti-magazine.com/magazine>

<sup>49</sup> <https://www.babelio.com/livres/Ciravegna-Chichois-de-la-rue-des-Mauvestis/459091>

la compréhension, plusieurs tâches existent mais nous avons choisi d'utiliser des questionnaires à choix multiples (QCM).

Au niveau de la création des questions de compréhension, nous avons une contrainte majeure qui était de créer des questions de compréhension qui suivent le même protocole et qui évaluent la lecture des deux types de textes (Figure 16). Nous nous sommes imposé plusieurs règles : les cinq questions portent sur l'ensemble du texte en suivant l'ordre chronologique d'apparition des informations ou évènements dans le texte ; elles portent sur les éléments essentiels à mémoriser pour construire le sens du texte ; les questions font appel à la compréhension globale et aux inférences du texte. Les propositions de réponses faites pour chaque question des QCM sont toutes vraisemblables et en relation avec le sujet du texte.

Figure 16 : Exemples de questions à choix multiples avec les propositions de réponses pour chacune. La réponse correcte est présentée en bleu.

**QCM Littéraire**

**Que se passe-t-il lorsque les garçons se rencontrent ?**

- Ils s'arrêtent pour discuter de la rentrée et dire ce qu'ils en pensent
- Ils s'arrêtent pour raconter leurs vacances et parler des jeux de cour
- *Ils se poussent, tombent et chacun repart avec le cartable de l'autre*
- Ils se montrent leurs nouveaux cartables et leurs affaires neuves

**QCM Documentaire scientifique**

**Qu'est-ce que les oiseaux d'aujourd'hui ont perdu qui existait chez les dinosaures ?**

- Les pattes, ils ne peuvent que voler
- *Les dents et la longue queue de lézard*
- Les dents uniquement
- La longue queue uniquement

#### 4.1.2 Pré-tests

Avant de proposer les textes dans la phase expérimentale sur le terrain de recherche, nous avons choisi de réaliser des prétests afin d'évaluer le choix des textes et leur simplification ainsi que les QCM créés. Les pré-tests ont été réalisés dans des circonscriptions du Var proches géographiquement du terrain de la recherche. Chaque année les pré-tests qui ont évalué le niveau des textes ont eu lieu sur janvier et février afin de pouvoir rendre les textes accessibles sur tablette pour le troisième trimestre. Le corpus, comprenant chaque année 10 documentaires scientifiques et 10 textes littéraires, a été évalué dans sa totalité bien que la demande auprès des

professeurs ait été de faire lire aux élèves deux textes littéraires et deux documentaires scientifiques au minimum. Dans cette phase de pré-tests, ils ont été lus dans cinq classes de cinq écoles différentes, sur les communes de Saint-Mandrier, La Seyne, Ollioules. Dans les faits, cinq ou six textes de chaque catégorie ont été choisis et lus par chaque classe, sur la base de leur titre, thème en lien avec la vie de la classe, les apprentissages en cours. Dans quatre classes sur cinq, le public appartenait à un milieu socio-économique sensiblement identique à celui des écoles où les expérimentations principales se sont déroulées. Le contexte familial des élèves relevait de la classe moyenne. La 5<sup>ème</sup> classe était située dans une école classée en REP<sup>+</sup>, sur la commune de La Seyne sur Mer.

Le protocole est resté identique dans chaque classe : lecture silencieuse du texte format papier puis le texte est retourné pour passer à l'étape suivante, QCM en format papier aussi. Les professeurs avaient pour consigne de ne pas avoir travaillé le thème abordé en amont pour ne pas biaiser l'évaluation de la compréhension du texte. Les textes simplifiés ont été donnés à lire aux élèves choisis par les professeurs. Les élèves ne savaient pas quel format était le leur (ORIG ou SIMP). Les consignes étaient identiques à celles utilisées par la suite, c'est-à-dire que l'enseignant ne pouvait pas aider les élèves, intervenir lors des lectures, ni dans la partie QCM. Les modalités d'évaluation étaient connues des élèves. Absence de réponse : 0 étoile ; réponse erronée : 1 étoile ; réponse incorrecte mais plausible car contenant des éléments du texte : 2 étoiles ; réponse correcte : 3 étoiles. La correction des QCM s'est faite en différé en classe entière. Pour suivre leur progression, après la correction de leurs réponses au QCM par l'adulte et la correction collective, les élèves renseignaient un tableau avec le nombre d'étoiles obtenues. Les pré-tests ont permis de retravailler certaines simplifications des textes mais aussi de modifier certaines formulations dans les QCM. Nous avons laissé les enseignants libres quant à la correction et au travail possible sur l'analyse de la langue après la lecture des textes.

### 4.1.3 Le corpus

Nous avons créé, au fil des trois années d'expérimentation un corpus parallèle (original vs simplifié) de 69 textes qui ont été simplifiés principalement sur trois niveaux : lexical, syntaxique et discursif. Parmi ces 69 textes originaux, 5 ont été simplifiés sur un seul niveau : lexical ou syntaxique ou lexico-syntaxique en CE1. Pour simplifier manuellement ce corpus, nous avons utilisé principalement trois outils en langue française. Tout d'abord l'outil *Résyf*<sup>50</sup> (Billami, François & Gala, 2018 ; Gala & Javourey-Drevet, 2020c), qui est une ressource

---

<sup>50</sup> <https://cental.uclouvain.be/resyf/>

lexicale de synonymes monolingues classés selon leur difficulté à être lus et compris par les apprenants natifs du français. Les synonymes proviennent d'un réseau lexical de synonymes existant, créé grâce un jeu sérieux appelé *JeuxdeMots*, (Lafourcade, 2007)<sup>51</sup>, et ont été sémantiquement désambiguïsés (François et al., 2016). L'ordonnancement des synonymes dans Resyf (Billami, François & Gala, 2018) se fait par le biais d'un algorithme qui prend en compte des variables linguistiques telles que la fréquence des mots tirée de Manulex (Lété et al. 2004), la longueur des mots, la polysémie, la structure syllabique, la régularité phonème-graphème, etc. Le système d'ordonnancement a été validé par 40 annotateurs avec un taux d'accord inter-annotateur fort (Kappa de Cohen) de  $k = .63$  (François, Billami, Gala & Bernhard, 2016). Une autre ressource utilisée pour le choix de synonymes est le Dictionnaire Électronique des Synonymes (DES), provenant du Centre de Recherches Interlangues sur la Signification en Contexte<sup>52</sup>(CRISCO), qui propose des relations synonymiques ou antonymiques pour les mots recherchés (Morel & François, 2015). Cette ressource a été créée à partir de sept dictionnaires dans les années 90 et mise à jour depuis. Enfin l'outil Manulex<sup>53</sup> (Lété et al. 2004) propose des listes de fréquences de mots en fonction de leur catégorie sur trois niveaux CP, CE1 et CE2-CM2. Les fréquences présentées ont été calculées sur la base de 54 manuels scolaires.

Le corpus constitué pour l'expérimentation a été inclus dans le corpus ALECTOR (Gala et al., 2020b). Grâce au travail d'équipe réalisé au sein du projet ANR ALECTOR<sup>54</sup>, un protocole a été défini sur la base des opérations réalisées avec des recommandations pour des transformations de textes, afin d'améliorer leur lisibilité et leur compréhension pour des enfants dyslexiques et faibles lecteurs (Gala, Todirascu Javourey-Drevet, Bernhard, Wilkens & Meyer, 2020d). Ce travail, présenté en Annexe B, résume des règles simples pour que d'autres textes puissent être simplifiés.

## 4.2 Calculer la difficulté d'un texte

La notion de difficulté est ardue à définir, car elle est multidimensionnelle ; elle dépend de l'interaction de nombreux facteurs, notamment sociaux (situation de lecture), individuels (lecteur) et linguistiques (texte). Le champ de la lisibilité se concentre sur les caractéristiques du texte et non pas celles du lecteur (Boyer, 1992). La lisibilité d'un texte désigne le degré de facilité avec laquelle celui-ci peut être lu. Il convient d'ajouter que le terme lisibilité ne porte

---

<sup>51</sup> <http://www.jeuxdemots.org/jdm-accueil.php>

<sup>52</sup> <https://hal-ujm.archives-ouvertes.fr/CRISCO/medihal-01823024v1>

<sup>53</sup> <http://www.manulex.org/fr/home.html>

<sup>54</sup> <https://alectorsite.wordpress.com/equipas/>

pas sur l'ensemble des propriétés d'un texte, mais uniquement sur les aspects linguistiques et conceptuels. L'effet de la mise en page, les caractéristiques typographiques telles que la taille ou la couleur de la police, de la couleur du fond ou encore la présence de graphiques ne sont pas prises en compte mais relèvent d'un domaine, dénommé « lisibilité typographique » par Richaudeau (1969). Suite à la création des textes simplifiés chaque année, nous avons choisi d'évaluer le niveau de difficulté de chacune des versions puisque la simplification a pour objectif de rendre le texte plus lisible.

#### **4.2.1 Formules de lisibilité**

Lively et Pressley (1923) ont conçu ce qui est généralement reconnu comme la première formule de lisibilité, laquelle ne prenait en compte que le vocabulaire, en s'appuyant sur une liste de fréquences lexicales compilée par Thorndike (1921). Elle constitue donc une perspective assez réduite des difficultés que rencontre un lecteur confronté à un document écrit, puisque seules les caractéristiques du texte sont prises en compte comme facteurs explicatifs de cette difficulté. La formule de Flesch (1948), prenant en considération le nombre moyen de syllabes par 100 mots et la longueur moyenne des phrases en nombre de mots, a été développée pour l'anglais. Elle ne considère qu'un nombre très restreint de caractéristiques textuelles, à savoir deux facteurs explicatifs. Les premières formules de lisibilité pour le français sont celles de Kandel et Moles (1958) et celles de De Landsheere (1963), qui constituent toutes deux une adaptation de la formule de Flesch (1948).

La notion de lisibilité a évolué au cours du 20<sup>e</sup> siècle : des seuls aspects lexicaux et syntaxiques, elle englobe aujourd'hui également les aspects de cohésion et de cohérence (Mesnager, 2002). Les formules de lisibilité, qui ont fait l'objet de nombreuses critiques dans les années 1980 et 1990, ont récemment regagné une certaine popularité, comme en témoignent des numéros de revues dédiées, notamment dans des secteurs tels que le traitement automatique des langues (François & Bernhard, 2014). L'approche de la lisibilité via l'intelligence artificielle (IA) a relancé le domaine, ouvrant de nouvelles pistes de recherche au travers d'algorithmes utilisés pour paramétrer des dimensions plus complexes de texte ou pour agencer plus clairement des caractéristiques linguistiques dans un modèle de lisibilité (François & Bernhard, 2014). Les analyses de lisibilité basées sur des connaissances en psychologie cognitive (Kintsch, 1988), qui ont pris davantage en compte la cohérence et les relations d'inférences dans le texte, ont inspiré la création d'outils, par exemple Coh-Metrix (Graesser et al., 2004) et Word Net (Fellbaum, 1998), en langue anglaise. Ce sont des descripteurs multi-dimensionnels de la langue qui appréhendent au mieux la complexité des textes. Actuellement

le domaine de la lisibilité évolue encore, notamment avec son utilisation dans le champ de l'éducation. La prise en compte du niveau du lecteur, adulte ou apprenti-lecteur, pose des questions, celles de savoir si la méthode de lisibilité est adaptée au public de lecteurs et aux textes, celle de sa validité prédictive notamment au niveau de la compréhension pour les lecteurs. Évaluer la difficulté d'un texte pour des apprentis-lecteurs représente un nouveau défi dans le domaine de la lisibilité (Benjamin, 2012).

Sheehan, Flor et Napolitano (2013) ont ainsi analysé plus en détail les différences entre les textes littéraires et informatifs et ont remarqué que les formules de lisibilité ont tendance à surestimer la difficulté des textes informatifs et à la sous-estimer pour les textes littéraires. Cela est dû au fait que les textes littéraires incluent plus de vocabulaire de base de la langue, tandis que les documents informatifs contiennent un contenu plus spécifique (terminologie de spécialité). D'autres études (Dell'Orletta, Montemagni & Venturi, 2014 ; Vajjala & Meurers, 2014) ont confirmé l'effet du genre textuel sur les performances de prédiction de lisibilité. Il existe un besoin de formules spécialisées, à la fois pour des genres de textes spécifiques et pour des populations de lecteurs spécifiques. La raison principale qui empêche la construction de modèles plus spécialisés est, selon eux, le manque de corpus appropriés.

On comprend que la lisibilité du texte dépend de diverses variables. Alors que les facteurs lexico-sémantiques et syntaxiques ont été largement utilisés dans la littérature, des propriétés discursives et cognitives de plus haut niveau telles que la cohésion et la cohérence ont reçu moins d'attention. Todirascu, François, Gala, Fairon, Ligozat et Bernhard (2013) ont évalué l'efficacité de 41 mesures de cohésion et de cohérence de texte en tant que prédicteurs de la lisibilité du texte. Ils comparent les résultats obtenus manuellement sur deux corpus comprenant des textes avec différents niveaux de difficulté et montrent que certaines caractéristiques cohésives sont en effet des prédicteurs utiles.

Par ailleurs, il existe quelques outils en ligne qui mesurent la lisibilité d'un texte et qui donnent une analyse en fonction d'éléments lexicaux ou grammaticaux. Nous pouvons mentionner TextBalancer 1.5.0<sup>55</sup> créé par Spiezia et Annunziata, destiné à des apprenants de français italianophones. Cet outil permet d'avoir une analyse pour les textes en français, mais uniquement basée sur la formule de Kandel et Moles donc les deux variables de Flesch. Il en est de même pour l'outil du Centre d'Investigation Clinique Necker Cochin<sup>56</sup>. D'autres outils utilisent des variables plus fines pour calculer la lisibilité, par exemple pour les textes administratifs : AMesure (François, Müller, Degryse & Fairon, 2018) du Centre pour le

---

<sup>55</sup> <http://textbalancer.it/>

<sup>56</sup> [https://www.recherchecliniquepariscentre.fr/?page\\_id=3169](https://www.recherchecliniquepariscentre.fr/?page_id=3169)

Traitement Automatique du Langage (CENTAL) à Louvain-la-Neuve<sup>57</sup>. Toutefois, il existe un nombre limité d'outils permettant de calculer automatiquement la lisibilité ou la complexité des textes pour le français (Dascalu et al., 2013 ; François & Fairon, 2012).

#### **4.2.2 Variables textuelles choisies pour analyser la lisibilité**

Pour mieux décrire en quoi nos versions simplifiées différaient des versions originales du même texte, nous avons utilisé certaines des variables standard qui sont généralement utilisées dans les formules de lisibilité (par exemple, Flesch, 1948 ; Kandel & Moles, 1958 ; pour une revue, voir François, 2015). Nous avons commencé avec un ensemble de 406 variables développées par François (2011). Un grand nombre de ces variables sont des variantes les unes des autres et mesurent donc des propriétés de textes similaires (par exemple, diverses mesures de la fréquence des mots). Parmi toutes ces variables, nous avons d'abord sélectionné les plus pertinentes et nous avons calculé leurs valeurs pour chacun de nos textes simplifiés et originaux, et ce pendant les trois années où nous avons mené nos expériences de terrain.

#### **4.2.3 Analyse de la lisibilité du corpus avec les variables choisies**

Nous avons sélectionné chaque année les variables les plus discriminantes entre les deux versions : originale et simplifiée pour les 60 textes originaux (10 littéraires et 10 documentaires scientifiques pour chaque année). Cette procédure a abouti à une sélection de 22 variables qui peuvent être organisées en trois familles : lexicale, morphologique et syntaxique. Nous avons ensuite utilisé un test de Wilcoxon apparié pour évaluer si une variable différenciait significativement un texte simplifié d'un texte original. Nous avons ensuite vérifié les distributions de toutes les variables significatives. Les variables sont décrites ci-dessous et leurs propriétés statistiques pour les textes simplifiés et originaux se trouvent dans le Tableau 2.

---

<sup>57</sup> <https://cental.uclouvain.be/amesure/>

Tableau 2 : Variables de lisibilité sélectionnées pour les versions simplifiées et originales des textes, sur les trois années d'expérimentation. Ecart types entre parenthèses.

<b>CE1- CE2 - CM1</b>	<b>Textes simplifiés</b>	<b>Textes originaux</b>	<b>Sig.</b>
<b>Variables lexicales</b>			
75 <sup>ème</sup> percentile de la distribution des fréquences des lemmes			
CE1	183,9 (176,2)	119,4 (96,8)	,000***
CE2	120,2 (80,9)	87,8 (56,5)	,000***
CM1	83,97 (62,9)	65,3 (52,9)	,000***
90 <sup>ème</sup> percentile de la distribution des fréquences des lemmes			
CE1	18,3 (11,7)	11,8 (8,0)	,000***
CE2	13,6 (9,6)	8,3 (6,1)	,000***
CM1	9,5 (6,6)	7,1 (5,5)	,000***
Fréquence moyenne des adjectifs (lemme)			
CE1	1592,2 (1013,4)	1122,2 (808,5)	,001**
CE2	1309,0 (940,0)	949,8 (669,2)	,006**
CM1	1489,1 (829,0)	982,8 (684,9)	,000***
Médiane de la distribution des fréquences des adjectifs (lemmes)			
CE1	479,3 (474,1)	367,0 (445,8)	,023*
CE2	455,5 (424,0)	315,4 (355,5)	,007**
CM1	251,2 (241,9)	188,0 (190,9)	,004**
Proportion de mots de plus de 8 lettres			
CE1	0,058 (0,031)	0,073 (0,030)	,003**
CE2	0,064 (0,024)	0,091 (0,026)	,000***
CM1	0,068 (0,033)	0,087 (0,039)	,000***

Proportion de mots absents de la liste de 2 000 mots de Gougenheim			
CE1	0,150 (0,051)	0,172 (0,045)	,001**
CE2	0,160 (0,050)	0,182 (0,045)	,000***
CM1	0,192 (0,049)	0,210 (0,053)	,001**
<b>Variables morphologiques</b>			
Proportion de verbes dans la forme future			
CE1	0,015 (0,030)	0,018 (0,023)	,033*
CE2	0,006 (0,009)	0,008 (0,011)	,052
CM1	0,014 (0,028)	0,011 (0,029)	,038*
Proportion de verbes dans la forme du participe passé			
CE1	0,093 (0,054)	0,124 (0,065)	,007**
CE2	0,115 (0,059)	0,131 (0,070)	,028*
CM1	0,175 (0,098)	0,182 (0,101)	,550
Proportion de verbes au présent de l'indicatif			
CE1	0,425 (0,246)	0,387 (0,242)	,006**
CE2	0,462 (0,263)	0,430 (0,252)	,001**
CM1	0,467 (0,232)	0,448 (0,226)	,033*
Proportion de verbes dans la forme du participe présent			
CE1	0,012 (0,015)	0,028 (0,020)	,002**
CE2	0,015 (0,023)	0,026 (0,022)	,006**
CM1	0,027 (0,027)	0,033 (0,035)	,215
<b>Variables syntaxiques</b>			
75 <sup>ème</sup> percentile de la distribution du nombre de mots par phrase			
CE1	16,8 (3,8)	20,9 (6,6)	,000***
CE2	17,4 (4,2)	20,3 (5,0)	,000***
CM1	16,9 (2,8)	20,8 (4,3)	,000***

90 <sup>ème</sup> percentile de la distribution du nombre de mots par phrase			
CE1	20,4 (4,1)	27,2 (11,3)	,000***
CE2	21,9 (4,6)	24,9 (5,5)	,000***
CM1	21,6 (2,6)	26,4 (4,6)	,000***
Nombre médian de mots par phrase			
CE1	11,9 (2,4)	15,2 (5,4)	,001**
CE2	12,3 (3,4)	15,0 (4,8)	,000***
CM1	12,9 (2,5)	14,9 (3,9)	,000***
Nombre moyen de mots par phrase			
CE1	13,2 (2,5)	16,9 (6,3)	,000***
CE2	13,4 (3,0)	15,7 (4,2)	,000***
CM1	13,8 (2,3)	16,2 (3,3)	,000***
Rapport entre les prépositions et l'ensemble des articles			
CE1	0,762 (0,220)	0,809 (0,204)	,048*
CE2	0,775 (0,242)	0,803 (0,187)	,062
CM1	0,804 (0,248)	0,856 (0,268)	,005**
Proportion de pronoms dans le texte			
CE1	0,111 (0,046)	0,121 (0,044)	,003**
CE2	0,123 (0,048)	0,123 (0,047)	,823
CM1	0,103 (0,040)	0,110 (0,042)	,001**
Rapport des articles par rapport à l'ensemble des mots grammaticaux			
CE1	0,360 (0,085)	0,342 (0,082)	,007**
CE2	0,353 (0,101)	0,346 (0,092)	,198
CM1	0,367 (0,073)	0,346 (0,063)	,000***
Rapport entre les noms propres et tous les mots lexicaux			
CE1	0,032 (0,024)	0,027 (0,024)	,004**
CE2	0,014 (0,025)	0,013 (0,023)	,041*
CM1	0,019 (0,024)	0,018 (0,022)	,005**

Rapport entre les noms propres et les adjectifs			
CE1	0,214 (0,192)	0,184 (0,193)	,001**
CE2	0,098 (0,175)	0,083 (0,133)	,047*
CM1	0,127 (0,156)	0,122 (0,151)	,018*
Rapport entre les noms propres et les pronoms			
CE1	0,173 (0,153)	0,129 (0,124)	,000***
CE2	0,069 (0,110)	0,067 (0,103)	,011*
CM1	0,110 (0,140)	0,097 (0,122)	,002**
Rapport entre les pronoms et les mots lexicaux			
CE1	0,195 (0,083)	0,214 (0,082)	,001**
CE2	0,220 (0,085)	0,221 (0,085)	,550
CM1	0,181 (0,075)	0,195 (0,081)	,001**
Rapport entre les pronoms et les noms communs			
CE1	0,555 (0,343)	0,604 (0,344)	,004**
CE2	0,584 (0,334)	0,595 (0,323)	,376
CM1	0,449 (0,230)	0,493 (0,254)	,000***

#### 4.2.4 Discussion pour donner suite aux analyses textuelles

##### Variables lexicales

Le premier ensemble de six variables reflète les propriétés lexicales liées à la fréquence et à la longueur des mots. Les variables de fréquence ont été calculées à l'aide de Lexique 3 (New, Pallier, Ferrand & Matos, 2001). Nous avons trouvé des différences significatives pour la fréquence moyenne des adjectifs. Ils ont souvent été remplacés par des synonymes plus courts et plus fréquents. De plus, les textes simplifiés contiennent moins de mots que les originaux, avec par exemple l'élimination des adverbes et donc la fréquence des adjectifs est plus importante dans les textes simplifiés que dans les originaux. La fréquence des 75<sup>ème</sup> et 90<sup>ème</sup> percentiles de la distribution de fréquences des lemmes augmente dans la version simplifiée. Cela s'explique par le fait que plusieurs mots considérés comme compliqués, donc moins fréquents, ont été remplacés par des mots plus fréquents. Par conséquent, la fréquence des lemmes correspondant au 75<sup>ème</sup> et 90<sup>ème</sup> percentile va logiquement avoir tendance à augmenter. Les textes simplifiés contenaient également moins de mots qui étaient absents d'une liste de référence des mots français les plus fréquents (Gougenheim, 1964). En termes de longueur des mots, les textes simplifiés avaient une proportion significativement plus faible de mots longs (nous avons utilisé > 8 lettres, mais nous avons eu des effets significatifs également pour > 7 et > 9 lettres).

##### Variables morphologiques

Nous avons sélectionné quatre variables qui représentent la flexion des verbes par rapport aux temps verbaux et aux modes. La présence d'un morphème marquant le présent de l'indicatif est une variable significative pour les textes utilisés, les autres variables ont, quant à elles, une significativité différente selon les années.

Dans les textes utilisés la proportion de verbes à la forme simple du présent a augmenté dans les textes simplifiés. En revanche, pour le futur, on ne peut pas généraliser car cette variable n'est significative que pour les textes utilisés au CE1 et CM1. La proportion du morphème de futur entre les textes originaux et les simplifiés n'évolue pas dans le même sens. La proportion de morphèmes désignant des participes présents et de participes passés a baissé significativement pour les textes simplifiés les deux premières années, CE1 et CE2. Pour les textes du CM1, la baisse des participes dans la forme simplifiée par rapport à la forme originale est visible mais n'est pas significative.

## Variables syntaxiques

Nous avons retenu 12 variables syntaxiques. En termes de difficulté syntaxique, nous avons sélectionné un certain nombre de variables liées au nombre de mots par phrase, telles que la moyenne, la médiane et les 75<sup>ème</sup> ou 90<sup>ème</sup> percentiles. Nous constatons pour ces quatre variables une diminution significative entre le texte original et le simplifié sur les trois années. Nous avons également constaté que huit ratios des parties de discours distinguaient les textes simplifiés des textes originaux. Les ratios des parties du discours sont souvent pris comme une approximation de la complexité syntaxique des phrases (Bormuth, 1966 ; Crossley, Louwerse, McCarthy & McNamara, 2007). Suite aux analyses, plusieurs modifications ont été relevées. Le nombre de pronoms est significativement moins important dans les textes simplifiés que dans les originaux car ils ont été remplacés par des noms communs ou des noms propres :

- La proportion de pronoms dans le texte est plus importante dans l'original que dans le simplifié.
- Le nombre de pronoms par rapport à l'ensemble des mots lexicaux (*i.e.* noms, verbes, adverbess et adjectifs) est plus important dans l'original que dans le simplifié.
- Le nombre de pronoms par rapport aux noms communs est plus important dans l'original que dans le simplifié.
- Le nombre de noms propres par rapport aux pronoms est moins important dans l'original que dans le simplifié et est significatif sur les trois années.

Le nombre de noms propres par rapport à l'ensemble des mots lexicaux et des adjectifs a augmenté dans les textes simplifiés :

- Le nombre de noms propres par rapport aux adjectifs est moins important dans l'original que dans le simplifié et est significatif sur les trois années
- Le nombre de noms propres par rapport à l'ensemble des mots lexicaux est moins important dans l'original que dans le simplifié et est significatif sur les trois années

L'augmentation des noms communs de la version simplifiée explique l'augmentation du nombre d'articles par rapport à l'ensemble des mots grammaticaux. La diminution du nombre de prépositions par rapport à l'ensemble des articles s'explique par l'augmentation de la part des articles dans le texte simplifié :

- Le nombre de prépositions par rapport à l'ensemble des articles est plus important dans l'original que dans le simplifié.
- Le nombre d'articles par rapport à l'ensemble des mots grammaticaux est moins important dans l'original que dans le simplifié.

## Chapitre 4

### Développement d'un corpus simplifié de textes scolaires

Dans ce chapitre nous avons présenté le corpus utilisé dans le cadre de nos expérimentations. Il s'agit d'un corpus de 69 textes du domaine littéraire et scientifique adressé à des enfants de 7 à 9 ans. Nous avons abordé les caractéristiques de ce corpus en termes de simplifications mises en œuvre. Nous avons aussi décrit les questionnaires de compréhension et la méthodologie pour les créer. Enfin, nous avons abordé la question de la lisibilité de textes et l'impact de différentes variables linguistiques pour identifier les niveaux de lisibilité. Dans le chapitre suivant, nous présentons les questions spécifiques des trois études que comporte la partie empirique.

## Chapitre 5

### Objectifs des expérimentations

Nous avons vu dans le chapitre 2 que comprendre un texte nécessite de rapidement aller de la reconnaissance des mots à l'élaboration d'une représentation, en s'appuyant sur des analyses linguistiques et des idées venant du texte, en lien avec les connaissances du monde. Malgré la place importante de la compréhension écrite dans notre société et les programmes scolaires, les résultats des évaluations nationales et internationales (chapitre 1) ont mis en évidence que le niveau de compréhension de textes des enfants français est préoccupant. La simplification de textes (chapitre 3) peut offrir un outil pour améliorer la compréhension de textes afin de faire entrer les élèves en difficulté dans le cercle vertueux de l'auto-apprentissage (Share, 1995 ; Ziegler et al., 2014). La simplification de textes s'opère à différents niveaux linguistiques : lexical, morphosyntaxique et discursif. Son but est de rendre les textes plus lisibles tout en préservant leur contenu original (Saggion, 2017). La simplification de textes semble donc une piste intéressante pour un public scolaire classique dont certains élèves peuvent rencontrer des difficultés tels que les faibles lecteurs, les faibles 'compreneurs' ou les dyslexiques identifiés ou non. La simplification de textes pourrait potentiellement restaurer un plaisir à la lecture qui encouragerait une exposition aux textes écrits, laquelle est facteur de progression.

L'objectif de cette thèse était de répondre à la problématique suivante : « **Est-ce que la simplification de textes littéraires et de documentaires scientifiques peut améliorer la fluidité et la compréhension en lecture chez des élèves français entre 7 et 9 ans ?** ». Ces lecteurs sont dans une phase décisive d'acquisition des processus de reconnaissance des mots écrits qui est centrale pour la compréhension (voir chapitre 2).

Nous avons réalisé trois grandes études qui répondent à cinq questions de recherche.

1. La première question est de savoir si la simplification de textes peut améliorer la fluidité en lecture et la compréhension de deux types de textes pour une cohorte d'élèves ayant été suivis sur trois ans. Nous avons fait l'hypothèse que la simplification serait efficace pour la fluidité de lecture mais aussi pour la compréhension. Cette question est abordée dans le chapitre 6.

2. La deuxième question vise à déterminer si la simplification est bénéfique pour tous les élèves ou si le gain est plus important pour certains lecteurs, selon leurs compétences initiales. Nous avons fait l'hypothèse que la simplification devrait être davantage bénéfique pour les élèves ayant un faible niveau de lecture (Gentaz et al., 2015). Cette question est abordée dans le chapitre 7.

3. La troisième question cherche à analyser les facteurs textuels qui produisent les gains de vitesse et de compréhension chez les lecteurs. Nous avons fait l'hypothèse que les facteurs de longueur de mots, de fréquence, de suppression de pronominalisation et de morphologie verbale seraient les facteurs les plus discriminants entre les versions originales et simplifiées parce que ce sont les variables qui ont le poids le plus important dans les formules de lisibilité (chapitre 4). Cette question est abordée dans le chapitre 8.

4. La quatrième question est d'étudier si un niveau de simplification est plus efficace que les autres, notamment le lexical versus le syntaxique. Nous avons fait l'hypothèse que le niveau lexical serait aussi important que le niveau syntaxique et que la simplification lexico-syntaxique serait la plus efficace (Safari & Mohaghegh Montazeri, 2017). Cette question est abordée dans le chapitre 9.

5. La cinquième question est de savoir si un élève, notamment s'il est faible lecteur, pourrait augmenter durablement son niveau de fluidité ou de compréhension de lecture après un entraînement de trois mois avec des lectures de textes simplifiés ou originaux. L'hypothèse sous-jacente était que la lecture de textes simplifiés permettrait à des faibles lecteurs de mieux entrer dans le cercle vertueux de l'auto-apprentissage (Share, 1995) et par ce biais d'augmenter leur capacité de compréhension de manière générale. Cette question est abordée dans le chapitre 10.

## Chapitre 6

### Effets de la simplification sur la fluidité et la compréhension : suivi de cohorte sur trois années

Dans ce chapitre, nous avons analysé l'effet de la simplification de textes littéraires et de documentaires scientifiques, sur trois niveaux : lexical, syntaxique et discursif. La cohorte de 138 élèves a été suivie du CE1 au CM1. Les enfants ont lu des textes originaux et simplifiés en classe sur iPad et nous avons mesuré grâce à un outil informatisé leur vitesse de lecture et la compréhension de chaque texte. Dans ce chapitre, nous allons tout d'abord présenter le terrain de recherche, puis le protocole expérimental, le traitement des données, les résultats en fluence par mots, les résultats en compréhension et enfin nous allons discuter des résultats.

#### 6.1 Terrain de recherche et participants

Nous avons suivi une cohorte d'élèves sur trois ans du CE1 au CM1 de quatre écoles sur trois communes, Sanary sur Mer, La Cadière d'Azur, St Cyr sur Mer. Les trois communes appartiennent au département du Var. Sanary-sur-Mer, 17070 habitants, 29000 euros en moyenne de revenus annuels par ménage, compte un collège, trois écoles élémentaires et trois maternelles. La Cadière-d'Azur, 5260 habitants, 37500 euros en moyenne de revenus annuels par ménage compte une école maternelle et une élémentaire. Saint-Cyr-sur-Mer, 11830 habitants, 32700 euros en moyenne de revenus annuels par ménage, compte deux collèges, trois écoles élémentaires et trois écoles maternelles.

Nous avons 165 élèves la première année (81 filles, 84 garçons), mais suite aux mouvements des élèves sur chaque année scolaire, les analyses des données de la cohorte suivie ont été réalisées avec 138 élèves (65 filles et 73 garçons) présents sur les trois années. L'échantillon ne comptait pas de primo-arrivant et comptait trois enfants bilingues (portugais, arabe et bulgare) qui ont participé à l'expérimentation. Tous les parents des élèves ayant participé aux expériences ont signé une autorisation parentale pour les tests en individuel ainsi qu'une autorisation pour captation de l'image et de la voix dans le cadre des captures vidéo réalisées en classe. Parmi les 22 professeurs des écoles volontaires pour l'expérimentation, certains ont participé plusieurs années, à la suite de changements de niveau ou parce qu'ils avaient des cours double niveau.

## 6.2 Protocole expérimental conduit sur trois ans

Afin de tester l'hypothèse selon laquelle la simplification de textes améliore la fluidité et la compréhension de textes, nous avons opté pour une méthode expérimentale, qui permet d'étudier les effets de la simplification et de la nature de textes en intra-sujet, à savoir que chaque sujet a « vu » toutes les conditions (original vs simplifié et littéraire vs documentaire). Les élèves ont lu chaque année, du CE1 au CM1, dix textes littéraires (LIT) et dix textes documentaires scientifiques (SCI) en version originale ou en version simplifiée.

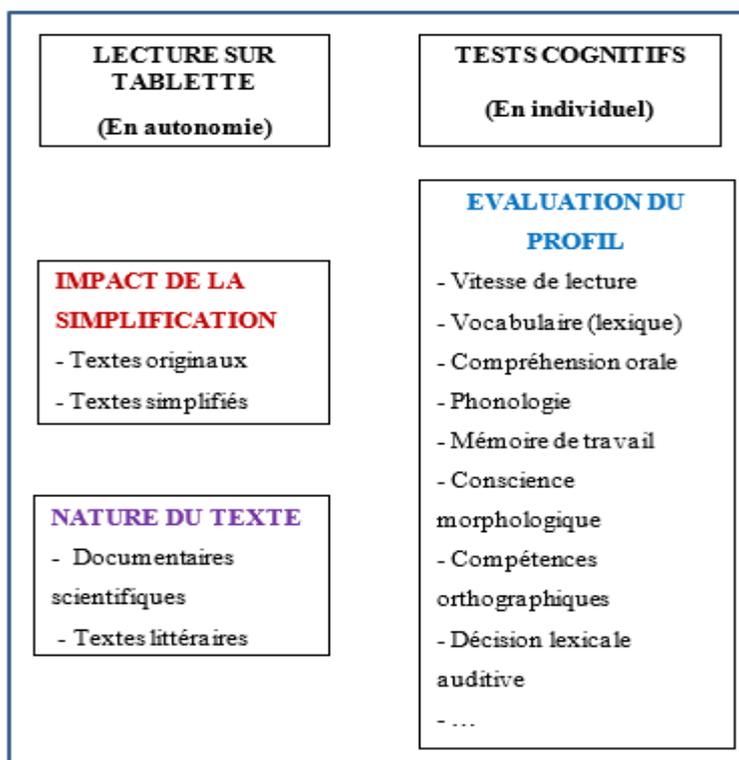
### Design expérimental

L'ensemble de l'expérimentation a été réalisé en situation classes sauf les tests en individuel qui ont été effectués au sein de l'école mais dans un local distinct de la salle de classe. L'étude comprenait deux phases (Figure17) :

- la phase de tests individuels composée de 4 sessions de 30 minutes la première année, puis 3 séances de 30 minutes la seconde année et enfin deux séances d'environ 35 minutes la dernière année. La phase des tests en individuel s'est déroulée chaque année sur janvier et février. Le but était d'obtenir une variété de mesures langagières, cognitives et psycholinguistiques pour chaque participant.

- la phase de lecture en classe en autonomie et en individuel. Les élèves ont lu les textes en silence et de manière autonome sur un iPad. La phase de lecture s'est déroulée de mars à juin, chaque année scolaire sur les 3 ans depuis 2017.

Figure 17: Protocole expérimental conduit en deux phases chaque année.



## Organisation de la répartition des textes

Le carré latin a été utilisé pour organiser l'ordre des textes au sein des écoles afin de contrebalancer les effets liés à l'ordre des textes. Nous avons pris en compte l'évolution des élèves au sein de l'année en lecture de mars à juin. Les dix textes de chaque type (LIT/SCI) ont été classés en fonction de leur complexité puis organisés en deux groupes de quatre et un groupe de deux. La version des textes (original *versus* simplifié) et l'ordre des textes ont été contrebalancés à travers les quatre écoles. Aucune classe n'a vu le même texte dans sa version originale et simplifiée. Chaque année, chaque élève a lu 20 textes, 10 étaient des textes littéraires (5 originaux et 5 simplifiés) et 10 étaient des documentaires scientifiques (5 originaux et 5 simplifiés).

Après la lecture de chaque texte, les élèves devaient répondre à un questionnaire à choix multiples (QCM) destiné à mesurer leur compréhension du texte. Les questions portaient uniquement sur des éléments qui ne reposaient pas directement sur des connaissances culturelles. Il est important de noter qu'elles étaient identiques dans la version originale et dans la version simplifiée. Le QCM comprenait cinq questions de compréhension, chacune ayant trois alternatives (une correcte ; une incorrecte avec des éléments du texte mais ne répondant pas à la question, une incorrecte sans rapport avec le texte) pour la première année. Pour la deuxième et troisième année, le QCM comprenait toujours cinq questions de compréhension mais chacune avait quatre alternatives (une correcte ; deux incorrectes avec des éléments du texte mais ne répondant pas à la question, une incorrecte sans rapport avec le texte). Les questions ainsi que les propositions, sont toujours présentées dans le même ordre, pour l'ensemble des élèves. En revanche, les propositions de réponses ont été randomisées sur les 100 questions.

Pour les niveaux CE2 et CM1, pour donner suite au QCM un résumé à trous était présenté avec cinq propositions à replacer. Pour chaque texte nous avons toujours la même catégorie grammaticale de propositions, par exemple uniquement des noms ou uniquement des verbes. Nous avons équilibré également les différentes catégories grammaticales entre les textes littéraires et les textes scientifiques.

Afin de motiver les élèves, nous avons mis en place pour chaque lecture un système de bonifications ludiques, dans lequel ils pouvaient gagner des récompenses (étoiles) proportionnel au nombre de réponses correctes fournies. En cas d'hésitation, l'élève pouvait changer sa réponse avant de la valider et de passer à la question suivante. Les étoiles en or correspondaient aux réponses correctes, les étoiles en argent correspondaient à des réponses

incorrectes mais en lien avec un élément du texte et aucune étoile correspondait à la réponse absurde. Ce système nous a permis de créer des questions de plus en plus complexes, d'incorporer de plus en plus de questions inférentielles entre le niveau CE1 et CM1 au fur et à mesure des années afin que les résultats des bons 'compreneurs' ne plafonnent pas.

Nous avons mis en place les lectures de texte et la tâche de compréhension sur les iPads en utilisant le modèle d'iBook. Les textes ont été présentés phrase par phrase. Toutefois cette découpe des phrases a tenu compte du contenu sémantique. Par exemple, lorsque deux phrases se succédaient mais que la seconde était fortement liée par le sens (ex. dans un dialogue), nous ne les avons pas séparées dans la présentation sur les tablettes. Dans les discours, lorsque c'était toujours le même personnage qui parlait, nous ne présentions pas les phrases séparément. Comme il était impossible pour les élèves de relire les phrases et comme la présentation phrase par phrase avec les tablettes ne facilitait pas la lecture, nous avons pris en compte la division sémantique des textes. Nous avons appliqué le principe, « toutes choses restant égales par ailleurs » : lorsqu'une phrase du texte initial a été divisée en deux phrases dans le texte simplifié, l'élève qui lisait la version simplifiée avait les deux phrases présentées à l'écran. Nous avons ainsi créé un corpus parallèle de 69 textes sur les trois ans, simplifiés au niveau lexical, morphosyntaxique et discursif. La division entre les deux versions simplifiée et originale a permis de placer chaque phrase l'une en face de l'autre. C'est-à-dire que si une longue phrase de la version originale était divisée en deux phrases dans la version simplifiée, les deux phrases apparaissaient en même temps à l'écran.

Au niveau pédagogique les enseignants ont été libres de choisir leur mise en œuvre et plusieurs pratiques sont apparues. La lecture sur support numérique s'est faite en classe, individuellement et en autonomie. Elle s'est déroulée sur différents temps : en Activités Pédagogiques Complémentaires (APC), en autonomie sur un temps de classe choisi et en rituel du matin. Ces variations d'organisation pédagogique n'ont pas impacté directement les résultats obtenus ; en effet nous avons observé les scores en vitesse de lecture et les scores en compréhension directement lors de la lecture réalisée en individuel. Les enseignants avaient pour consigne de ne pas intervenir tant que l'ensemble des élèves n'avaient pas répondu aux cinq questions et réalisé le placement des mots dans le résumé afin de ne pas biaiser les résultats.

### **Récolte des données**

Les temps de lecture et les résultats des QCM ainsi que les scores des résumés ont été automatiquement transférés au laboratoire. Cela a permis de suivre les enfants au quotidien.

L'ensemble de la programmation pour la collecte des données a été réalisé par Stéphane Dufau, Ingénieur Recherche, responsable du service expérimentation et analyses au sein du LPC ; ce dispositif a permis une récolte de données pour l'ensemble de l'expérimentation.

### **6.3 Tests en individuel**

Les tests ont été réalisés dans le même ordre pour chaque participant, à raison d'une ou deux séances par semaine. Les élèves n'ont jamais réalisé deux séances dans la même journée et l'ordre de passage a été randomisé afin que les élèves du début de l'alphabet ne passent pas les tests toujours le matin par exemple. Certains tests ont été maintenus sur les trois années alors que d'autres ont été proposés sur une seule année. Nous vous présentons les tests réalisés en CE1 car nous n'avons pas utilisé les tests réalisés en CE2 et CM1 dans nos analyses.

#### **Conscience phonémique**

Un test informatisé (Evalect) a été utilisé pour évaluer les compétences en matière de conscience phonémique (Sprenger-Charolles et al., 2005 ; 2010). Nous avons utilisé le sous-test de suppression de phonèmes, dans lequel les enfants se voyaient présenter des pseudo-mots parlés de type Consonne-Consonne-Voyelle (CCV) et de type Consonne-Voyelle-Consonne (CVC). Les élèves devaient supprimer le premier phonème. Le test comportait 12 items par tâche, donc 24 au total. Le taux d'erreurs moyen pour chaque tâche et le taux d'erreurs moyen pour les deux tâches ont été utilisés dans les analyses.

#### **Lecture à haute voix**

Nous avons utilisé la version française du Test d'efficacité en lecture de mots (TOWRE) adapté au français par Gentaz et ses collègues (2015). Ce test contenait une liste de mots familiers qui devaient être lus à haute voix en 1 minute (60 mots, cinq par ligne) et une liste de pseudo-mots (60 mots, cinq par ligne). Ces mots figuraient parmi les 1000 mots les plus fréquents dans les manuels scolaires de première année (Lété et al., 2004). Les pseudo-mots ont été appariés aux mots en termes de longueur, de structure syllabique et d'orthographe. Vingt des pseudo-mots avaient un graphème dont la prononciation dépendait du contexte (s = /s/ ou /z/...). Pour les besoins de nos analyses, nous avons regroupé les mots et les pseudo-mots en une seule mesure de la vitesse de lecture.

### **Vocabulaire**

Le vocabulaire a été évalué à l'aide du Test de Vocabulaire Actif/Passif (TVAP), un test de vocabulaire standardisé pour les enfants âgés de cinq à huit ans (Deltour & Hupkens, 1980). Dans ce test, l'élève devait sélectionner une image sur six qui correspondait au mot fourni par l'examineur. Le test était composé de 30 mots, 2 points par réponse correcte et 1 point pour une réponse approximative. Nous avons calculé un score global sur 60.

### **Compréhension orale**

Pour évaluer la compréhension des informations syntaxiques et sémantiques, nous avons utilisé un test de français standardisé appelé Epreuve de Compréhension Syntaxico-Sémantique Etalonnée (ECOSSE) développé par Lecocq (1998). Dans ce test, une phrase orale est donnée à l'élève, qui doit choisir l'image correspondante parmi quatre possibilités, dont deux contiennent des pièges lexicaux ou grammaticaux. Le test est conçu pour utiliser une variété de structures syntaxiques, dont la complexité augmente au cours de l'épreuve. Nous n'avons utilisé que les huit items les plus discriminants (M11, M24, M42, N21, N44, P22, P33, R33). Le test a été informatisé, de sorte que chaque élève a été testé dans les mêmes conditions.

### **Mémoire de travail**

Les tests Digit Span (nombres répétés) ont été développés pour permettre une meilleure mesure de la mémoire de travail (Wechsler, 2004). Nous avons utilisé la répétition des nombres en avant avec 16 items. Les deux premiers éléments ont deux chiffres et les deux derniers éléments ont neuf chiffres. Les enfants devaient se rappeler oralement la série de chiffres dans le même ordre.

### **Intelligence non verbale**

Nous avons utilisé le test des matrices progressives de Raven (Raven, 1998), qui consiste en 36 items divisés en trois ensembles de 12 items (ensemble A, ensemble AB et ensemble B). Chaque item représente une figure géométrique avec une partie manquante et le participant devait choisir une pièce sur 6 pour compléter la figure.

### **Morphologie**

Nous avons utilisé une tâche d'achèvement de phrase pour mesurer la sensibilité morphologique (Casalis et al., 2004). Dans ce test, les enfants devaient compléter une phrase en utilisant une forme dérivée d'un mot dont la racine était fournie dans la première partie de la

phrase (par exemple, "quelqu'un qui chasse est un ..." ... "chasseur"). Nous avons utilisé la forme suffixée qui contient 12 éléments.

### **Reconnaissance visuelle des mots**

Nous avons utilisé un test standardisé de reconnaissance visuelle, le Test d'identification de mots écrits (Timé 3) développé par Écalle (2006). Le test permet d'évaluer le niveau de lecture d'enfants de 7 à 15 ans. Dans ce test, 40 mots au total, présentés visuellement, doivent être identifiés dans le cadre de deux tâches ayant chacune 20 propositions. La première tâche consiste à associer un mot choisi parmi cinq à une image. La seconde tâche consiste à associer un mot choisi parmi cinq à un mot-cible, ces deux mots étant reliés sur le plan sémantique. Par exemple pour la tâche 1, pour le mot-cible *moufle*, les cinq propositions de mots comprennent: la réponse correcte (*moufle*), un pseudo-mot homophone (*mouffle*), une syllabe réduite (moufe), un mot visuellement proche (*monfle*), un voisin orthographique (*moule*). Par exemple pour la tâche 2, pour le mot-cible *livre*, les cinq propositions de mots comprennent : la réponse correcte (*album*), un pseudo-mot, visuellement proche (atbum), un pseudo-mot orthographiquement proche, phonologiquement éloigné (alebum), un voisin orthographique (albinos), un pseudo-mot orthographiquement éloigné, phonologiquement identique (*albome*). Le score obtenu est le nombre total de réponses correctes dans les deux tâches.

### **Orthographe**

Nous avons utilisé un test d'orthographe normalisé du français, le Langage oral - Langage écrit - Mémoire, Attention test (L2MA, 2ème édition ; Chevrie-Muller et al., 2010). Dans ce test, sept phrases ont été dictées à l'enfant et la précision a été codée sur deux dimensions, la morphologie (45 points) et la grammaire (31 points).

### **Dénomination rapide automatisée**

Il a été présenté aux enfants 4 lignes de 9 chiffres sur un écran d'ordinateur (même ordre pour tous les participants). Ils devaient les nommer le plus rapidement possible. Nous avons mesuré le temps total pour nommer les 36 chiffres, lequel a été converti en une mesure de vitesse (chiffres par seconde).

### **Traitement des chaînes de caractères**

Nous avons utilisé une tâche de recherche de lettres, qui demandait aux participants de rechercher une lettre cible dans un chapelet de consonnes (voir Ziegler et al., 2008 ; 2010).

Nous avons utilisé des chaînes de consonnes pour limiter les processus lexicaux et phonologiques. Chaque participant a vu 120 éléments. Les éléments sont restés à l'écran jusqu'à la réponse du participant. Ainsi, la variable dépendante critique était le temps de réponse pour des réponses correctes. Nous avons mesuré le temps de réaction.

### Statistiques descriptives des tests en individuel réalisés en CE1

En CE1, onze tests ont été utilisés pour évaluer le profil cognitif et linguistique individuel de chaque participant (voir Tableau 3).

Tableau 3 : Statistiques descriptives pour les 11 tâches cognitives et langagières ainsi que les mesures de l'asymétrie et de l'aplatissement en CE1.

Variabes	Min-Max	Moy	SD	Skewness	Kurtosis
Conscience phonémique : taux d'erreurs (%)	20,83-100	72,85	18,31	-0,76	-0,26
Conscience phonémique : taux d'erreurs CCV (%)	0-100	58,79	27,50	-0,64	-0,65
Conscience phonémique : taux d'erreurs CVC (%)	8,33-100	86,97	18,48	-2,02	4,10
Lecture à haute voix (mots par sec)	0,27-1,57	0,79	0,22	0,57	0,69
Vocabulaire (/60)	28-56	45,28	5,26	-0,43	0,19
Compréhension orale (/8)	1-8	6,37	1,53	-1,19	1,52
Mémoire de travail (/16)	3-13	7,10	1,70	0,59	0,52
Intelligence non verbale (/36)	6-32	22,15	4,80	-0,27	-0,03
Morphologie (/12)	0-12	5,82	2,89	0,05	-0,83
Reconnaissance visuelle des mots (/40)	5-33	18,11	5,24	0,03	0,16
Dictée : Morphologie (/45)	12-45	35,93	7,18	-1,27	1,64
Dictée : Grammaire (/31)	1-26	11,75	5,30	0,20	-0,61
RAN (sec)	10,2-27,4	15,54	3,17	0,95	1,01
Traitement des chaînes de caractères (ms)	881-3712	1959	520	0,74	0,54

### 6.4 Analyses des Questions à Choix Multiples

La première année (CE1), la fiabilité interne de la mesure des réponses au QCM était élevée avec un alpha de Cronbach de 0,954. La proportion de questions de type 'prélevé' était de 82 %, tandis que 18 % des questions demandaient aux élèves de faire des inférences. La seconde année, l'alpha de Cronbach était de 0,949 prouvant la fiabilité interne des items du questionnaire. Le pourcentage de questions de type 'prélevé' était de 74 %, tandis que 26 % des

questions demandaient aux élèves de faire des inférences. La troisième année la fiabilité interne de la mesure des réponses au QCM était élevée avec un alpha de Cronbach de 0,964. Le taux des questions de type 'prélevé' était de 38 %, tandis que 62% des questions demandaient aux élèves de faire des inférences (Cain et al., 2001). Nous avons choisi d'augmenter la quantité de questions de type 'inféré' chaque année, ce qui correspond à une attente du développement lié à la compréhension et permet de ne pas avoir un effet de plafonnement dans les résultats, notamment pour les élèves n'ayant aucune difficulté de compréhension.

## **6.5 Traitement des données**

### **Vitesse par mot**

Comme nous connaissons le nombre exact de mots par phrase et le temps passé sur chaque phrase, nous avons calculé un temps de réaction moyen par mot. Comme étape de prétraitement des données collectées, nous avons d'abord effectué une élimination des temps de réaction inférieurs à 200 ms par mot. La valeur de 200 ms était basée sur une analyse visuelle préliminaire des données. Les éliminations correspondent à des situations dans lesquelles la phrase n'a pas pu être lue correctement, soit par manque d'intérêt, soit en raison d'une interaction involontaire avec l'interface de l'iPad. Nous avons également éliminé les données lorsque les temps de réaction étaient supérieurs à 3 écarts-types par rapport à la moyenne personnelle du sujet. Nous avons ensuite transformé les temps de réaction par mot (RT) en vitesse par mot ( $1/RT$ ), ce qui correspond au nombre de mots lus par seconde. Les vitesses de lecture en nombre de mots lus par seconde ont été utilisées dans toutes les analyses ultérieures.

### **Compréhension**

Nous avons calculé le pourcentage de réponses correctes pour chaque élève aux QCM. Les élèves ont répondu à 100 questions chaque année (20 textes : 5 questions/texte).

### **Résumé**

Le résumé a été proposé en CE2 et CM1. Nous avons calculé un score moyen sur 5 pour chaque élève car chaque résumé comporte 5 propositions à placer dans le texte.

## **6.6 Propriétés statistiques des variables dépendantes**

Dans le Tableau 4, nous présentons les caractéristiques descriptives des variables dépendantes comportant le minimum (Min), le maximum (Max), la moyenne (Moy), l'écart-type (SD), l'asymétrie (skewness) et l'aplatissement (kurtosis) pour les 3 années.

Tableau 4 : Statistiques descriptives des variables dépendantes du test de compréhension et des lectures de textes sous format numérique (moyenne sur l'ensemble des 20 textes) pour chaque année (CE1, CE2, CM1).

CE1-CE2-CM1	Min	Max	Moy	SD	Skewness	Kurtosis
Vitesse (mots par seconde)						
CE1	0,49	2,50	1,19	0,39	0,88	0,69
CE2	0,68	3,08	1,54	0,48	0,96	1,00
CM1	0,78	3,03	1,76	0,50	0,63	-0,08
Compréhension (% réponses correctes)						
CE1	28,0	98,0	75,63	16,45	-0,93	-0,01
CE2	23,0	96,0	67,26	18,07	-0,45	-0,68
CM1	18,0	98,0	67,37	20,98	-0,46	-0,85
Résumé (/5)						
CE2	0,85	5,0	3,20	0,98	-0,367	-0,717
CM1	0,65	4,9	3,32	1,05	-0,502	-0,630

## 6.7 Effets de la simplification sur la fluidité

Pour déterminer si la simplification avait des effets pour les différents types de textes sur les trois ans, nous avons analysé les vitesses de lecture en mots par seconde en utilisant une analyse de variance (ANOVA) 2 \* 2 \* 3 avec la simplification (originale vs simplifiée), le type de texte (littéraire vs scientifique) et le niveau scolaire (CE1, CE2 et CM1) comme facteurs intra-sujet. Les résultats de l'ANOVA sont présentés dans le Tableau 5.

Tableau 5 : Résultats de l'ANOVA Simplification\*Type de textes\*Niveau en intra sujet<sup>58</sup> pour la fluidité.

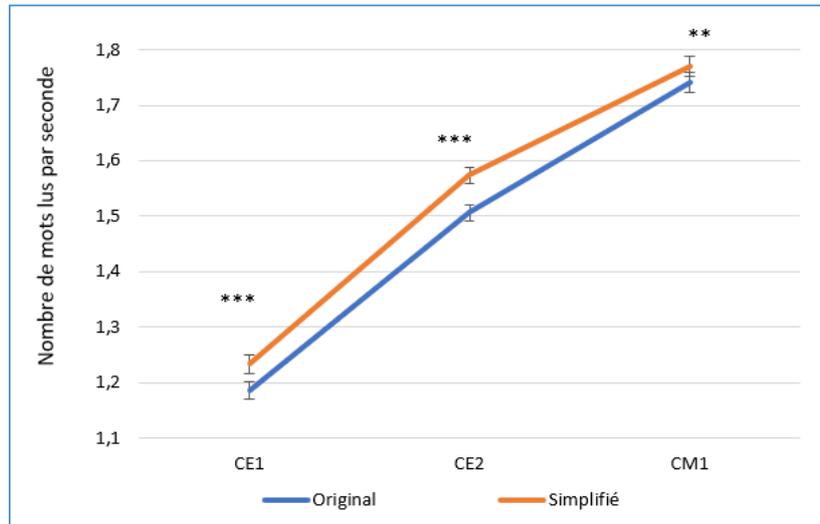
Effets	df	F	p	$\eta_p^2$	sign
Simplification	1, 137	81,327	,000	,373	***
Texte	1, 137	53,986	,000	,283	***
Niveau	2, 274	208,129	,000	,603	***
Simplification * Texte	1, 137	0,071	,790	,001	
Simplification * Niveau	2, 274	3,353	,036	,024	*
Texte * Niveau	2, 274	0,424	,655	,003	
Simplification * Texte * Niveau	2, 274	8,969	,000	,061	***

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

<sup>58</sup> Sont présentés les effets, le degré de liberté (df), le Fisher (F), la valeur de probabilité (p), l'éta-carré partiel, proportion de la variance spécifique expliquée par le facteur lorsque l'effet des autres facteurs est contrôlé ( $\eta_p^2$ ), la significativité, (sign).

L'effet de la simplification est fortement significatif. Les textes simplifiés sont lus plus rapidement que les textes originaux. Il y a également un effet du type de texte (les textes littéraires sont lus plus rapidement que les textes scientifiques) et un effet du niveau scolaire (la vitesse de lecture progresse d'année en année). L'interaction entre simplification et niveau est significative. Elle est présentée dans la Figure 18.

Figure 18 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour la fluidité. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Pour décomposer cette interaction, nous avons calculé les contrastes par niveau scolaire pour l'effet de la simplification (Tableau 6). Nous constatons que l'effet de la simplification est fortement significatif pour chaque niveau scolaire.

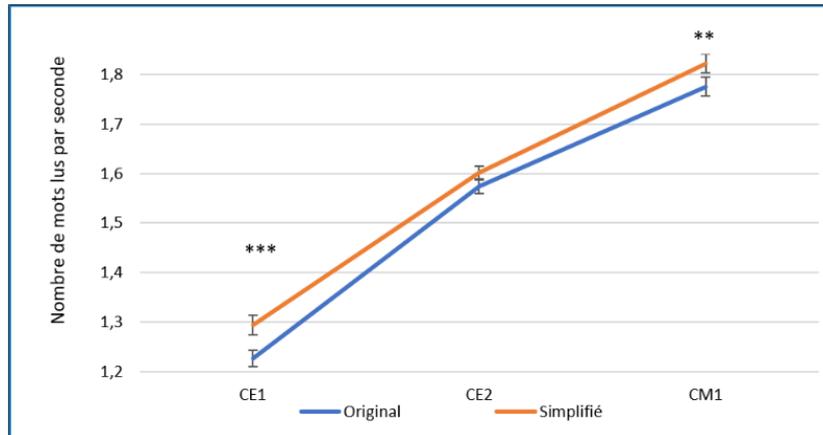
Tableau 6 : Effet de la simplification en fonction des niveaux scolaires pour la fluidité.

Effet de la simplification	df	F	p	$\eta_p^2$	sign
CE1	1, 137	19,945	,000	,127	***
CE2	1, 137	55,276	,000	,287	***
CM1	1, 137	10,229	,002	,69	**

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

Enfin, les analyses montrent une triple interaction (Simplification \* Texte \* Niveau) ce qui suggère que les effets de la simplification varient en fonction de la nature du texte et en fonction du niveau scolaire. Pour décomposer cette triple interaction, nous présentons les effets de la simplification par niveau scolaire pour les textes littéraires dans la Figure 19 et pour les documentaires scientifiques dans la Figure 20.

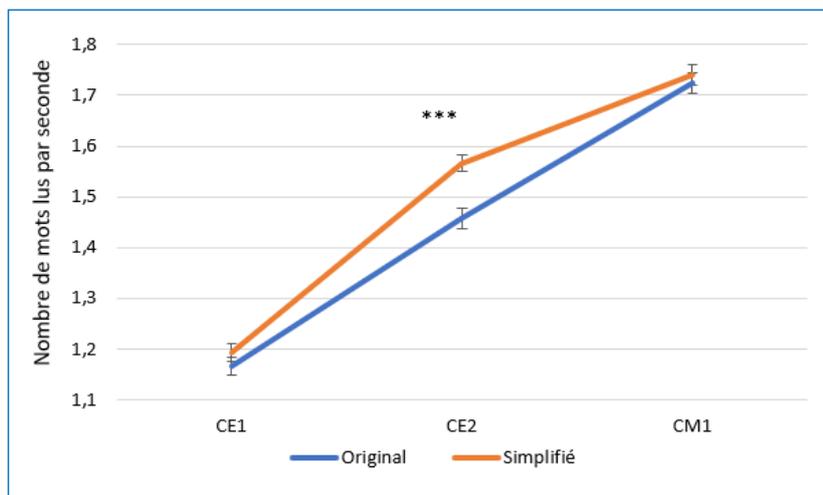
Figure 19 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour les textes littéraires pour la fluidité. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Pour les textes littéraires on constate un effet de la simplification, pour le CE1 ( $F(1,137) = 16,651, p < ,000$ ) et le CM1 ( $F(1,137) = 9,010, p < ,003$ ).

L'effet de la simplification par niveau scolaire pour les textes scientifiques est présenté dans la Figure 20. On constate que, pour les textes scientifiques, la simplification n'est significative que pour le CE2 ( $F(1,137) = 75,69, p < ,000$ ) mais pas pour le CE1 ( $F(1,137) = 3,77, p = ,054$ ) et le CM1 ( $F(1,137) = 1,13, p = ,289$ ).

Figure 20 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour les documentaires scientifiques. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



## 6.8 Effets de la simplification sur la compréhension

Afin d'évaluer la compréhension nous avons choisi d'utiliser deux mesures, le Questionnaire à Choix Multiples (QCM) et le résumé.

### 6.8.1 Compréhension : Questions à Choix Multiples

Pour déterminer si la simplification avait des effets pour les différents types de textes en longitudinal, nous avons analysé les scores de compréhension au QCM en utilisant une analyse de variance (ANOVA)  $2 \times 2 \times 3$  avec la simplification (originale vs simplifiée), le type de texte (littéraire vs scientifique) et le niveau scolaire (CE1, CE2 et CM1) comme facteurs intra sujet. Les résultats de l'ANOVA sont présentés dans le Tableau 7.

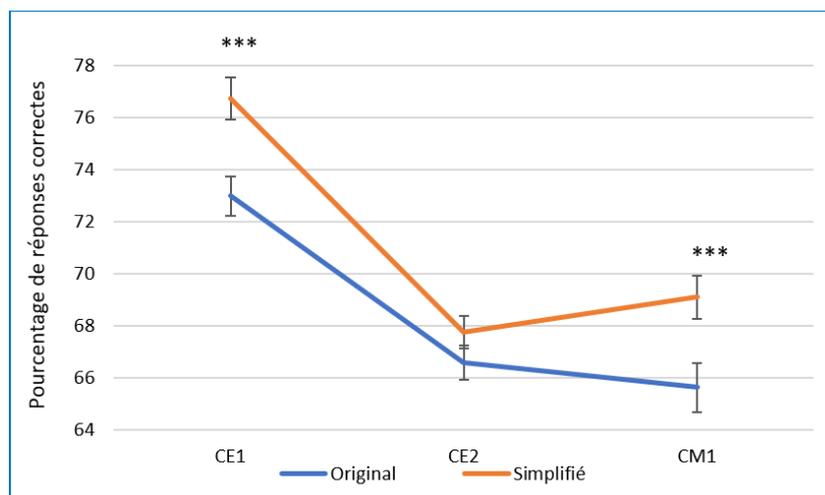
Tableau 7 : Résultats de l'ANOVA Simplification\*Type de textes\*Niveau en intra sujet pour le QCM.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 137	32,020	,000	,189	***
Texte	1, 137	30,245	,000	,181	***
Niveau	2, 274	28,376	,000	,172	***
Simplification * Texte	1, 137	1,037	,310	,008	
Simplification * Niveau	2, 274	3,479	,032	,025	*
Texte * Niveau	2, 274	6,396	,002	,045	**
Simplification * Texte * Niveau	2, 274	2,972	,053	,021	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

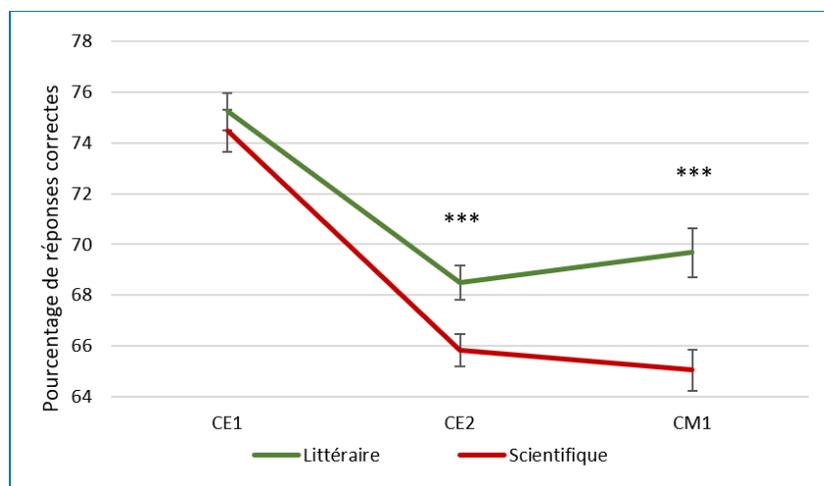
Comme pour la vitesse, nous avons obtenu un effet significatif de la simplification (les textes simplifiés ont été mieux compris que les textes originaux), un effet de la nature du texte ( $F(1,137) = 30,25$ ,  $p < ,000$ ) et un effet du niveau scolaire ( $F(2,274) = 28,38$ ,  $p < ,000$ ). L'interaction entre les effets de la simplification et le niveau était significative. Cette interaction est présentée dans la Figure 21. L'analyse des contrastes montre que la simplification était significative pour le CE1 ( $F(1,137) = 23,80$ ,  $p < ,000$ ) et le CM1 ( $F(1,137) = 15,77$ ,  $p < ,000$ ) mais pas pour le CE2 ( $F(1,137) = 2,53$ ,  $p = ,114$ ).

Figure 21: Effet de la simplification par niveau scolaire en pourcentage de réponses correctes au QCM. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Nous avons également obtenu une interaction significative entre les effets de nature de texte et le niveau scolaire reflétant le fait que la différence entre les deux types de textes augmente d'année en année. Cette interaction est présentée dans la Figure 22. Les analyses de contrastes confirment que la différence entre les deux types de textes est significative en CE2 ( $F(1,137) = 11,90, p < ,001$ ) et en CM1 ( $F(1,137) = 27,44, p < ,000$ ) mais pas en CE1 ( $F(1,137) = 1,13, p < ,290$ ).

Figure 22 : Effet du type de texte par niveau scolaire en pourcentage de réponses correctes au QCM. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Il est important de noter que si les scores en compréhension diminuent entre le CE1 et le CM1, ce n'est pas pour autant que le niveau de compréhension a baissé. En effet, les questions ont été construites pour qu'elles soient de plus en plus complexes afin que les scores des meilleurs élèves ne plafonnent pas. Nous avons donc proposé davantage de questions de type inférentiel, ce qui pourrait expliquer cette « chute » apparente des scores de compréhension.

### 6.8.2 Compréhension : Résumé

Dans le Tableau 8, nous présentons les statistiques descriptives pour la variable « résumé ». L'inspection des moyennes suggère que la tâche n'était ni trop facile (effet de plafond) ni trop difficile (effet de plancher).

Tableau 8 : Statistiques descriptives de la variable dépendante « résumé » du test de compréhension sous format numérique en fonction des deux conditions (simplification et type de textes) pour chaque année, CE2 et CM1.

CE2-CM1	Min	Max	Moy	SD	Skewness	Kurtosis
<b>CE2</b>						
LIT-ORIG	0,8	5,0	3,61	1,07	-0,717	-0,217
LIT-SIMP	0,4	5,0	3,59	1,17	-0,656	-0,391
SCI-ORIG	0,4	5,0	2,80	1,10	0,001	-0,702
SCI-SIMP	0,6	5,0	2,80	1,11	-0,187	-0,924
<b>CM1</b>						
LIT-ORIG	1,0	5,0	3,35	1,21	-0,308	-1,040
LIT-SIMP	0,6	5,0	3,52	1,06	-0,635	-0,103
SCI-ORIG	0,8	5,0	3,24	1,24	-0,353	-0,859
SCI-SIMP	0,0	5,0	3,16	1,23	-0,388	-0,737

Les scores dans la tâche de résumé ont été analysés en utilisant une ANOVA 2 \* 2 \* 2 intra-sujets avec la simplification (originale vs simplifiée) et le type de textes (littéraire vs scientifique) pour les deux niveaux scolaires car les élèves ont réalisé cette tâche en CE2 et en CM1. Les résultats de l'ANOVA sont présentés dans le Tableau 9.

Tableau 9 : Résultats de l'ANOVA Simplification\*Type de textes\*Niveau en intra sujet pour les résumés.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 137	0,240	,625	,002	
Texte	1, 137	196,796	,000	,590	***
Niveau	1, 137	3,987	,048	,028	*
Simplification * Texte	1, 137	2,016	,158	,015	
Simplification * Niveau	1, 137	0,446	,506	,003	
Texte * Niveau	1, 137	63,326	,000	,316	***
Simplification * Texte * Niveau	1, 137	3,322	,071	,024	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

Les résultats ne montrent aucun effet de la simplification et aucune interaction au niveau de la simplification. Il nous semble que cette tâche n'a pas été sensible pour évaluer le niveau de compréhension et donc les effets de la simplification.

## 6.9 Discussion

L'objectif général de cette première étude était d'examiner si, sur trois ans, la simplification de textes pouvait améliorer la compréhension et la fluidité en lecture. Les

résultats ont montré que la simplification a amélioré la vitesse de lecture pour les trois niveaux scolaires mais pas de manière similaire pour les deux types de textes sur les trois années. Ce constat conduit à être prudent sur les conclusions. En ce qui concerne la compréhension (QCM), une meilleure compréhension a été obtenue pour les textes simplifiés par rapport aux originaux mais pas de manière similaire chaque année. Ces résultats correspondent à ceux d'études antérieures, montrant que la simplification a un effet positif surtout sur la fluidité, présentées dans la section 3.4.2. Il est important de noter que les élèves ont répondu exactement aux mêmes QCM après la lecture de textes originaux ou simplifiés, ce qui permet une comparaison contrôlée. Ces résultats ont montré que la simplification ne réduisait pas la compréhension et que les éléments essentiels du texte, nécessaires pour construire la compréhension, sont présents dans le texte simplifié comme dans l'original (Siddharthan, 2014). En revanche pour les résumés à trous, la simplification n'a pas été efficace. Les tâches de construction de résumés, d'organisation de données afin de créer un résumé auraient pu être envisagées après la lecture des textes. Mais dans notre étude, en raison du nombre important d'élèves, il n'était pas possible d'informatiser la collecte des données issues d'autres tâches. Pour autant dans le cadre scolaire, le résumé est proposé afin de permettre aux élèves de mieux se représenter la nature de l'activité de compréhension. C'est une tâche qui favorise l'organisation en mémoire des informations lues et incite à une auto-évaluation de la compréhension (Cèbe, Goigoux & Thomazet, 2014).

Nous avons choisi de réaliser un suivi de cohorte sur trois ans afin d'examiner les effets de la simplification pendant le temps du développement de l'apprentissage de la lecture. L'hypothèse sous-jacente était que les effets de la simplification pouvaient diminuer puisque le facteur décodage est de moins en moins prédominant dans la fluidité (Kim & Wagner, 2015) et dans la compréhension en lecture (Morais, 1993). Nous avons constaté que les effets de la simplification étaient présents du CE1 au CM1 pour la fluidité et la compréhension. Cela suggère que la simplification peut être efficace sur l'ensemble du cursus scolaire élémentaire.

Pour conclure, bien que des études précédentes aient principalement examiné les effets de la simplification des textes dans des populations ayant des besoins particuliers, comme les lecteurs analphabètes (Aluísio & Gasperin, 2010) ou dyslexiques (Gala & Ziegler 2016 ; Rello et al., 2013a), nos résultats semblent suggérer que la simplification de textes est bénéfique pour un public de jeunes lecteurs en classe ordinaire sans difficultés particulières. En effet, nos résultats montrent que la simplification peut apparaître comme bénéfique pour la fluidité et la compréhension.

## Chapitre 7

### Différences inter-individuelles : qui bénéficie de la simplification ?

Dans ce chapitre, nous avons analysé si la simplification de textes était bénéfique pour tous les élèves ou si le gain était plus important pour certains lecteurs en fonction de leurs compétences initiales. Afin de traiter cette question nous avons recherché quelles seraient les variables cognitives ou langagières qui expliqueraient les gains en vitesse de lecture et les gains en compréhension. Les analyses sont réalisées sur les données de l'étude présentée dans le chapitre 6.

#### 7.1 Méthodologie du calcul des gains

Les scores des gains ont été calculés pour chaque variable dépendante pour chaque élève et pour chaque année. Nous aurions pu travailler en scores absolus ( $X_{\text{SIMP}} - X_{\text{ORIG}}$ ) mais cette formule de calcul ne tient pas compte du score initial de l'élève. Mais le gain relatif de la simplification dépend des scores obtenus dans les textes originaux et donc des performances initiales de l'élève. Le score de gain relatif est la différence entre les moyennes sur le texte simplifié et le texte original divisée par la moyenne de l'élève sur le texte original. Le gain de la simplification a été calculé grâce à cette formule :

$$\text{Gain de simplification} = (X_{\text{SIMP}} - X_{\text{ORIG}}) / X_{\text{ORIG}}$$

Ensuite nous avons réalisé pour chaque élève une moyenne des gains sur les trois années au niveau des variables : vitesse de lecture par mot et compréhension sur les QCM. Nous avons cherché à savoir, selon les difficultés initiales des élèves au niveau cognitif et langagier, quels étaient les élèves qui bénéficiaient de la simplification de textes. Nous avons donc effectué des analyses de corrélations bivariées entre les scores de gain sur la vitesse de lecture et la compréhension et les facteurs cognitifs et langagiers par rapport aux tests initiaux réalisés en CE1 par chacun des élèves de la cohorte (section 6.3).

#### 7.2 Lecteurs bénéficiant de la simplification

Dans le Tableau 10, nous présentons les corrélations de Pearson entre les gains en vitesse de lecture et de compréhension et les performances de l'ensemble des élèves pour chacun des 11 tests en individuel (section 6.3) ainsi que leurs subdivisions pour certains tests.

Tableau 10 : Corrélations bivariées entre les variables cognitives et linguistiques individuelles et les gains en simplification relatifs aux scores pour la vitesse de lecture et la compréhension.

Variabes	Gains de vitesse en lecture par mot	Gains Compréhension
Conscience phonémique : taux d'erreurs	-,145	,201**
Conscience phonémique : taux d'erreurs CCV	-,122	,150
Conscience phonémique : taux d'erreurs CVC	-,109	,177*
Lecture à haute voix	,010	-,144
Lecture à haute voix (mots familiers)	-,008	-,167*
Lecture à haute voix (mots inventés)	-,050	-,118
Vocabulaire	,033	-,170*
Compréhension orale	,092	-,147
Mémoire de travail	-,099	-,034
Intelligence non verbale	,023	-,132
Morphologie	,060	-,037
Reconnaissance visuelle des mots	-,081	-,135
Dictée : Morphologie	,003	-,185*
Dictée : Grammaire	-,027	-,103
RAN	-,102	,147
Traitement des chaînes de caractères	,019	,106

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$

Aucun test cognitif et langagier n'affiche de corrélation significative avec les gains en vitesse de lecture (Tableau 10). Seuls les gains en compréhension sont associés à des variables individuelles. Le taux d'erreurs moyen en conscience phonémique et le taux d'erreurs moyen en CVC sont faiblement corrélés avec les gains en compréhension, respectivement ( $\rho = ,201$  et  $\rho = ,177$ ). Ces deux corrélations suggèrent que la simplification améliore la compréhension en longitudinal des apprentis lecteurs ayant des difficultés en phonologie. La corrélation entre l'orthographe (morphologie) et les gains en compréhension sur les trois années ( $\rho = - ,185$ ) fait apparaître que les élèves ayant des difficultés en orthographe bénéficient de la simplification. La corrélation ( $\rho = - ,167$ ) entre la vitesse de lecture à haute voix des mots familiers et le gain en compréhension montre que les lecteurs débutants les plus faibles ont enregistré des gains importants en compréhension sur les trois années (voir Figure 23, ci-dessous). La corrélation entre le vocabulaire et les gains en compréhension ( $\rho = - ,170$ ) fait apparaître également que les élèves ayant un niveau plus faible en vocabulaire bénéficient de la simplification, voir Figure 24, ci-dessous).

Figure 23: Corrélation entre les gains en compréhension (gain relatif) pour les textes simplifiés et originaux et un test de lecture à haute voix normalisé (lecture en 1 minute).

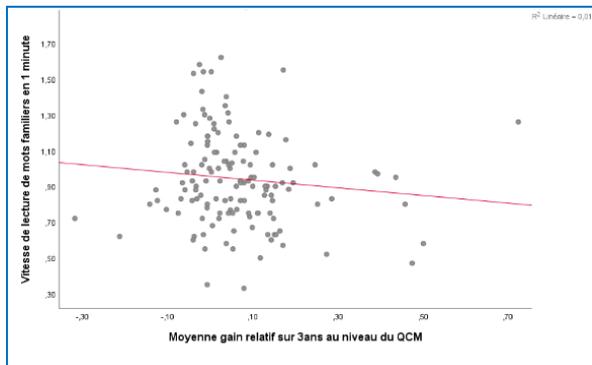
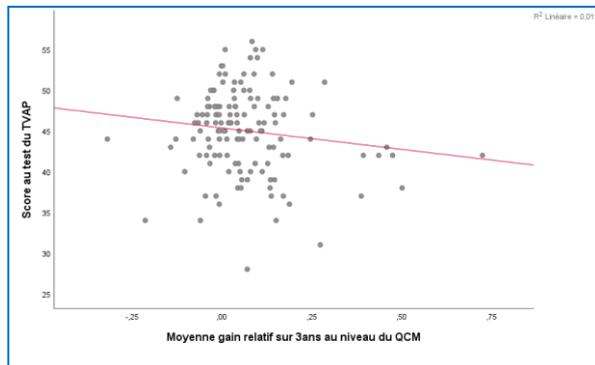


Figure 24: Corrélation entre les gains en pourcentage de réponses correctes pour les textes simplifiés et originaux et un test standardisé de vocabulaire (Échelle de vocabulaire TVAP).



Nous constatons dans le Tableau 10 que ce sont les élèves de l'ensemble de la cohorte ayant le plus faible niveau en conscience phonémique, orthographe (morphologie), lecture de mots familiers (Figure 23) et vocabulaire (Figure 24) la première année, qui bénéficient le plus de la simplification au niveau de la compréhension sur l'ensemble des trois ans.

### 7.3 Effets de la simplification suivant le niveau initial en vitesse de lecture de mots familiers et en vocabulaire sur la fluidité et la compréhension

Ensuite, nous avons analysé, suivant le niveau initial des élèves, quels étaient les effets de la simplification de textes pour les deux variables dépendantes : vitesse de lecture au niveau des mots et compréhension. Nous avons constitué des groupes en fonction de leur niveau initial au moyen du calcul des scores Z, dans les deux compétences : lecture à haute voix de mots familiers et niveau de vocabulaire. Nous avons créé à partir des résultats en score Z, pour chaque test, trois groupes de performances : les élèves ayant un faible niveau (groupe 1), les lecteurs ayant un niveau moyen (groupe 2) et les élèves ayant le meilleur niveau (groupe 3). Nous avons pu alors comparer l'impact de la simplification au fil du temps suivant le niveau initial en lecture à haute voix de mots familiers et en vocabulaire. Nous avons analysé cette évolution pour la cohorte des 138 élèves présents sur les trois années d'expérimentation.

### 7.3.1 Effets de la simplification suivant le niveau initial en vitesse de lecture à haute voix de mots familiers sur la fluence et la compréhension

Nous avons conçu les trois groupes en fonction du score Z de chaque élève pour le test de lecture à haute voix en une minute de mots familiers de manière assez classique en fonction de la courbe gaussienne<sup>59</sup>.

Le groupe 1 est constitué de 21 élèves ayant un score Z entre - 2,3 et - 1, niveau faible ;

Le groupe 2 est constitué de 96 élèves ayant un score Z entre - 1 et + 1, niveau moyen ;

Le groupe 3 est constitué de 21 élèves ayant un score Z entre + 1 et + 3,5, niveau supérieur.

### Résultats pour la fluence et la compréhension

L'évolution pour la fluence suivant le niveau initial en vitesse de lecture à haute voix de mots familiers a été analysée au moyen d'une ANOVA 2\*2\*3\*3 avec la simplification de textes (original vs simplifié), le type de texte (littéraire vs scientifique) et les 3 niveaux scolaires (CE1-CE2-CM1) comme facteur en intra sujet et les 3 groupes en facteur inter-sujets. Les résultats ne montrent pas d'interaction de la simplification de textes par groupe constitué ( $F(2,135) = 0,103, p = 0,903$ ). Il n'y a donc pas de différence significative en fluidité pour les gains de la simplification entre les trois groupes constitués. Nous avons procédé de la même manière avec la variable de compréhension QCM. Nous n'avons obtenu aucun effet de la simplification par groupe ( $F(2,135) = 1,971, p = 0,381$ ). Ces résultats suggèrent qu'il n'y a pas de différence significative entre le gain en compréhension et les trois groupes constitués en fonction du niveau initial en vitesse de lecture de mots familiers.

### 7.3.2 Effets de la simplification suivant le niveau initial en vocabulaire sur la fluence et la compréhension

L'évolution suivant le niveau initial en vocabulaire a été analysée en constituant trois groupes en fonction du score Z de chaque élève pour le test de vocabulaire (TVAP), de manière assez classique en fonction de la courbe gaussienne<sup>60</sup>.

Le groupe 1 est constitué de 20 élèves ayant un score Z entre - 3,3 et - 1, niveau faible ;

Le groupe 2 est constitué de 100 élèves ayant un score Z entre - 1 et + 1, niveau moyen ;

Le groupe 3 est constitué de 18 élèves ayant un score Z entre + 1 et + 2,2, niveau supérieur.

---

<sup>59</sup> Nous aurions pu constituer les 3 groupes de manière différente, pour avoir des groupes aux effectifs plus équivalents. Nous avons réalisé les analyses avec les deux configurations de groupes et les résultats sont similaires pour la simplification au niveau de la vitesse par mot et pour la compréhension dans les deux configurations de groupes.

<sup>60</sup> Idem

### Résultats pour la fluence

L'évolution pour la fluence suivant le niveau initial en vocabulaire, a été analysée au moyen d'une ANOVA 2\*2\*3\*3 avec la simplification de textes (original vs simplifié), le type de texte (littéraire vs scientifique) et les 3 niveaux scolaires (CE1-CE2-CM1) comme facteur en intra sujet et les 3 groupes en facteur inter-sujets. Au niveau de la variable fluence, les résultats sont présentés dans le Tableau 11, ci-dessous.

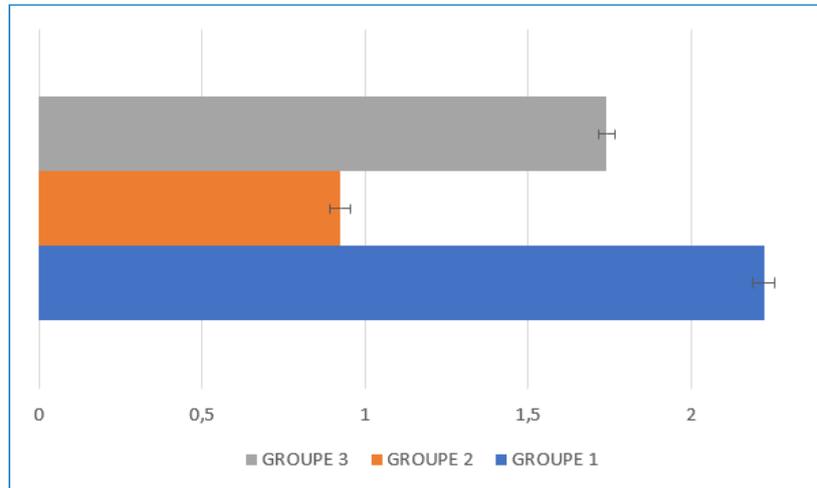
Tableau 11 : Résultats de l'ANOVA Simplification\*Type de textes\*Niveau en intra sujet\* inter-sujets pour la variable fluence.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 137	89,576	,000	,399	***
Texte	1, 137	23,863	,000	,150	***
Niveau	2, 274	121,146	,000	,473	***
Texte*Groupe	2, 274	2,060	,131	,030	
Niveau*Groupe	4, 548	0,966	,427	,014	
Simplification*Groupe	2, 274	7,314	,001	,098	***
Simplification * Texte	1, 137	1,214	,273	,009	
Simplification * Niveau	2, 274	0,900	,408	,007	
Texte * Niveau	2, 274	0,645	,525	,005	
Texte*Simplification*Groupe	2, 274	1,837	,163	,026	
Texte*Niveau *Groupe	4, 548	0,404	,806	,006	
Simplification*Niveau *Groupe	4, 548	0,360	,837	,005	
Simplification * Texte * Niveau	2, 274	3,898	,021	,028	*
Texte*simplification*Niveau*Groupe	4, 548	1,294	,273	,019	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\* $p < 0,001$

Nous avons constaté une triple interaction entre la simplification, le type de textes et les trois niveaux scolaires ce qui suggère que les effets de la simplification ont été différents selon le type de textes et le niveau. Cette observation a déjà été décrite dans la section 6.7 (voir Figures 19 et 20). Nous avons constaté une interaction entre la simplification et les groupes, ce qui suggère que la simplification a été bénéfique selon le niveau initial en vocabulaire des élèves (Tableau 11 et Figure 25). Les effets principaux étaient tous significatifs : la simplification, le type de textes, le niveau.

Figure 25 : Taille d'effet en score Z de la simplification pour les gains en fluence en fonction du niveau initial en vocabulaire des élèves. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



La majorité des élèves a bénéficié de la simplification, quel que soit leur niveau initial en vocabulaire. Ce constat corrobore les conclusions du chapitre 6. Le plus faible bénéfice de la simplification est pour les élèves du groupe 2 alors que les élèves du groupe 3, les plus performants initialement en vocabulaire, ont un bénéfice plus important. La Figure 25 montre que les élèves ayant le niveau le plus faible en vocabulaire en CE1 sont ceux qui ont le plus bénéficié de la simplification au cours des trois ans.

### Résultats pour la compréhension

L'évolution pour la compréhension (réponses aux QCM) suivant le niveau initial en vocabulaire, a été analysée au moyen d'une ANOVA 2\*2\*3\*3 avec la simplification de textes (original vs simplifié), le type de texte (littéraire vs scientifique) et les 3 niveaux scolaires (CE1-CE2-CM1) comme facteur en intra sujet et les 3 groupes en facteur inter-sujets. Au niveau de la variable compréhension, les résultats sont présentés dans le Tableau 12.

Tableau 12 : Résultats de l'ANOVA Simplification\*Type de textes\*Niveau en intra sujet\* inter-sujets pour la variable compréhension.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 137	30,737	,000	,185	***
Texte	1, 137	18,527	,000	,121	***
Niveau	2, 274	14,467	,000	,097	***
Texte*Groupe	2, 274	1,775	,173	,026	
Niveau*Groupe	4, 548	1,936	,105	,028	
Simplification*Groupe	2, 274	1,915	,151	,028	
Simplification*Texte	1, 137	0,206	,650	,002	
Simplification*Niveau	2, 274	2,356	,097	,017	
Texte*Niveau	2, 274	3,302	,038	,024	*
Texte*Simplification*Groupe	2, 274	0,693	,502	,010	
Texte*Niveau*Groupe	4, 548	1,795	,130	,026	
Simplification*Niveau*Groupe	4, 548	0,389	,817	,006	
Simplification*Texte*Niveau	2, 274	1,758	,174	,013	
Texte*simplification*Niveau*Groupe	4, 548	1,277	,279	,019	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

L'interaction du type de texte en fonction du niveau est significative. Ce résultat fait ressortir qu'il y a une différence entre les types de textes et les trois niveaux scolaires. Cet effet a déjà été décrit dans le chapitre 6, la sous-section 6.8.1, Figure 22. Les effets principaux étaient tous significatifs : la simplification, le type de textes, le niveau scolaire. Aucun effet de la simplification avec les autres facteurs n'est significatif. L'ensemble des groupes, quel que soit le niveau de vocabulaire initial, bénéficie de la simplification, quel que soit le type de textes et les années.

## 7.4 Discussion

Afin de savoir si la simplification de textes était bénéfique pour tous les élèves ou si le gain était plus important pour certains lecteurs, selon leurs compétences initiales, pour l'ensemble de la cohorte, nous avons choisi 11 tests qui permettaient de déterminer quelles étaient les différences inter-individuelles qui pourraient expliquer le bénéfice de la simplification. Aucune corrélation n'est significative entre les variables cognitives et langagières des élèves et les gains en vitesse de lecture. Dans nos résultats, comme on peut le voir dans le Tableau 10, de faibles corrélations significatives existent suggérant que les élèves

ayant un faible niveau en phonologie, orthographe (morphologie), lecture de mots familiers et vocabulaire la première année sont ceux qui ont le plus bénéficié de la simplification au niveau de la compréhension sur l'ensemble des trois ans. D'après Lervåg et al., (2018) les variations de la compréhension orale et écrite ont pour puissants prédicteurs en longitudinal le vocabulaire, la grammaire, la mémoire de travail verbale et les compétences inférentielles. Gentaz, Sprenger-Charolles & Theurel (2015) ont montré que la variance dans la compréhension de la lecture en français pour des élèves de CP a été expliquée par les quatre prédicteurs suivants : les compétences de décodage, se révélant être le meilleur prédicteur, la compréhension orale étant le deuxième, le vocabulaire le troisième et la conscience phonémique le dernier. Dans notre étude, les résultats de corrélation entre les gains en compréhension de lecture grâce à la simplification, les facteurs vitesse de lecture à haute voix et le niveau en vocabulaire et phonologie vont dans le même sens que l'étude de Gentaz et al. (2015). La simplification a soutenu la tâche de compréhension des élèves ayant des capacités faibles en phonologie, orthographe (morphologie), lecture de mots familiers et vocabulaire la première année.

Quel que soit le niveau initial en lecture à haute voix de mots familiers, la majorité des élèves ont gagné en fluidité de lecture grâce à la simplification. L'analyse des gains de la simplification en fonction du niveau initial en lecture de mots familiers à haute voix n'est pas significative. Les analyses précédentes, qui se sont focalisées uniquement sur les performances au CE1, ont montré des corrélations faibles mais significatives entre la lecture à haute voix et les gains de fluidité (Javourey-Drevet, François, Gala, Dufau, Ginestier & Ziegler, en révision). Pour autant, sur les trois années, nous ne retrouvons pas de corrélation significative au niveau des gains en vitesse de lecture par mot selon le niveau initial en lecture à haute voix de mots familiers.

En fonction du niveau initial en vocabulaire, l'analyse de l'effet de la simplification a montré une corrélation significative avec le gain en fluidité pour tous les élèves et en particulier pour les élèves ayant un faible niveau initial de vocabulaire. La simplification appliquée aux textes originaux a permis de contrebalancer les difficultés en vocabulaire. L'importance du vocabulaire s'accroît avec le temps dans la tâche de compréhension. Les modifications apportées en simplifiant, par l'utilisation de mots moins longs, l'utilisation de mots plus fréquents, le remplacement des pronoms par leur référents (voir Tableau 5, chapitre 4), ont permis d'aider les élèves ayant un faible niveau initial en vocabulaire en CE1. De plus, d'après Lervåg et al. (2018), le facteur 'langage' dépend du vocabulaire avec lequel il est fortement corrélé ( $.93^{**}$ ) et il est interdépendant avec le décodage des mots ( $.23^{**}$ ). En aidant les élèves qui ont des difficultés en vocabulaire avec la simplification, ils gagnent en vitesse de décodage

des mots et donc en fluidité. Toutefois, un gain est observé au niveau de la compréhension avec la simplification pour l'ensemble des élèves, quel que soit le niveau initial en vocabulaire. Chez le lecteur expert, la relation est très forte entre la compréhension écrite et orale alors que chez le lecteur novice, la relation est très importante entre la compréhension écrite et le décodage (Morais, 1993).

En conclusion, les gains obtenus grâce à la simplification, en fluence pour les élèves ayant un faible niveau initial en vocabulaire et en compréhension pour les élèves ayant de faibles performances en phonologie, orthographe, vitesse de lecture à haute voix et vocabulaire, soulignent l'importance d'adapter le matériel de lecture aux besoins des élèves (Kendeou, 2014).

## Chapitre 8

### Variables textuelles expliquant les gains en fluidité et en compréhension sur les trois années

Dans ce chapitre, nous présentons une analyse des variables textuelles qui pourront expliquer un gain de fluidité ou de compréhension entre la version simplifiée et la version originale. Les analyses ont été faites pour chaque niveau scolaire (CE1-CE2-CM1). La simplification de textes, littéraires et documentaires scientifiques, a été réalisée sur trois niveaux, lexical, syntaxique et discursif. Les analyses ont été réalisées sur les données de l'étude présentée dans le chapitre 6. Nous avons étudié trois séries de variables (Tableau 2, chapitre 4) qui distinguent les textes simplifiés des textes originaux : lexicales, morphologiques et syntaxiques. Nous avons fait l'hypothèse que les facteurs de longueur de mots, de fréquence, de suppression de pronominalisation et de morphologie verbale seraient les facteurs les plus discriminants entre les versions originales et simplifiées parce que ce sont les variables qui ont le poids le plus important dans les formules de lisibilité (chapitre 4).

#### 8.1 Variables textuelles qui prédisent les gains en fluidité et en compréhension

Afin de déterminer les variables textuelles qui prédiraient les gains en fluence et en compréhension de lecture, les scores de gains ont été calculés pour chaque texte et pour chaque variable textuelle choisie. Le gain de simplification absolu a été obtenu en soustrayant la moyenne du texte original de la moyenne de la version simplifiée [Gain de simplification =  $X_{SIMP} - X_{ORIG}$ ]. Nous avons utilisé ici des scores de gain absolu car les distributions sont meilleures pour les scores de gain absolu que pour les scores de gain relatif. Pour chacune des variables textuelles et pour chaque texte, nous avons obtenu une différence entre un texte simplifié et un texte original sur une variable donnée. Nous avons procédé ainsi pour les 22 variables textuelles décrites dans le Tableau 2, chapitre 4. Nous avons corrélé sur l'ensemble des 20 textes lus chaque année les scores de différence des variables textuelles avec les scores de différence des gains de simplification pour la fluence et pour la compréhension. Les résultats pour les 22 variables textuelles sont présentés dans le Tableau 13. En raison du petit nombre de textes lus chaque année, toutes les corrélations ont été inspectées visuellement et les valeurs aberrantes (outliers) ont été exclues. Le nombre final de textes utilisés dans chaque corrélation est présenté dans le Tableau 13.

Tableau 13 : Corrélations entre les variables textuelles [XSIMP - XORIG] et les gains de simplification pour tous les textes en termes de vitesse de lecture et de compréhension.

Variables textuelles [XSIMP – XORIG] CE1-CE2-CM1	Nbre de textes	Gains en vitesse de lecture par mot	Gains en compréhension
<b>Variables lexicales</b>			
75 <sup>ème</sup> percentile de la distribution des fréquences des lemmes			
CE1	19	,128	,514*
CE2	20	,131	-,077
CM1	19	,430	,352
90 <sup>ème</sup> percentile de la distribution des fréquences des lemmes			
CE1	19	-,100	,268
CE2	20	,234	-,008
CM1	19	-,131	,023
Fréquence moyenne des adjectifs (lemme)			
CE1	20	,022	,193
CE2	20	,033	-,150
CM1	20	,411	,456*
Médiane de la distribution des fréquences des adjectifs (lemmes)			
CE1	19	,645**	,009
CE2	19	-,017	-,433
CM1	20	,460*	,386
Proportion de mots de plus de 8 lettres			
CE1	20	,007	,179
CE2	20	-,585**	-,134
CM1	20	-,304	,301

Chapitre 8  
Variables textuelles expliquant les gains en fluidité et en compréhension sur les trois années

Proportion de mots absents de la liste de 2 000 mots de Gougenheim			
CE1	20	-,071	-,462*
CE2	20	-,453*	,001
CM1	18	-,351	,235
<b>Variables morphologiques</b>			
Proportion de verbes dans la forme future			
CE1	19	,063	-,380
CE2	18	,328	,490*
CM1	20	-,344	-,432
Proportion de verbes dans la forme du participe passé			
CE1	20	-,330	,081
CE2	20	-,235	-,096
CM1	20	-,229	-,186
Proportion de verbes au présent de l'indicatif			
CE1	19	-,018	-,052
CE2	20	,031	-,036
CM1	20	,162	,268
Proportion de verbes dans la forme du participe présent			
CE1	18	,214	-,023
CE2	20	,021	-,265
CM1	20	,099	,056
<b>Variables syntaxiques</b>			
75 <sup>ème</sup> percentile de la distribution du nombre de mots par phrase			
CE1	20	-,314	,074
CE2	20	-,375	-,214
CM1	20	-,124	-,092

Chapitre 8  
Variables textuelles expliquant les gains en fluidité et en compréhension sur les trois années

90 <sup>ème</sup> percentile de la distribution du nombre de mots par phrase			
CE1	19	,151	-,197
CE2	20	-,597**	-,352
CM1	20	-,149	-,196
Nombre médian de mots par phrase			
CE1	20	-,410	-,017
CE2	20	-,374	-,200
CM1	19	-,341	-,159
Nombre moyen de mots par phrase			
CE1	20	-,395	,079
CE2	20	-,552*	-,289
CM1	20	-,106	-,066
Rapport entre les prépositions et l'ensemble des articles			
CE1	20	,590**	,040
CE2	20	-,048	-,323
CM1	20	-,023	,007
Proportion de pronoms dans le texte			
CE1	20	-,156	-,261
CE2	20	,312	-,018
CM1	19	-,602**	-,437
Rapport des articles par rapport à l'ensemble des mots grammaticaux			
CE1	20	-,454*	,082
CE2	20	-,079	,279
CM1	20	,323	,315

Chapitre 8  
Variables textuelles expliquant les gains en fluidité et en compréhension sur les trois années

Rapport entre les noms propres et tous les mots lexicaux			
CE1	20	-,153	,224
CE2	20	,131	-,139
CM1	18	,401	,171
Rapport entre les noms propres et les adjectifs			
CE1	20	-,177	,377
CE2	20	,195	-,030
CM1	20	-,004	-,159
Rapport entre les noms propres et les pronoms			
CE1	20	,137	,607**
CE2	20	-,107	-,349
CM1	19	,579**	,528*
Rapport entre les pronoms et les mots lexicaux			
CE1	20	-,104	-,262
CE2	20	,301	-,025
CM1	19	-,661**	-,465*
Rapport entre les pronoms et les noms communs			
CE1	20	-,177	-,220
CE2	20	,235	-,082
CM1	19	-,476*	-,359

## 8.2 Discussion

En ce qui concerne les gains en fluence, 11 variables textuelles ont été associées de manière significative aux gains de vitesse par mot.

Au niveau des variables lexicales, les variables explicatives sont la médiane de la distribution des fréquences des adjectifs ( $\rho = ,645$ , pour les CE1 ;  $\rho = ,460$ , pour les CM1) et la proportion des mots de plus de 8 lettres ( $\rho = - ,585$ , pour les CE2). Pour les mots de plus de 8 lettres une corrélation négative exprime le fait que la baisse de mots longs est une aide pour la vitesse de lecture par mot (Gala & Ziegler, 2016). De même la proportion des mots absents de la liste de 2000 mots de Gougenheim ( $\rho = - ,453$ , pour les CE2) présente une corrélation négative car les textes simplifiés contenaient moins de mots absents de cette liste de fréquence des mots français. Par conséquent, si dans le texte simplifié il y a un plus grand nombre de mots absents de la liste de Gougenheim cela signifie que les mots du texte simplifié sont plus fréquents et permettent un gain en vitesse de mots lus par seconde. Ces résultats rejoignent ceux de Rello et al., (2013b).

Au niveau des variables morphologiques, aucune des quatre variables observées n'explique un gain en vitesse de lecture.

Au niveau des variables syntaxiques qui ont expliqué le gain en vitesse de lecture, deux sont significatives : le 90<sup>ème</sup> percentile de la distribution du nombre de mots par phrase ( $\rho = - ,597$ , pour les CE2) et le nombre moyen de mots par phrase ( $\rho = - ,552$ , pour les CE2). Ces corrélations suggèrent que plus les phrases sont courtes, plus la fluence progresse.

Au niveau des ratios des parties du discours, six variables sont significatives sur les huit variables retenues mais pas sur les trois années. Les deux qui n'expliquent pas le gain sur aucune des trois années sont : le rapport entre les noms propres et les mots lexicaux (*i.e.* noms, verbes, adverbes et adjectifs) et le rapport entre les noms propres et les adjectifs. Le rapport entre les prépositions et l'ensemble des articles ( $\rho = ,590$ ) est significatif uniquement pour le CE1, et de façon positive. Il semblerait que l'ajout de prépositions accélère la vitesse de lecture et ce constat est différent des attentes (Crossley, Louwerse, McCarthy & McNamara, 2007). Cependant nous constatons cette augmentation uniquement sur le CE1. Pour le CE2 et le CM1, la corrélation est négative mais non significative, ce qui montre que le nombre des prépositions a diminué entre les versions originales et simplifiées. Toutefois pour les quatre variables relatives aux pronoms, la proportion de pronoms dans le texte ( $\rho = - ,602$ , pour le CM1), le rapport entre les pronoms et les noms communs ( $\rho = - ,476$ , pour le CM1), le rapport entre les

pronoms et les mots lexicaux ( $\rho = - ,661$ , pour les CM1) ainsi que le rapport entre les noms propres et les pronoms ( $,579$ , pour le CM1), suggèrent que le remplacement des pronoms permet une augmentation de la vitesse de lecture et notamment avec le remplacement des pronoms par leur référent (Brouwers et al, 2014).

En ce qui concerne la compréhension, nous avons constaté que six variables textuelles expliquaient les gains mais pas de façon constante sur les trois années.

Au niveau des variables lexicales, trois variables ont des corrélations significatives avec les gains en compréhension de lecture : une corrélation négative avec le nombre de mots absents de la liste des 2000 mots les plus fréquents en français ( $\rho = - ,462$ , pour le CE1), le 75<sup>ème</sup> percentile de la distribution de fréquence des lemmes ( $\rho = ,514$ , pour le CE1) et la fréquence moyenne pour les adjectifs ( $\rho = ,456$ , pour les CM1). Ces résultats montrent que les textes simplifiés produisent des gains de compréhension car ils ont tendance à contenir des mots plus fréquents. Plus les mots sont fréquents et connus, plus la compréhension est facilitée (Perfetti & Stafura, 2014).

Au niveau des variables morphologiques, le futur fait ressortir une corrélation ( $\rho = ,490$ , pour les CE2) mais nous supposons que ces résultats sont un artéfact car la corrélation entre les textes originaux et les textes simplifiés pour les CE2 n'est pas significative. De plus une vérification manuelle a été effectuée afin de valider l'hypothèse de l'artéfact.

Au niveau des variables syntaxiques, aucune des quatre variables spécifiques au nombre de mots par phrase n'est significative et ne peut expliquer un gain en compréhension.

Pour les variables relatives aux parties du discours, deux variables affichent des corrélations significatives pour le gain en compréhension : le rapport entre les pronoms et les mots lexicaux ( $\rho = - ,465$ , pour le CM1) et le rapport entre les noms propres et les pronoms ( $\rho = ,607$ , pour le CE1) et ( $\rho = ,528$ , pour le CM1), sur deux années. Ces résultats mettent en évidence que ce sont les textes qui privilégient les noms propres par rapport aux pronoms et les textes dans lesquels les pronoms ont été remplacés par les mots lexicaux, donc leur référent, qui favorisent la compréhension en lecture. Ces résultats correspondent au remplacement dans les textes simplifiés des pronoms par leur référent et sont conformes à la littérature (Lima & Bianco, 1999 ; Martins & Le Bouédec, 1998).

Les résultats sont à prendre avec précaution car les corrélations que nous avons observées pour les gains de vitesse et de compréhension varient en fonction des trois années. De plus, les corrélations sont basées sur un nombre limité de textes (20) ce qui réduit leur fiabilité.

En résumé, ce chapitre met en avant l'importance des variables lexicales et syntaxiques dans les gains attendus en fluidité et en compréhension, de par la simplification. C'est pourquoi nous avons poursuivi nos investigations avec les trois types de simplifications : lexicale, syntaxique et lexico-syntaxique.

## Chapitre 9

### Effets de la simplification : lexicale, syntaxique et lexico-syntaxique chez les lecteurs débutants

Le but de cette deuxième étude était de comparer les effets de la simplification de textes sur plusieurs niveaux : lexical versus syntaxique versus lexico-syntaxique. Nous souhaitons déterminer si un type de simplification était plus efficace par rapport aux deux autres, notamment entre le lexical et le syntaxique pour la fluidité et la compréhension. Nous avons fait l'hypothèse que le facteur lexical serait aussi important que le facteur syntaxique et que la simplification lexico-syntaxique serait la plus efficace (Safari & Mohaghegh Montazeri, 2017).

#### 9.1 Population

Nous avons réalisé cette expérimentation auprès d'élèves de CE1. La moyenne d'âge était de 90,55 mois. Tous les élèves étaient scolarisés dans les 4 écoles présentées dans la section 6.1 sur les trois communes : Sanary sur Mer, La Cadière d'Azur, Saint-Cyr sur Mer. Nous avons constitué de façon aléatoire trois groupes d'élèves :

Le premier groupe (Lex) comportait 118 élèves qui ont lu des textes originaux et des textes simplifiés uniquement au niveau lexical, durant l'année scolaire 2017-2018

Le deuxième groupe (Syntax) comportait 51 élèves qui ont lu des textes originaux et des textes simplifiés uniquement au niveau syntaxique, durant l'année 2018-2019

Le troisième groupe (LexSyntax) comportait 83 élèves qui ont lu des textes originaux et des textes simplifiés sur deux niveaux lexical et syntaxique, durant l'année scolaire 2018-2019.

#### 9.2 Design expérimental

Nous avons utilisé 10 textes, cinq littéraires et cinq documentaires scientifiques originaux, qui ont été simplifiés avec une simplification uniquement lexicale ou uniquement syntaxique ou lexico-syntaxique. Chaque élève a lu autant de textes originaux que de textes simplifiés avec le même type de simplification. Ainsi nous avons pu analyser les effets de la simplification en intra sujet et les effets des trois types de simplification en inter-groupes. Dans les classes participant à l'expérimentation, nous avons contrebalancé l'ordre de lecture des textes littéraires et des documentaires scientifiques ainsi que les textes originaux et les textes

simplifiés en fonction de la simplification testée par chacun des groupes. Chaque texte a été lu dans une version simplifiée et originale mais pas par les mêmes élèves. La collecte des données a été réalisée en "temps réel" par WIFI comme dans les chapitres précédents.

### 9.3 Tests en individuel

Chaque élève au cours du deuxième trimestre (février-mars) a passé deux tests en classe afin d'évaluer son niveau de compétence en reconnaissance visuelle des mots, en morphologie et en grammaire. Ces tests nous ont permis de connaître le niveau initial de chacun des groupes.

#### Reconnaissance visuelle des mots

Nous avons utilisé un test standardisé de reconnaissance visuelle, le Test d'identification de mots écrits (Timé 3) développé par Écalle (2006), décrit dans la section 6.3.

#### Orthographe

Nous avons utilisé un test normalisé d'orthographe du français, le Langage oral - Langage écrit - Mémoire, Attention test (L2MA, 2ème édition ; Chevrie-Muller et al., 2010), décrit dans la section 6.3.

Nous avons réalisé une ANOVA avec le groupe comme facteur inter-sujets pour chaque test afin de vérifier si les groupes avaient un niveau équivalent (Tableau 14).

Tableau 14 : Description des variables pour chacun des tests et pour chaque groupe.

Variables	Min-Max	Moyenne	SD	Skewness	Kurtosis
<b>Reconnaissance visuelle des mots (/40)</b>					
Groupe 1 (Lex)	9-33	17,64	5,34	0,561	-0,284
Groupe 2 (Syntax)	8-27	16,59	4,45	0,321	0,094
Groupe 3 (LexSyntax)	6-29	17,23	4,97	0,158	-0,114
<b>Dictée : Morphologie (/45)</b>					
Groupe 1 (Lex)	12-45	37,55	6,30	-1,598	3,291
Groupe 2 (Syntax)	16-45	36,25	6,47	-1,182	1,264
Groupe 3 (LexSyntax)	2-45	34,96	9,34	-1,266	1,254
<b>Dictée : Grammaire (/31)</b>					
Groupe 1 (Lex)	1-28	14,31	5,78	0,346	-0,794
Groupe 2 (Syntax)	2-24	13,55	6,11	0,208	-1,131
Groupe 3 (LexSyntax)	0-28	13,14	5,79	0,170	-0,315

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

Nous avons comparé la performance des trois groupes pour les tests choisis. En ce qui concerne la reconnaissance visuelle des mots, il n'y a pas de différence significative entre les trois groupes ( $F(2,250) = 0,788, p = ,456$ ). De même pour la dictée pour laquelle nous n'avons pas trouvé de différences entre les groupes ni pour la partie morphologique ( $F(2,250) = 2,940, p = ,055$ ) ni pour la partie grammaire ( $F(2,250) = 1,004, p = ,368$ ). Toutefois, le groupe 3 est légèrement inférieur sur le test morphologique avec un effet de groupe marginalement significatif.

#### 9.4 Résultats en fluidité et en compréhension

Les résultats bruts ont été traités pour les deux facteurs, vitesse de lecture en nombre de mots lus par seconde et compréhension (QCM). Les prétraitements en vitesse de lecture de mots lus par seconde ont été identiques à l'étude 1, comme explicité dans la section 6.5. Les vitesses de lecture en nombre de mots lus par seconde ont été utilisées dans toutes les analyses ultérieures.

Nous avons calculé la moyenne des réponses correctes aux QCM à la suite de la lecture des textes pour chaque élève. Les élèves ont répondu à 5 questions après chaque texte littéraire et documentaire scientifique avec 3 propositions de réponses. Chaque élève a répondu à 50 questions.

##### 9.4.1 Vitesse de lecture de mots

Les résultats en vitesse de lecture de mots lus par seconde ont été analysés en utilisant une ANOVA trois-factorielles avec types de texte (littéraire vs scientifique) et simplification (originale vs simplifiée) comme facteur intra sujet et groupes (lexicale, syntaxique et lexico-syntaxique) comme facteur inter-sujets (Tableau 15).

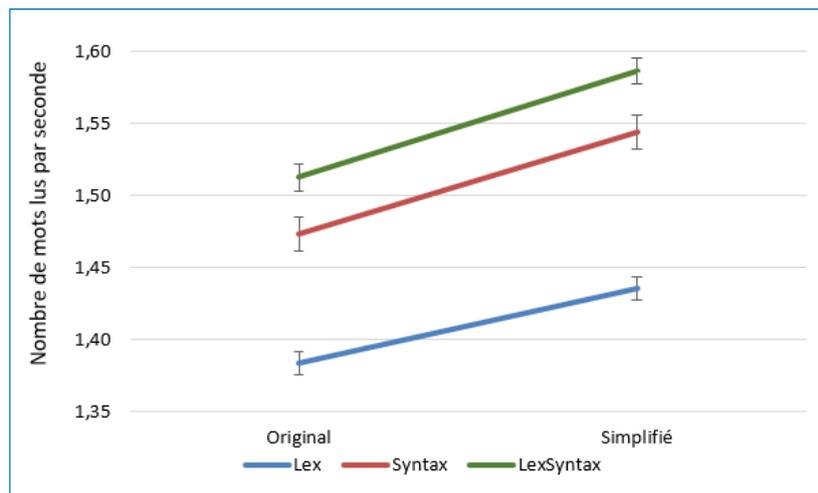
Tableau 15 : Résultats de l'ANOVA Simplification\*Type de textes \*Groupe pour la variable nombre de mots lus par seconde.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 251	33,888	,000	,120	***
Texte	1, 251	1,028	,312	,004	
Texte*Groupe	2, 250	4,999	,007	,039	**
Simplification*Groupe	2, 250	0,462	,631	,004	
Simplification*Texte	1, 251	15,566	,000	,059	***
Texte*Simplification*Groupe	2, 250	0,037	,964	,000	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

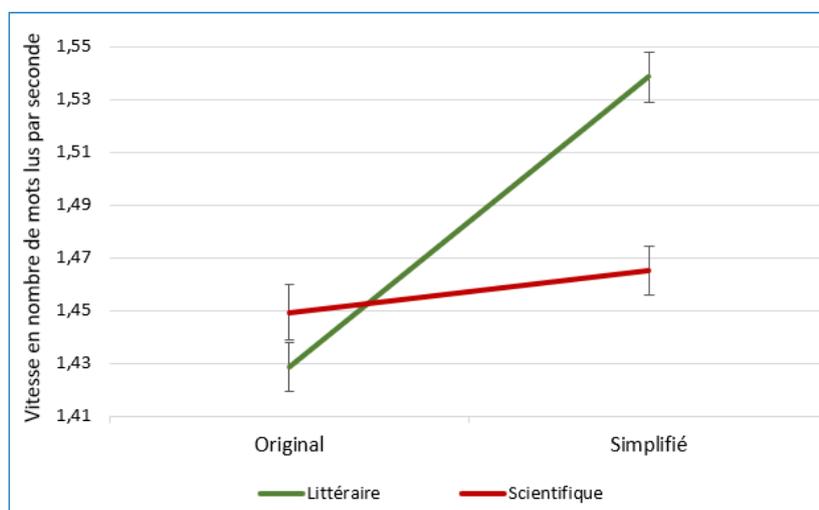
Les analyses montrent un effet fortement significatif de la simplification mais pas d'interaction significative de la simplification avec groupe ce qui suggère que les effets de la simplification ne varient pas en fonction du facteur groupe, donc de la nature de la simplification (Figure 26).

Figure 26: Interaction non-significative de la simplification en fonction des 3 groupes de simplification (Lexicale, Syntaxique, Lexico-syntaxique) pour la fluidité de lecture. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Nous avons obtenu une interaction significative entre la simplification et le type de texte, qui montre que les effets de la simplification ont été différents en fonction du type de texte avec un effet plus important pour les textes littéraires que pour les documentaires scientifiques (Figure 27). Enfin, l'interaction significative entre le facteur groupes et le type de texte indique que les trois groupes (Lex, Syntax, LexSyntax) n'ont pas lu à la même vitesse selon le type de texte.

Figure 27 : Effet de la simplification en fonction du type de textes au niveau de la fluidité de lecture. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



### 9.4.2 Résultats en compréhension : Questions à Choix Multiples

Les résultats en compréhension (QCM) ont été analysés en utilisant une ANOVA trois-factorielles avec types de texte (littéraire vs scientifique) et simplification (originale vs simplifiée) comme facteur intra sujet et groupes (lexicale, syntaxique et lexico-syntaxique) comme facteur inter-sujets (Tableau 16).

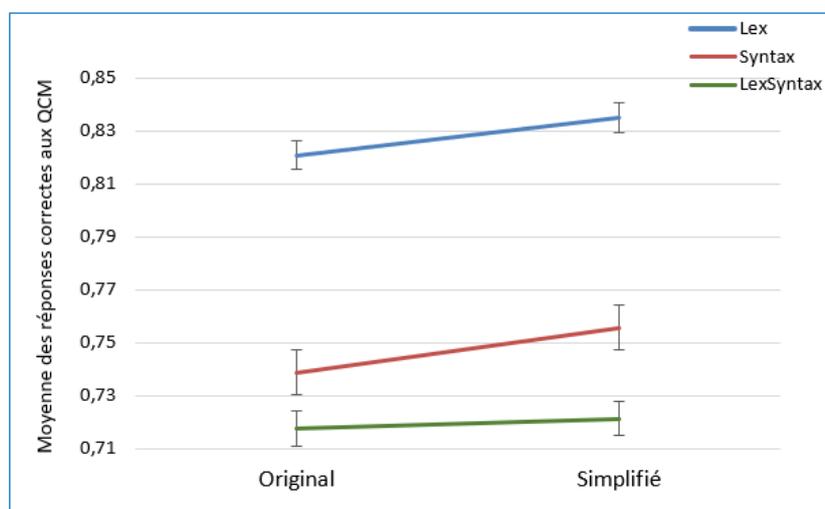
Tableau 16 : Résultats de l'ANOVA Simplification\*Type de textes \*Groupe pour la variable compréhension.

Effet	df	F	p	$\eta_p^2$	sign
Simplification	1, 251	2,162	,143	,009	
Texte	1, 251	9,491	,002	,037	**
Texte*Groupe	2, 250	3,821	,023	,030	*
Simplification*Groupe	2, 250	0,254	,776	,002	
Simplification*Texte	1, 251	5,323	,022	,021	*
Texte*simplification*Groupe	2, 250	0,540	,583	,004	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

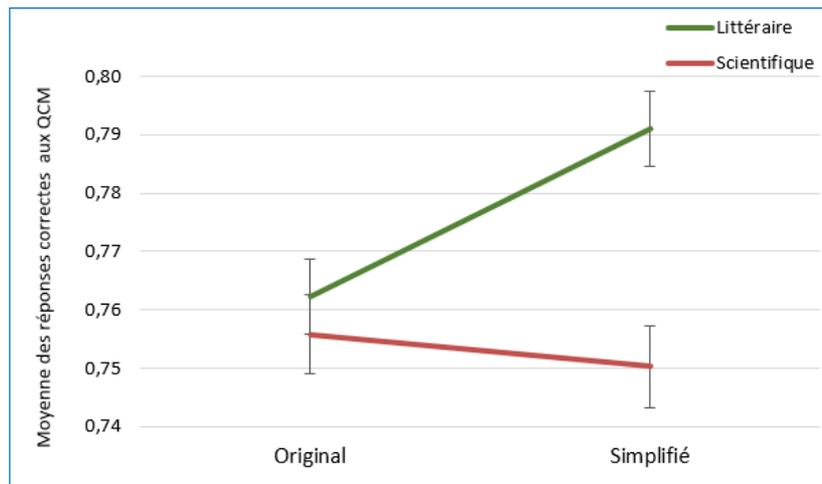
Les analyses des scores en compréhension montrent que l'effet principal de la simplification n'était pas significatif, ce qui indique que la simplification n'a eu aucun effet global sur la compréhension en lecture. L'interaction entre l'effet de la simplification en fonction des groupes n'était pas significative non plus, ce qui suggère qu'aucun groupe n'a montré un effet de simplification (Figure 28).

Figure 28 : Interaction non significative de la simplification en fonction des 3 niveaux de simplification : lexical, syntaxique et lexico-syntaxique en compréhension. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



En revanche, nous avons obtenu une interaction significative entre la simplification et le type de texte. Cette interaction reflète le fait que les effets de la simplification ont été différents suivant le type de texte avec un effet positif pour les textes littéraires mais négatif pour les documentaires scientifiques, sans être significatif (Figure 29).

Figure 29 : Effet de la simplification en fonction du type de texte au niveau de la compréhension. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



L'effet de la simplification étant différent suivant le type de textes, nous avons voulu savoir s'il y avait un effet significatif de la simplification pour les textes littéraires et pour les documentaires scientifiques. Nous avons réalisé une ANOVA pour chaque type de texte, en ayant comme seul facteur intra sujet la simplification.

Nous avons obtenu un effet significatif de la simplification pour les textes littéraires ( $F(1,251) = 7,773$ ,  $p = ,006$ ) qui indique que la simplification a été efficace sur les textes littéraires au niveau de la compréhension. L'interaction simplification par groupe n'est pas significative ( $F(2,250) = 0,616$ ,  $p = ,541$ ) ce qui indique que la simplification au niveau des textes littéraires n'est pas significativement différente pour les trois groupes. Pour les documentaires scientifiques, nous n'avons pas obtenu d'effet de la simplification ( $F(1,251) = 0,241$ ,  $p = ,624$ ) ce qui montre que la simplification n'est pas efficace de manière significative pour la compréhension. De plus, l'effet de la simplification n'est pas significatif ( $F(2,250) = 0,200$ ,  $p = ,818$ ) en fonction des groupes, ce qui montre que les trois types de simplification (Lex, Syntax et LexSyntax) sont identiques pour la compréhension.

## 9.5 Discussion

Dans cette deuxième étude, nous avons comparé les effets de la simplification de textes sur plusieurs niveaux : lexical, syntaxique, lexico-syntaxique auprès d'élèves en CE1 afin d'évaluer si un niveau de simplification était plus efficace par rapport aux deux autres. Nous avons fait l'hypothèse que le facteur lexical serait aussi important que le facteur syntaxique et que la simplification lexico-syntaxique serait la plus efficace (Safari & Mohaghegh Montazeri, 2017).

Concernant la fluidité en lecture, nous avons eu un effet significatif de la simplification mais aucun effet significatif de la simplification par groupe ce qui indique que les trois niveaux de simplification (Lex, Syntax et LexSyntax) ont permis un gain en fluidité de lecture. Ce constat suggère que la simplification uniquement lexicale ou uniquement syntaxique est déjà bénéfique chez le lecteur débutant. Ces résultats correspondent aux résultats de l'étude de Nandiegou et Reboul (2018) auprès de 21 enfants dyslexiques, faibles lecteurs ainsi qu'à ceux de Rello et ses collaborateurs (2013a) auprès de 96 adultes. Le gain en fluidité de lecture a été plus important sur les textes littéraires que sur les documentaires scientifiques. Les travaux de Sheehan et al. (2013) en lisibilité montrent que les textes littéraires, incluant plus de vocabulaire de base de la langue, sont plus facilement adaptables aux besoins des lecteurs. De plus, le vocabulaire spécifique des documentaires scientifiques, souvent inconnu des jeunes lecteurs, reste plus difficilement modifiable car le vocabulaire dans les textes scientifiques est spécialisé et demande des connaissances antérieures précises (Marin, Crinon, Legros & Avel, 2007).

Concernant la compréhension, la simplification était significative au niveau du texte littéraire mais pas au niveau du texte scientifique. Nous n'avons pas trouvé d'effet significatif entre la simplification et les groupes. Cela peut aussi être expliqué par l'effectif des trois groupes. Nous rappelons que le groupe 1 qui a lu les textes simplifiés au niveau lexical comptait 118 élèves alors que le groupe 2 qui a lu des textes simplifiés au niveau syntaxique comptait 83 élèves et que le groupe 3 avait 51 élèves.

Pour construire le sens d'un texte, un enfant doit connaître le sens des mots mais également posséder des connaissances syntaxiques et s'appuyer sur la structure de phrases (Bentolila, 2019). Les interactions significatives entre le type de texte par groupe et le type de texte par la simplification ainsi que l'effet principal par type de textes suggèrent que la compréhension des textes scientifiques est plus difficile pour les jeunes lecteurs. La nature du texte influence la façon dont le lecteur va intégrer les informations explicites du texte, c'est à

dire faire appel à ses connaissances antérieures (Wolfe, 2005). Dans le texte informatif, la complexité conceptuelle des mots et leur mise en relation rend plus difficile la compréhension que dans les textes narratifs (Hiebert & Cervetti, 2012). Nos résultats sont légèrement différents de l'étude menée en langue anglaise par Safari et Mohaghegh Montazeri (2017) qui ont comparé les performances en compréhension écrite de trois groupes d'adultes comparables en niveau de lecture. Le groupe ayant bénéficié des simplifications lexicales et syntaxiques combinées a obtenu les meilleures performances, de manière significative, en comparaison des deux autres groupes, alors qu'aucune différence significative de performance n'existait entre les groupes ayant bénéficié uniquement de la simplification lexicale ou uniquement de la simplification syntaxique. Le bénéfice apporté par la simplification au niveau de la fluidité a été plus important pour les textes littéraires que pour les textes scientifiques. De plus, au niveau de la compréhension en lecture le bénéfice de la simplification est apparu uniquement pour les textes littéraires.

## Chapitre 10

### Étude d'entraînement : la simplification peut-elle durablement influencer les performances de lecture ?

Nous avons vu dans les chapitres 6 et 7 que la simplification était bénéfique en particulier pour les faibles lecteurs. Dans cette troisième étude, nous nous sommes intéressés aux effets de généralisation de la lecture de textes simplifiés. La question était de savoir si un élève, notamment s'il s'agit d'un faible lecteur, pourrait augmenter son niveau de fluidité ou de compréhension après un entraînement de quelques mois avec des textes simplifiés uniquement. L'hypothèse sous-jacente était que la lecture de textes simplifiés permettrait à des faibles lecteurs de mieux entrer dans le cercle vertueux de l'auto-apprentissage (Share, 1995) et par ce biais d'augmenter leurs capacités et stratégies de compréhension de manière générale (effet de transfert ou de généralisation). Pour tester cette hypothèse, nous avons proposé un entraînement pendant trois mois à trois groupes d'élèves de CE2 : un groupe qui n'a lu que des textes simplifiés, un groupe qui a lu ces mêmes textes uniquement en version originale et un groupe témoin (« business-as-usual ») qui a fait les mêmes pré et post-tests sans bénéficier d'un entraînement spécifique avec nos textes.

#### 10.1 Population

Nous avons travaillé avec les CE2 des 4 écoles décrites dans la section 6.1 plus deux autres classes dans deux écoles différentes sur la commune du Beausset. Le Beausset est une commune du département du Var (83330) comptant 8970 habitants dont les revenus sont supérieurs à 32800 euros par ménage. La commune du Beausset compte trois groupes scolaires: deux écoles maternelles, deux écoles élémentaires et un collège.

L'ensemble des élèves de CE2 qui ont participé à cette expérience avaient une moyenne d'âge de 102,78 mois et étaient répartis dans 11 classes car certaines classes étaient des cours doubles et ne comptaient par exemple que sept élèves du niveau CE2. L'effectif total était de 188 élèves pour cette expérience.

## 10.2 Design expérimental

### Méthodologie

L'ensemble des 188 élèves ont lu deux textes en pré-tests puis deux textes en post tests, chaque test étant suivi d'un QCM avec 10 questions de compréhension. Ces tests nous ont servi à évaluer si l'entraînement avec des textes simples versus originaux avait un effet bénéfique (effet de transfert) sur la lecture de textes plus longs ou plus compliqués. Les textes choisis pour les pré-tests et les post-tests étaient tirés du fichier, *Le Moniteur de lecture CE2*, ensemble de textes permettant une évaluation des compétences en lecture des élèves en fonction des périodes de l'année scolaire et de deux types de textes, « fictionnel et technique », selon les auteurs. Chaque élève a aussi passé des tests cognitifs, langagiers et psycholinguistiques, en complémentarité des tests de lecture. Nous avons réparti les élèves en trois groupes expérimentaux de façon aléatoire afin de constituer des groupes équivalents au niveau de leur effectif. Le groupe témoin (TEM) a participé aux pré-tests et aux post-tests, les élèves n'ont pas lu les textes originaux ou simplifiés de notre corpus sur tablette mais l'enseignement délivré en classe leur permettait de travailler la compréhension en lecture avec des outils choisis par l'enseignant, y compris les outils nécessitant l'utilisation de la tablette. Le groupe TEM comprenait 62 élèves. Le groupe (ORIG) qui a lu les 24 textes en version originale a eu trois mois pour lire les textes, répondre aux questions et compléter les résumés après chaque lecture. Le groupe ORIG comprenait 61 élèves. Chaque texte était suivi de 5 questions ayant chacune 4 propositions de réponses et un résumé à trous. Le groupe (SIMP) a lu les mêmes textes en version simplifiée sur trois niveaux (lexical, syntaxique et discursif) dans un design expérimental similaire au groupe ORIG. Le groupe SIMP comprenait 65 élèves.

### Tests de lecture avant et après l'entraînement

Le pré-test était basé sur la lecture de deux textes, un littéraire et un texte informatif et chaque texte était suivi de 10 questions de compréhension. L'évaluation de la compréhension s'est faite sous une forme de QCM et chaque question avait trois propositions de réponses. Nous avons choisi des textes extraits du fichier « *Le moniteur de lecture CE2* », Nathan-Retz, 1989, inscrits sur la première et la troisième période de l'année scolaire.

« Cet outil était initialement destiné aux maîtres dans la classe pour identifier et évaluer les performances de ses élèves en matière de lecture. [...]. Les capacités de compréhension du texte sont évaluées au niveau global du texte comme au niveau d'une saisie d'informations fines. » (Bentolila, & Richaudeau, 1989, p.2-3).

Les auteurs de cet outil d'évaluation des compétences en lecture ont précisé que la population expérimentale qui a participé à la création de l'outil était de 450 élèves (CE2, CM1 et CM2). Les résultats des élèves ont permis de juger de la difficulté relative des textes proposés dans l'outil et de la pertinence de l'ensemble des grilles de questionnement attaché à ces textes. Le livre de perfectionnement de lecture appelé A.R.T.H.U.R. a ensuite été conçu pour apporter des solutions spécifiques après les évaluations proposées aux enseignants dans « *Le moniteur de lecture* ». Les propositions d'activités de lecture faites dans A.R.T.H.U.R. avaient toutes pour objectif le perfectionnement de la lecture avec des temps réservés au décodage et à la vitesse de lecture, aux stratégies de compréhension au travers d'ateliers de questionnement de textes, à la maîtrise de la langue, vocabulaire et connaissances linguistiques. Dans notre expérimentation, les élèves ont lu, en pré-tests, « *Nicolas a des ennuis* » pour le texte littéraire et « *Comment écrit-on en japonais ?* » pour le texte informatif. Les élèves ont lu, en post-tests, « *Rira bien qui rira le dernier* » pour le texte littéraire et « *La lune est-elle habitée ?* » pour le texte informatif. Des tests en individuel ont été proposés à l'ensemble des élèves en amont des pré-tests.

### **Les trois groupes d'élèves de l'expérimentation**

Le groupe témoin (TEM), comprenait des élèves qui ont réalisé la lecture et répondu aux questions des textes servant de pré-tests et de post-tests mais qui n'ont pas eu accès aux 69 textes simplifiés ou originaux du corpus via l'iBook. Ces élèves ont suivi un enseignement en classe respectant les programmes avec un entraînement à la lecture et à la compréhension, choisi par l'enseignant via des supports et dispositifs différents. Leurs professeurs avaient entre 10 et 15 ans d'expérience. Ces enseignants ont déclaré utiliser le rallye lecture qui se réalise en ligne, et/ou l'atelier a.r.t.h.u.r, et/ou Lectorino-Lectorinette, en format papier. Les classes utilisent tout ou partie de ces outils, avec travail sur les compétences de décodage, les connaissances de la langue, les compétences narratives, les compétences inférentielles. Ces élèves sont aussi équipés en iPad comme les deux autres groupes. Les enseignants ont pu travailler avec eux les différentes compétences en lecture et en compréhension via des applications numériques de leur choix ou sur support papier. Le groupe dit témoin comptait 62 élèves.

Le deuxième groupe (ORIG) était constitué d'élèves qui ont lu 24 textes originaux, 12 textes littéraires et 12 documentaires scientifiques. La programmation des lectures, choisie pour l'iBook, a été de donner à lire un texte littéraire suivi d'un documentaire scientifique à chaque élève. La progression des textes a été réalisée en fonction de l'évolution présumée des élèves entre mars et mai. Cette cohorte n'a lu que des textes originaux via l'application sur des

modalités pédagogiques choisies par leurs enseignants qui ont donc pu travailler les compétences liées à la lecture et à la compréhension comme ils le désiraient. Néanmoins, les enseignants ont déclaré que le temps imposé de trois mois pour lire les 28 textes proposés dans cette expérimentation les a contraints à modifier leurs habitudes de travail de la lecture et notamment à réduire le temps de travail sur l'explicitation des stratégies de compréhension. Les consignes données aux enseignants pour effectuer les lectures des textes originaux ont été les mêmes que dans le chapitre 6 ; les professeurs ne devaient pas intervenir jusqu'à ce que l'ensemble des élèves ait fini de lire le texte, de répondre aux questions et de compléter le résumé. Chaque lecture a été réalisée en individuel et en autonomie en situation classe, même lors des pré-tests et post-tests. Le groupe comprenait 61 élèves.

Le troisième groupe (SIMP) a suivi les mêmes conditions que le groupe 2 mais a lu les 24 textes en version simplifiée. Ce groupe a donc lu les même 24 textes, 12 littéraires et 12 documentaires scientifiques exactement dans le même ordre que le groupe 2. Les enseignants ont suivi le même protocole avec des règles identiques. Le groupe était constitué de 65 élèves.

### **10.3 Tests en individuel pour les trois groupes**

Nous avons utilisé 9 tests pour évaluer le profil cognitif et langagier individuel de chaque participant. Ils ont été réalisés dans le même ordre pour chaque élève, en deux séances.

#### **Lecture à haute voix**

Nous avons utilisé le test de la lecture à haute voix mais pour le niveau G3-G9 (Pourcin, Sprenger-Charolles, El Ahmadi & Colé, 2016). Les auteurs ont élaboré 4 catégories de mots familiers : des mots réguliers courts, des mots réguliers longs (15 items par catégorie) et des mots irréguliers courts, des mots irréguliers longs (10 items par catégorie). Le même design a été proposé pour la lecture de mots inventés (pseudo-mots) mais chacune des quatre catégories comprenait 10 items. Le nombre d'items dans chacune des deux listes (mots familiers et mots inventés) a été fixé de façon à permettre à la plupart des enfants de lire tous les items en moins d'une minute. Nous avons obtenu un score sur 50 pour les mots familiers et un score sur 40 pour les pseudo-mots. Nous avons calculé la vitesse de lecture moyenne pour ces deux tâches et calculé la vitesse moyenne globale pour l'ensemble des mots familiers et des mots inventés. Cette moyenne définit la variable nommée lecture à haute voix en mots par seconde dans le tableau 17.

### **Vocabulaire**

Le vocabulaire a été évalué à l'aide du Test de Vocabulaire Actif/Passif (TVAP), (Deltour & Hupkens, 1980), décrit dans la section 6.3.

### **Compréhension orale**

Pour évaluer la compréhension des informations syntaxiques et sémantiques, nous avons utilisé un test de français standardisé appelé Epreuve de Compréhension Syntaxico-Sémantique Etalonnée (ECOSSE) développé par Lecocq (1998). Dans ce test, une phrase orale est donnée à l'élève, qui doit choisir l'image correspondante parmi quatre possibilités, dont deux contiennent des pièges lexicaux ou grammaticaux. Le test est conçu pour utiliser une variété de structures syntaxiques, dont la complexité augmente au cours de l'épreuve. Nous n'avons utilisé que les 24 items (H23, L44, M24, N11, N44, O34, P22, P33, Q14, Q34, R22, R33, S32, S43, T34, T42, U12, U23, U34, V12, V21, V32, W34, W42). Le test a été informatisé, de sorte que chaque élève a été testé dans les mêmes conditions.

### **Mémoire de travail**

Les tests Digit Span (nombres répétés) ont été développés pour permettre une meilleure mesure de la mémoire de travail (Wechsler, 2004). Nous avons utilisé la répétition des nombres en avant avec 16 items et en arrière avec 16 items également. Les deux premiers éléments sont composés de deux chiffres et les deux derniers éléments de neuf chiffres. Les enfants devaient se rappeler oralement la série de chiffres dans le même ordre.

### **Intelligence non verbale**

Nous avons utilisé le test des matrices progressives de Raven (Raven 1998), décrit dans la section 6.3. Nous avons obtenu un score sur 36.

### **Morphologie**

Nous avons utilisé une tâche d'achèvement de phrase pour mesurer la sensibilité morphologique (Casalis et al., 2004). Dans ce test, les enfants devaient compléter une phrase en utilisant une forme dérivée d'un mot dont la racine était fournie dans la première partie de la phrase (par exemple, "celui qui colorie fait du ..." ... "coloriage"). Nous avons utilisé la forme suffixée qui contient 12 éléments. La même tâche a été réalisée avec des pseudo-mots, elle comporte également 12 items.

### **Reconnaissance visuelle des mots**

Nous avons utilisé un test standardisé de reconnaissance visuelle, le Test d'identification de mots écrits (Timé 3) développé par Écalle (2006), décrit dans la section 6.3.

## Fluence

Ce test provient d'une étude sur la compréhension dans le cadre de ANR Devcomp<sup>61</sup>, Bianco, Nardy, Lima, Colé et Megherbi, (2014). L'enfant doit lire le texte à haute voix. L'adulte arrête l'enfant à une minute de lecture. La moyenne de la fluence est à 103,54 en une minute. Nous avons utilisé le texte intitulé « Le voyage de Georges ».

## Test des Cloches

Le test des cloches a été élaboré par Gauthier, Dehaut et Joanette (1989) afin d'évaluer la déficience visuelle chez les adultes cérébro-lésés. Il est ici utilisé afin de mesurer les capacités visuo-attentionnelles des enfants dans le cadre d'une épreuve de recherche de cibles parmi des distracteurs. Le test consiste à présenter à l'enfant une feuille renfermant 112 dessins d'objets (scie, pomme, cheval, voiture, nuage...) dont 35 dessins de cloches. La tâche consiste pour l'enfant à barrer le plus possible de cloches pendant un temps de 2 minutes. Le score (nombre de cloches barrées en deux minutes) est pris en compte. Nous obtenons un temps de réaction en seconde en fonction du nombre de cloches trouvées et du temps de réalisation.

Nous avons réalisé les analyses statistiques pour les tests effectués par les 188 élèves des trois groupes (Tableau 17). Les résultats de chacun des tests, certains étant subdivisés, sont présentés pour l'ensemble des 188 élèves.

Tableau 17 : Statistiques descriptives pour les 9 tâches cognitives et langagières ainsi que les mesures de l'asymétrie et de l'aplatissement pour les trois groupes de CE2 (N=188).

Variables	Min-Max	Moyenne	SD	Skewness	Kurtosis
Lecture à haute voix mots familiers (mots par sec)	0,17-1,52	0,66	0,28	0,79	0,26
Lecture à haute voix mots inventés (mots par sec)	0,13-0,83	0,46	0,13	0,16	0,07
Lecture à haute voix (mots par sec)	0,15-1,11	0,56	0,19	0,47	0,01
Vocabulaire (/60)	36-58	48,81	4,23	-0,35	0,19
Compréhension orale (/24)	11-23	17,67	2,25	-0,07	-0,21
Mémoire de travail en avant (/16)	4-16	7,97	1,90	0,91	1,86
Mémoire de travail en arrière (/16)	4-10	5,76	1,13	0,73	0,75
Intelligence non verbale (/36)	12-35	25,07	4,81	-0,20	-0,53
Morphologie suffixés (/12)	3 -12	8,94	1,88	-0,83	0,76
Morphologie pseudo mots (/12)	0-10	3,05	2,04	0,65	0,15
Reconnaissance visuelle des mots (/40)	10-36	24,40	5,72	-0,37	-0,40
Fluence	34-186	99,43	29,39	0,44	-0,09
Test des cloches	0,16-0,31	0,24	0,03	-0,21	-0,46

<sup>61</sup> <https://pdessus.fr/devcomp/membres.html>

Nous avons analysé les différences entre les groupes à l'aide d'une ANOVA à un facteur avec les trois groupes pour chaque test cognitif et langagier. Il n'y avait pas de différence significative entre les groupes pour aucune des variables (tous les  $p > ,108$ ).

#### 10.4 Analyses des effets de l'entraînement

Les résultats pour la vitesse, en nombre de mots lus par seconde, ont été traités comme explicité dans la section 6.5. Concernant la compréhension, une moyenne des bonnes réponses aux QCM a été réalisée, pour chaque élève, à la suite de la lecture des textes. En pré-test et post-test les élèves ont répondu à 10 questions après la lecture de chaque texte littéraire et documentaire scientifique. Chaque enfant a donc répondu à 20 questions lors des pré-tests et lors des post-tests, nous avons calculé une moyenne des réponses correctes pour chaque type de texte.

##### 10.4.1 Fluence par mot

Les résultats ont été analysés en utilisant une ANOVA trois-factorielles avec *types de texte* (littéraire vs scientifique) et *gain* (pré vs post) comme facteur intra-sujet et *groupe* (TEM-ORIG-SIMP) comme facteur inter-sujets (Tableau 18).

Tableau 18 : Résultats de l'ANOVA trois factorielles (types de texte\*gain\*groupe) pour la variable nombre de mots lus par seconde.

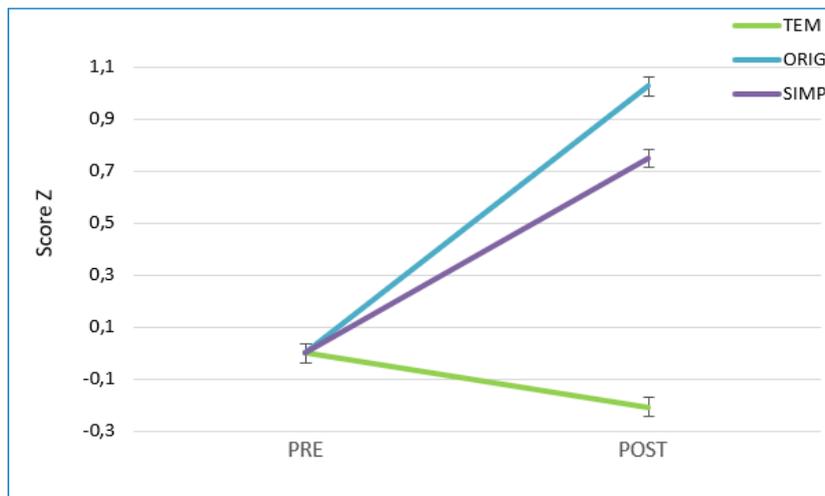
Effet	df	F	p	$\eta_p^2$	sign
Gain	1, 187	24,091	,000	,115	***
Texte	1, 187	34,410	,000	,157	***
Texte*Groupe	2, 186	0,774	,463	,008	
Gain*Groupe	2, 186	12,151	,000	,116	***
Gain*Texte	1, 187	30,831	,000	,143	***
Texte*Gain*Groupe	2, 186	0,961	,384	,010	

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\* $p < 0,001$

Les résultats de l'ANOVA montrent un gain significatif entre pré- et post-test et une interaction significative entre gain et groupe. Comme le montre la Figure 30, cette interaction reflète le fait que seuls les groupes qui ont lu nos textes sur tablette (ORIG et SIMP) ont progressé en fluidité entre le pré et le post test. Les analyses de contrastes montrent que la différence entre ces deux groupes (ORIG et SIMP) n'était pas significative ( $F(1,125) = 0,817$ ,  $p = ,368$ ). Le groupe

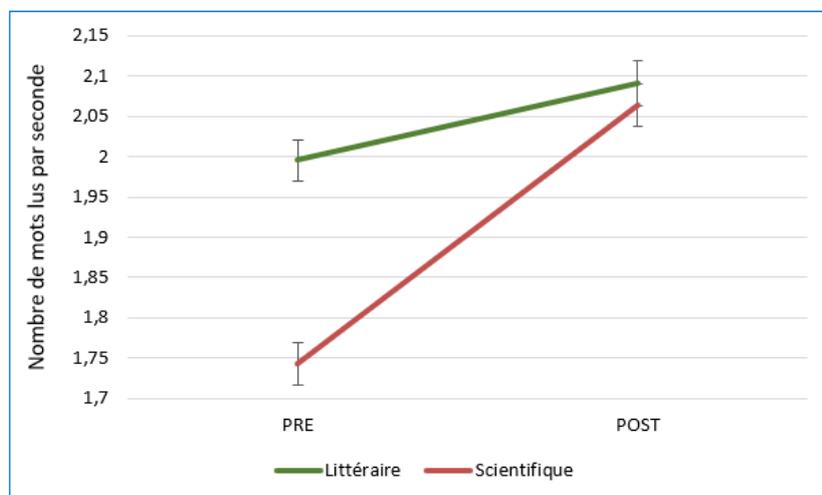
témoin a significativement moins progressé en fluidité que les deux groupes qui ont lu les textes sur tablette ( $ps < ,0001$ ).

Figure 30 : Taille d'effet en score Z de l'entraînement (pré-test-post-test) pour les gains en fluence en fonction des trois groupes. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



Nous avons également obtenu une interaction significative entre gain et type de texte, qui montre que les gains ont été différents en fonction du type de texte avec un gain plus important pour les textes scientifiques que pour les littéraires (Figure 31). Nous avons donc analysé les gains pour chaque type de texte séparément. Le gain était significatif pour les textes littéraires ( $F(1,187) = 4,128, p = ,044$ ) comme pour les textes scientifiques ( $F(1,187) = 47,170, p = ,000$ ).

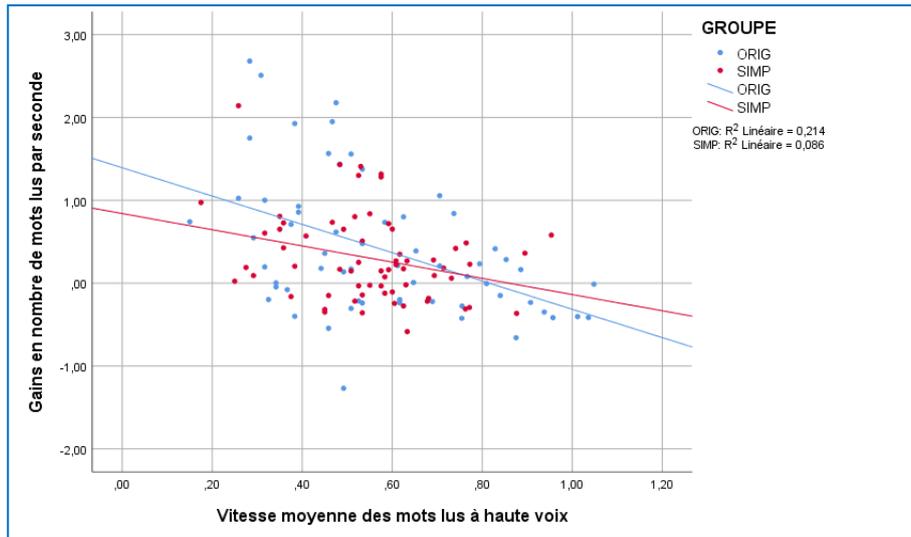
Figure 31 : Effet de l'entraînement entre pré-test et post-test en fluence pour les trois types d'entraînement en fonction du type de textes. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



### 10.4.2 Effets de l'entraînement sur la fluence en fonction du niveau initial dans les tests langagiers et cognitifs.

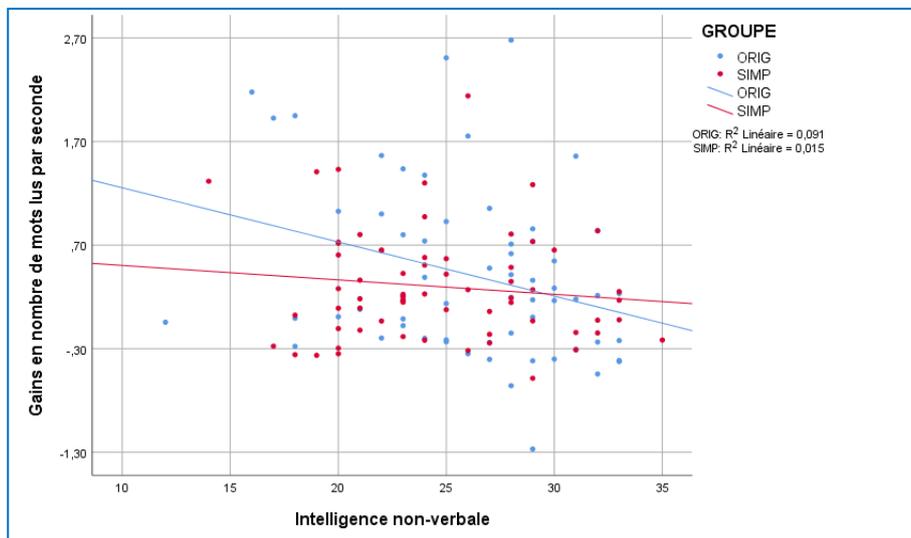
Afin de savoir si les gains du groupe ORIG ou SIMP étaient plus importants pour les plus faibles lecteurs, nous avons réalisé un modèle mixte linéaire afin de prédire les gains en fonction des groupes. Ce premier modèle fait apparaître que les gains ne dépendent pas du groupe ( $F(1,124) = 0,817, p = ,368$ ). Nous avons ensuite inclus dans ce premier modèle chacune des variables langagières (tests de lecture à haute voix de mots familiers, lecture à haute voix de mots inventés, vocabulaire, compréhension orale, mémoire de travail en avant, mémoire de travail en arrière, intelligence non-verbale, conscience morphologique de mots suffixés, conscience morphologique de pseudo-mots, reconnaissance visuelle des mots, fluence et test des cloches), ainsi que l'interaction entre les groupes et chacune des variables langagières testées afin de savoir si une variable langagière expliquait un gain et si cette variable interagissait avec groupe. Nous avons trouvé quatre variables langagières qui modèrent les gains. La première variable est la lecture à haute voix de mots familiers avec une influence de la variable négative ( $\beta = - 0,470, p = ,000$ ), ce qui veut dire que les enfants ayant un faible niveau en lecture à haute voix de mots familiers ont obtenu des gains plus importants que les élèves avec un bon niveau. Toutefois, l'interaction avec groupe n'était pas significative ( $p = ,174$ ) ce qui signifie que le niveau initial de l'élève en lecture des mots familiers impacte les gains quel que soit le groupe. La seconde variable expliquant le gain est la vitesse de lecture des mots inventés avec une modulation négative ( $\beta = - 1,472, p = ,000$ ). Donc les faibles lecteurs gagnent plus que les bons lecteurs. Une fois de plus, l'interaction avec groupe n'était pas significative ( $p = ,211$ ). La troisième variable, la vitesse moyenne des mots lus à haute voix par seconde, expliquant le gain, est modérée de façon négative ( $\beta = - 0,978, p = ,000$ ), ce qui montre que les enfants ayant les plus faibles capacités de décodage ont bénéficié de l'entraînement. Par ailleurs, l'interaction entre le groupe et la vitesse moyenne des mots lus à haute voix par seconde n'est pas significative ( $p = ,238$ ) ce qui suggère que la modulation des gains par cette variable n'est pas différente dans les deux groupes. Les gains des groupes en fonction de la variable vitesse de mots lus à haute voix par seconde sont présentés dans la Figure 32.

Figure 32 : Gains en nombre de mots lus par seconde entre les pré et post-tests en fonction de la vitesse moyenne du nombre des mots à haute voix en effet fixe.



La quatrième variable est l'intelligence non verbale, sa modulation des gains est encore négative ( $\beta = -0,014$ ,  $p = ,011$ ), ce qui veut dire que les enfants ayant un faible niveau en intelligence non verbale montrent des gains plus importants que les enfants ayant un fort potentiel quel que soit le groupe (interaction non-significative,  $p = ,136$ ). Les gains des groupes en fonction de la variable intelligence non verbale sont présentés dans la Figure 33.

Figure 33 : Gains en nombre de mots lus par seconde entre les pré et post-tests en fonction du test intelligence non verbale en effet fixe.



En résumé, les analyses montrent que les gains en nombre de mots lus par seconde entre les pré-tests et les post-tests sont significatifs. Cependant, les gains ne sont pas différents en fonction du groupe. Bien que les faibles lecteurs et les enfants avec un faible niveau cognitif tirent davantage bénéfice de l'entraînement que les bons lecteurs et les enfants ayant un fort

niveau cognitif, ce bénéfice n'est pas différent en fonction du groupe. Ces résultats vont à l'encontre de notre hypothèse selon laquelle les faibles lecteurs auraient dû bénéficier davantage de l'entraînement avec des textes simplifiés.

### 10.4.3 Compréhension : résultats aux Questions à Choix Multiples

Les résultats en compréhension, réponses correctes aux QCM, ont été analysés en utilisant une ANOVA trois-factorielles avec type de texte (littéraire vs scientifique) et gain (pré vs post) comme facteur intra sujet et les trois groupes (TEM-ORIG-SIMP) comme facteur inter-sujets (Tableau 19).

Tableau 19 : Résultats de l'ANOVA trois factorielles (types de texte\*gain\*groupe) pour la variable compréhension.

Effet	df	F	p	$\eta_p^2$	sign
Gain	1, 187	28,980	,000	,135	***
Texte	1, 187	89,046	,000	,325	***
Texte*Groupe	2, 186	4,375	,014	,045	*
Gain*Groupe	2, 186	7,049	,001	,071	**
Gain*Texte	1, 187	220,737	,000	,544	***
Texte*Gain*Groupe	2, 186	6,407	,002	,065	**

Note. \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$

Les résultats de l'ANOVA montrent que la triple interaction entre type de texte, gain et groupe était significative. Cette triple interaction reflète le fait que les effets de l'entraînement varient en fonction de la nature du texte et en fonction des groupes. Cette interaction est décomposée dans la Figure 34 pour les textes littéraires et dans la Figure 35 pour les documentaires scientifiques.

Figure 34 : Moyenne des réponses correctes au QCM des textes littéraires en fonction des pré et post-tests pour les trois types d'entraînement. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.

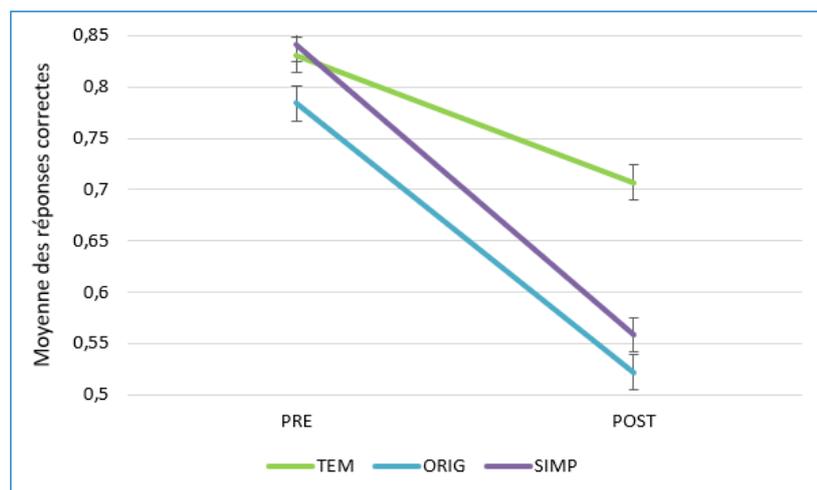
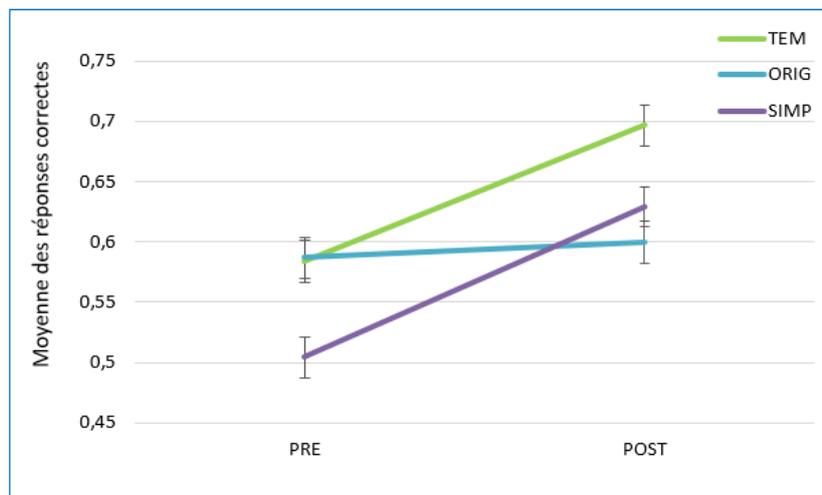


Figure 35 : Moyenne des réponses correctes au QCM des documentaires scientifiques en fonction des pré et post-test pour les trois types d'entraînement. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.



L'interaction entre le gain et le type de texte était significative, ce qui suggère que les gains étaient différents selon le type de texte : les scores sur les textes scientifiques se sont améliorés entre les pré et post-tests alors que les scores des textes littéraires ont baissé pour l'ensemble des groupes. L'interaction significative entre les groupes et les gains indique que les trois groupes n'ont pas eu la même progression entre les pré-tests et les post-tests.

Les analyses de contrastes montrent que la différence entre les deux groupes (ORIG et SIMP) n'était pas significative ( $F(1,125) = 1,958, p = ,164$ ). Le groupe témoin a significativement moins chuté en compréhension entre les pré-tests et les post-tests que les deux groupes qui ont lu les textes sur tablette. Les résultats montrent que l'interaction est plus importante entre le groupe témoin et le groupe qui a lu les textes originaux (TEM vs ORIG ( $F(1,122) = 12,467, p = ,001$ )) que l'interaction entre le groupe témoin et le groupe qui a lu les textes simplifiés (TEM vs SIMP ( $F(1,126) = 6,534, p = ,012$ )).

Enfin, nous avons une interaction significative entre le type de texte et les groupes, ce qui montre que les trois groupes n'ont pas eu les mêmes résultats selon le type de textes. Pour les textes littéraires, les analyses de contrastes montrent que la différence entre les deux groupes (ORIG et SIMP) n'était pas significative ( $F(1,125) = 0,242, p = ,623$ ). Le groupe TEM a significativement moins chuté en compréhension entre les pré-tests et les post-tests que les groupes qui ont lu les textes sur tablette ( $ps < ,001$ ).

Pour les documentaires scientifiques, les analyses de contrastes montrent que la différence entre les deux groupes (TEM et SIMP) n'était pas significative ( $F(1,126) = 0,097, p = ,755$ ). Le groupe ORIG a significativement moins progressé en compréhension entre les pré-tests et les

post-tests que les groupes TEM et SIMP. Les résultats montrent que la significativité est plus importante entre le groupe ORIG et le groupe SIMP (ORIG vs SIMP ( $F(1,125) = 8,129$   $p = ,007$ ) que la significativité entre le groupe TEM et le groupe ORIG qui a lu les textes originaux (TEM vs ORIG ( $F(1,122) = 4,775$   $p = ,031$ ).

#### **10.4.4 Effets de l'entraînement sur la compréhension en fonction du niveau initial dans les tests langagiers et cognitifs.**

Comme dans la section précédente, nous avons analysé si les variables langagières (profils des élèves) modulent les gains différemment dans les deux groupes (ORIG et SIMP). Pour cela, nous avons réalisé un modèle mixte linéaire afin de prédire les gains en fonction des groupes et chacune des variables cognitives ainsi que leurs interactions avec groupe. Les résultats montrent qu'aucune variable langagière ne module les gains. Donc les gains ne s'avèrent pas plus importants pour les élèves ayant un faible score aux tests cognitifs et langagiers.

### **10.5 Discussion**

Dans cette troisième étude, le but était de savoir si les élèves, en particulier les élèves les plus faibles, comprenaient mieux un texte de référence après avoir été entraînés pendant trois mois à la lecture de textes simplifiés versus originaux. Autrement dit, on a cherché à montrer un effet de transfert de la lecture de textes simplifiés à la lecture normale notamment pour des faibles lecteurs. Trois groupes d'élèves de CE2 ont réalisé les mêmes pré-tests et post-tests. Sur cette même période, parmi ces trois groupes, un groupe témoin (TEM) n'a réalisé que les pré-tests et post-tests, un groupe (ORIG) a lu des textes originaux et un dernier groupe (SIMP) a lu ces mêmes textes uniquement en version simplifiée. Notre hypothèse était que les faibles lecteurs devraient mieux entrer dans le cercle vertueux de l'auto-apprentissage (Share, 1995) grâce à l'entraînement de lecture de textes simplifiés.

Concernant la fluence, nous avons constaté que les élèves qui ont suivi les entraînements de lecture de textes simplifiés et originaux, sur tablette, ont vu leur fluence de lecture progresser significativement par rapport au groupe témoin. Les élèves, qui avaient les plus faibles résultats aux tests de lecture à haute voix et intelligence non verbale, sont ceux qui ont le plus progressé grâce à l'entraînement. Ces résultats correspondent en partie à l'hypothèse initiale qui était que l'entraînement à des lectures de textes simplifiés serait bénéfique aux élèves faibles lecteurs. On remarque que l'entraînement a permis aux faibles lecteurs une progression en fluence de

lecture quel que soit le type entraînement (ORIG vs SIMP). On peut conjecturer que cette activité, par sa fréquence et son rituel avec le dispositif de lecture phrase par phrase a permis un entraînement de la fluence de lecture. Cependant l'entraînement n'a pas eu d'effet spécifique pour les élèves ayant des difficultés en vocabulaire. Or nous avons vu dans la sous-section 7.3.2 que la simplification était plus spécifiquement bénéfique aux élèves ayant des difficultés en vocabulaire. Nous avons noté dans cette sous-section 7.3.2 que les élèves ayant un vocabulaire faible ont bénéficié de la simplification mais sur trois années. Sans doute que l'entraînement proposé sur trois mois n'a pas été assez long pour enregistrer un bénéfice pour les élèves ayant un déficit en vocabulaire. L'entraînement aura permis une progression des faibles lecteurs dans l'apprentissage de la lecture avec moins de dépendance au décodage alphabétique.

Concernant l'évolution de la compréhension, le groupe témoin, TEM, est celui qui a obtenu les meilleurs résultats en post-tests. Nous pouvons supposer que dans le groupe témoin, qui n'a pas suivi l'activité de lecture des textes proposés dans l'expérimentation mais néanmoins un enseignement en classe avec des activités de lecture choisies par leur professeur, les élèves ont pu développer leurs connaissances et leurs compétences en lecture sur des supports variés. Cela corrobore l'étude de Connor, Morrison et Petrella (2004), qui ont démontré, en suivant une cohorte d'élèves de CE2 sur une année scolaire, que l'enseignement des stratégies de compréhension permet une progression chez les plus faibles lecteurs et les lecteurs moyens alors qu'un enseignement implicite, entraînement avec un QCM après une lecture silencieuse, a permis un maintien des performances des bons lecteurs et une baisse en compréhension chez les très faibles et moyens lecteurs. L'enseignement explicite constitue un élément fondamental dans l'objectif d'une progression en compréhension. Pour les deux autres groupes ORIG et SIMP, la différence entre pré-tests et post-tests n'est pas significative, et aucun bénéfice spécifique pour les faibles lecteurs n'a été relevé.

Cependant, il faut rester prudent dans l'analyse de ces résultats car seulement deux textes en pré-tests et post-tests ont été lus. En effet, pour ne pas imposer aux classes participant à l'expérimentation la lecture de plus de trente textes, le nombre des textes proposés au départ et à la fin des trois mois a été limité. De plus, si les aspects didactiques et pédagogiques du contexte de classe n'ont pas été questionnés, et donc, n'apparaissent pas dans les résultats relevés lors des expérimentations, ils ne peuvent néanmoins pas être considérés comme inexistant. Dans le contexte de l'expérimentation, les enseignants ont été libres de programmer les activités pédagogiques qu'ils souhaitaient. La présentation de l'expérimentation ne leur a nullement imposé de modifier leur contexte d'enseignement mais on ne peut pas ignorer que celui-ci interagit entre quotidien et activité nouvelle lors d'expérimentations proposées dans le

milieu scolaire (Gentaz et al., 2013). D'autres investigations complémentaires pourraient être envisagées, notamment en augmentant le nombre de textes servant de pré et post tests et l'effectif des participants et en allongeant la durée d'exposition des élèves à l'entraînement de lectures spécifiques sur tablette. Comme d'autres l'ont souligné avant nous :

« Nous espérons que la présentation des apports, originalités et limites de la présente expérimentation sera utile aux futurs chercheurs intéressés par ce type d'approche, affreusement compliqué mais absolument indispensable pour faire progresser notre système éducatif. » (Gentaz et al., 2013, p 9)

## Chapitre 11

### Discussion générale

L'objectif de cette thèse était : Est-ce que la simplification de textes littéraires et de documentaires scientifiques peut améliorer la fluidité et la compréhension en lecture chez des élèves français entre 7 et 9 ans ? Rappelons que le niveau de compréhension en lecture des élèves français en CM1 a chuté de 14 points sur une période de 15 ans (PIRLS, 2016). La simplification de textes peut alors apparaître comme un moyen d'aider les élèves en difficultés de lecture et de compréhension.

Rappelons aussi que la compréhension en lecture dépend de l'identification des mots et de la compréhension orale mais aussi de traitements linguistiques et contextualisés de l'écrit (Bianco, 2015). Ces éléments sont présentés dans la Figure 1, dans l'introduction de la thèse. La maîtrise du décodage et de la reconnaissance des mots sont les conditions premières de l'accès à la compréhension d'un texte (Sprenger-Charolles & Ziegler, 2019). Avec l'automatisation du décodage, la dépendance au décodage alphabétique diminue progressivement (Wagner & Torgesen, 1987) donnant plus de poids à la compréhension orale dont fait partie le vocabulaire (Ouellette & Beers, 2010). La maîtrise de la morphosyntaxe (Bentolila, 2019) tout comme les inférences jouent alors un rôle indispensable dans la compréhension (Van den Broek, 1997). L'ensemble de ces éléments fait partie intégrante des mécanismes de la compréhension décrits notamment dans le modèle « Construction de structures » (Gernsbacher, 1990), dans lequel la suppression ou le renforcement des informations issues de l'écrit permettent d'accéder à la représentation mentale du texte. Le modèle de Construction-Intégration (Kintsch, 2005) insiste sur le contrôle de la représentation mentale exercé par le lecteur grâce à ces connaissances linguistiques et du monde.

Nous avons étudié les effets de la simplification auprès d'élèves français de sept à neuf ans car c'est la phase de l'automatisation de la lecture, de l'auto-apprentissage et des stratégies de compréhension (Bianco & Lima, 2017 ; Willingham, 2006). La simplification peut répondre aux besoins de lecteurs en difficultés grâce à la sélection et la conception de matériel pédagogique tenant compte notamment des aspects lexicaux, syntaxiques et de la cohérence dans les textes. Deux types de textes, textes littéraires et documentaires scientifiques, ont été simplifiés, sur lesquels les effets de la simplification ont été étudiés. Notons que les élèves français avaient une baisse plus importante sur les textes informatifs que sur les textes narratifs

dans les évaluations de PIRLS (2016). La nature du type de texte influence la compréhension car les textes narratifs sont plus facilement compris que les textes informatifs et notamment parce que les sujets traités dépendent de connaissances du monde précises (Best, Floyd & McNamara, 2008 ; McNamara, Ozuru & Floyd, 2017). Ces textes informatifs sont aussi moins fréquemment rencontrés à l'école (Duke, 2000).

Les résultats de nos études montrent que la simplification de textes a été bénéfique sur les trois années, entre le CE1 et le CM1, pour la majorité des élèves et dans les deux types de textes. Il est important de noter que la simplification de textes n'agit pas directement sur la compréhension de textes mais sur les facteurs qui influencent la compréhension de textes, comme la reconnaissance de mots, le vocabulaire, les inférences ou la morphosyntaxe. Elle le fait via les variables textuelles, telles que la fréquence des mots, la longueur des phrases ou la réduction du niveau de complexité d'un texte donné.

Concernant la fluidité de lecture, l'ensemble des élèves a tiré bénéfice de la simplification des textes et pas uniquement les élèves ayant des difficultés de lecture sur les trois années. Toutefois, en CE1, nous avons observé des corrélations entre le niveau de lecture et les gains de simplification suggérant que les faibles lecteurs tiraient davantage parti de la simplification (Javourey-Drevet, François, Gala, Dufau, Ginesté & Ziegler, en révision). Il est possible que les gains de la simplification se réduisent au fur et à mesure que le décodage s'automatise. En revanche, les élèves ayant un faible niveau initial en vocabulaire étaient ceux qui ont le plus bénéficié de la simplification au cours des trois années d'expérimentation. Selon Perfetti et Stafura (2014), la qualité lexicale se construit et se développe lors de l'exposition à l'écrit dans de multiples contextes. Notons également que la simplification sur les textes littéraires et sur les documentaires scientifiques n'a pas eu les mêmes effets chaque année.

Au niveau des variables textuelles, le gain en fluidité de lecture pour le CM1 peut être interprété par le remplacement des pronoms par leurs référents dans le texte. Ce gain peut aussi être justifié par le remplacement des adjectifs des textes originaux par des adjectifs plus fréquents dans les textes simplifiés, de façon plus importante sur deux années, en CE1 et en CM1. Le gain de la simplification en fluidité de lecture sur la seconde année d'expérimentation a été expliqué par l'utilisation de mots plus courts et plus fréquents lors de la simplification. Ces résultats sont conformes à la littérature sur les paramètres de fréquence et longueur des mots (Rello et al, 2013b). De plus, lors de cette seconde année, le nombre de mots par phrase a été réduit, ce qui explique aussi le gain constaté en fluidité (Gala et al., 2020a). Sur la variable fluidité, tous les niveaux de simplification (lexical, syntaxique, lexico-syntaxique) ont permis

un gain, de façon équivalente, pour des élèves de CE1. En effet, les aspects lexicaux, syntaxiques et la cohérence dans les textes participent à la lisibilité (Mesnager, 2002). De ce fait, ils ont un poids sur la fluidité de lecture du texte. Par conséquent, la simplification sur un des trois niveaux représente déjà une aide pour les apprentis-lecteurs. Cependant, ces gains ont été plus importants pour les textes littéraires que pour les documentaires scientifiques. Les textes scientifiques, même simplifiés, restent toujours plus difficiles à appréhender (Hiebert & Cervetti, 2012).

Concernant la compréhension en lecture, la simplification a été profitable pour l'ensemble des élèves et plus spécifiquement sur les textes littéraires que sur les documentaires scientifiques. Les élèves ayant le plus bénéficié de la simplification au cours des trois années d'expérimentation sont ceux qui avaient des difficultés initiales en phonologie, en lecture à haute voix de mots familiers, en vocabulaire et en orthographe morphologique. Nos résultats sont conformes à la littérature. Par exemple, Gentaz et ses collaborateurs (2015) précisent que les quatre principaux prédicteurs de la compréhension en lecture sont les compétences de décodage, la compréhension orale, le vocabulaire et la conscience phonémique. De plus, les faibles connaissances en vocabulaire pourraient être la cause de difficultés de traitement au niveau de l'écrit. Ainsi des enfants qui ont du mal à générer des inférences sont aussi moins en mesure d'utiliser le contexte pour découvrir la signification de nouveaux mots (Cain, Oakhill & Lemmon, 2004), ce qui entraîne des différences de connaissance de vocabulaire entre les enfants, tout au long de leur développement. Il est à noter que les élèves ayant bénéficié de la simplification sont ceux ayant de faibles capacités sur les facteurs vocabulaire et orthographe morphologique, qui sont pris en compte dans le modèle de Construction-Intégration de Kintsch (2005). En d'autres termes, dans les processus permettant d'aboutir à la construction du sens, un échange constant est réalisé entre les connaissances et les compétences du lecteur et les informations du texte afin qu'une représentation mentale cohérente soit validée. Ces processus de Construction-Intégration sont activés en s'appuyant sur les compétences linguistiques du lecteur dont le vocabulaire et l'orthographe morphologique font partie.

Le gain en compréhension de lecture grâce à la simplification de textes peut être expliqué par l'utilisation d'adjectifs et de mots plus fréquents dans les textes simplifiés et le remplacement des pronoms par leurs référents (Brouwers et al., 2014). La simplification permet la rencontre de mots qui sont pour le lecteur de haute qualité lexicale (Perfetti, 2007) alors qu'en conservant le mot du texte original le lecteur n'aurait peut-être pas eu accès au sens du mot. Le bénéfice de l'utilisation de la simplification lexicale ou syntaxique ou lexico-syntaxique sur la variable compréhension ne se retrouve que pour les textes littéraires. Cela peut s'expliquer par

le fait que les stratégies mises en œuvre dans les textes narratifs ou plus descriptifs sont différentes et plus appropriées car l'organisation des éléments du texte et les associations sémantiques influencent la mémorisation. Les textes documentaires informatifs sont plus difficiles à comprendre en général que dans les textes narratifs et leurs éléments sont plus complexes à mémoriser (Wolfe, 2005).

Suite à ces premiers résultats encourageants, nous avons voulu déterminer si un effet de généralisation était possible grâce à un entraînement de trois mois. Notre hypothèse était que les lecteurs en difficultés pourraient tirer parti de la lecture de textes simplifiés et ainsi entrer dans le cercle vertueux de l'auto-apprentissage (Share, 1995).

Concernant la fluidité de lecture, la lecture sur trois mois de textes originaux et simplifiés a été bénéfique pour l'ensemble des élèves par rapport aux élèves du groupe témoin qui n'ont pas suivi d'entraînement. De plus, les élèves ayant des difficultés en lecture à haute voix de mots et en intelligence non verbale, ont plus particulièrement bénéficié de l'entraînement. On peut conjecturer que l'entraînement a donné l'occasion à ces élèves d'automatiser les procédures de reconnaissance de mots qui sont à la base d'un gain en fluidité de lecture. Au départ un enfant peut lire lentement et avec beaucoup d'efforts des mots nouveaux, puis, en reconnaissant les mots décodés et en les comprenant il va par la suite les lire plus rapidement et efficacement, avec moins de recours au décodage, c'est le processus d'apprentissage orthographique (Ziegler et al., 2014). Ce processus se retrouve encore chez les lecteurs adultes experts (Castles et al., 2018), qui l'appliquent tout au long de leur vie face aux nouveaux mots écrits qu'ils découvrent, comme par exemple les mots récents Windows, WhatsApp, Snapchat. On peut supposer que l'entraînement aura permis aux élèves d'entrer dans le cercle vertueux de l'auto-apprentissage (Share, 1995) mais pas uniquement grâce à l'effet de la simplification.

Concernant la compréhension en lecture, aucun bénéfice spécifique lié à l'entraînement à la lecture des textes originaux ou simplifiés proposés sur tablette, n'a été relevé. Cela peut s'expliquer par le fait que le décodage et la compréhension ne sont pas directement liés comme décrit dans le triangle DVC (Perfetti et al., 2010). Pour aider les élèves à progresser en compréhension il est important de travailler le vocabulaire, par exemple, et les différents aspects de la langue. Ce constat montre les limites du dispositif proposé et l'importance de l'enseignement des stratégies de compréhension (Lima et al., 2016).

## Limites

Une des limites de la simplification connues des chercheurs est que la nuance des mots peut être perdue lors de la simplification. Les écrivains choisissent les mots qu'ils emploient afin de correspondre au mieux à leurs idées et leur intention de communication au travers d'une figure de style, comme par exemple, une métaphore, une répétition, une énumération .... En respectant la définition de Siddharthan (2014), l'objectif de la simplification est «la conservation de l'information et du sens original » plutôt que la conservation du style original. Le remplacement de mots complexes par des synonymes plus simples peut réduire les nuances sémantiques. Ainsi le verbe *concentrer* (réunir des éléments en un point, centraliser, augmenter la masse d'une substance dissoute dans une unité de volume d'un liquide solvant en chimie), peut être considéré comme synonyme du verbe *condenser* (rendre dense, liquéfier ou solidifier un corps à l'état gazeux ou épaissir en réduisant à un plus petit volume) mais l'auteur, utilisant l'un ou l'autre, apportera une nuance dans son message.

Notre réflexion nous a conduit aussi à interroger les conditions de l'expérimentation. Placer les élèves devant un résumé à compléter, alors que tous n'avaient peut-être pas encore connaissance de sa définition, de son intérêt et de son usage, ne l'avaient pas nécessairement étudié avec leurs professeurs, pose la question de l'expérimentation à grande échelle sans regard sur les pratiques de classe des professeurs. Ainsi, les résultats des expérimentations avec les résumés n'ont pas pu conduire à des résultats généralisables. Les expérimentations en classe posent un certain nombre de limites. L'impact des interventions des professeurs auprès de leurs élèves ne peut être occulté (Gentaz et al., 2013). De plus, l'étude de généralisation liée à un entraînement de lecture, n'a pu être menée que sur trois mois au regard du temps scolaire. Un entraînement plus long aurait peut-être eu d'autres effets.

## Valorisations

Le dispositif numérique utilisé a rendu possible la transmission directe des données au laboratoire. La première année, 165 élèves de CE1 ont débuté sur l'étude longitudinale. La seconde année, 330 élèves ont participé aux expérimentations : les CE1 devenus CE2 qui ont poursuivi l'étude longitudinale et les CE1 qui ont participé à l'étude sur les différents niveaux de la simplification. La dernière année, 583 élèves ont participé aux deux études citées précédemment ainsi qu'à l'étude sur l'entraînement.

Sur le plan pédagogique, l'utilisation des outils créés (voir chapitre 12), en permettant d'adapter les textes aux besoins de l'élève, pourrait soutenir l'entrée d'un plus grand nombre d'élèves dans le cercle vertueux de l'auto-apprentissage.

### **Perspectives**

L'objectif à moyen terme serait de poursuivre l'évaluation de l'effet de la simplification suivant le type de texte puisque des études ont été menées sur des documentaires historiques en adaptant le texte par une valorisation des relations de causalité (Beck et al., 1991). Nous pouvons aussi envisager une étude des effets de la simplification dans des écoles classées en REP, puisque le corpus et les questionnaires sont disponibles. À notre connaissance aucune étude n'a testé les effets de la simplification auprès d'enfants issus de milieux socialement et économiquement faibles ayant des difficultés en maîtrise de la langue. En effet, les travaux de cette thèse montrent que les lecteurs fragiles ou ayant un faible niveau en vocabulaire ont bénéficié de la simplification. Une autre possibilité est de tester les effets de la simplification pour des élèves en collège car les processus cognitifs engagés lors de la compréhension diffèrent de ceux de jeunes enfants (Van den Broek, 1997).

Les objectifs à long terme seraient de générer des textes simplifiés en français, de façon semi-automatique, grâce à la simplification automatique de textes afin de diversifier les possibilités de lectures et les types de texte proposés. De plus, grâce à l'apprentissage automatique et notre corpus (« big data »), l'objectif serait de proposer directement à l'élève le texte adapté en fonction, par exemple, de sa vitesse de lecture ou de sa compréhension d'un texte lu précédemment.

## Chapitre 12

### Ressources disponibles pour donner suite aux études

Dans ce chapitre, nous présentons deux ressources numériques conçues à partir de ce travail de thèse. Elles sont l'aboutissement d'un travail d'équipe regroupant plusieurs laboratoires et sont accessibles librement sur deux sites : le premier propose le corpus de textes ALECTOR, le deuxième héberge une application créée avec l'ensemble du matériel utilisé par les élèves lors des expérimentations présentées dans la thèse.

#### 12.1 Le corpus ALECTOR

Le corpus ALECTOR est issu d'un travail collaboratif au sein du projet ANR ALECTOR (ANR-16-CE28-0005). Sur le site web du projet sont mis à disposition l'ensemble des publications faisant suite aux travaux conduits sur ce projet, ainsi qu'un recueil de recommandations (*guidelines*) pour transformer des textes en français afin d'améliorer leur lisibilité et leur compréhension pour des enfants dyslexiques et faibles lecteurs<sup>62</sup>. Le corpus réalisé dans le cadre du projet est constitué de 79 textes, dont la plupart ont été utilisés dans le cadre des expérimentations de notre travail de thèse. L'interface, réalisée par Solange Lâm dans le cadre d'un stage en informatique, avec la mise en ligne sur la plateforme HumaNum par Tatsuya Watanabe, ingénieur d'études, permet de choisir un texte et sa version adaptée (simplifiée) en fonction de différents critères (difficulté du texte<sup>63</sup>, niveau, genre textuel et nombre de mots). Chaque texte est accompagné d'un questionnaire de compréhension (QCM) afin que le professionnel puisse le proposer à l'apprenant après chaque lecture. L'interface propose également l'ensemble des corrections des tests de compréhension (Figure 36).

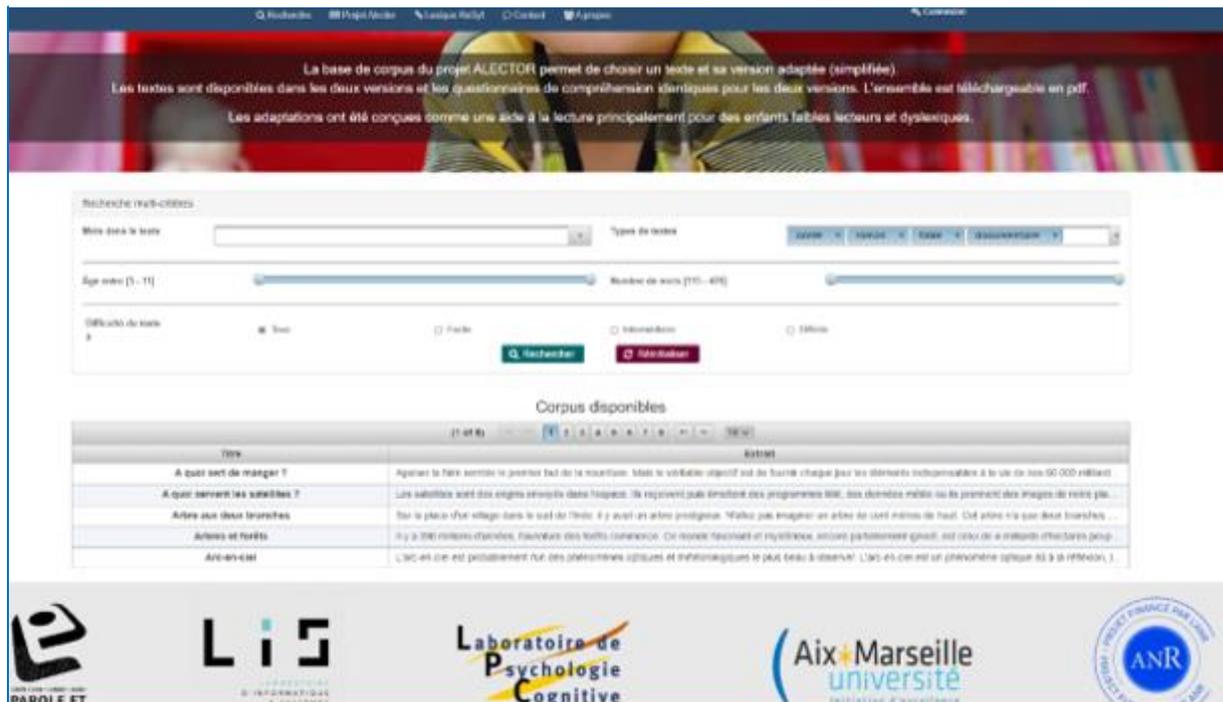
Nous avons choisi un système d'inscription sur le site afin de pouvoir avoir un minimum de renseignements sur les professionnels (orthophoniste, enseignant du premier ou du second degré, enseignant-chercheur ou autre) qui utilisent les textes et les questionnaires.

---

<sup>62</sup> <https://alectorsite.wordpress.com/>

<sup>63</sup>Faisant suite aux expérimentations vues précédemment au sein de cette thèse, nous avons calculé au sein de chaque classe d'âge un score Z en vitesse de lecture ainsi qu'un score Z pour la compréhension pour l'ensemble des textes. La somme de ces deux scores Z nous a permis d'évaluer de façon empirique la difficulté des textes et de proposer un ordre de difficultés au sein du corpus.

Figure 36 : Présentation de l'interface permettant une recherche multi-critères des textes disponibles.



Le résultat est une liste de textes correspondant aux critères sélectionnés. En accédant au texte choisi, il est possible de choisir d'augmenter la taille de la police, l'interligne et l'espacement entre les caractères pour aider les lecteurs en difficultés (Zorzi et al., 2012). La Figure 37 présente un texte en exemple.

Figure 37 : Présentation d'un texte parallèle avec la police Open Dyslexic.



Les textes et les questionnaires de compréhension sont téléchargeables en format pdf. Pour les textes la police OpenDyslexia a été choisie pour faciliter la lecture.

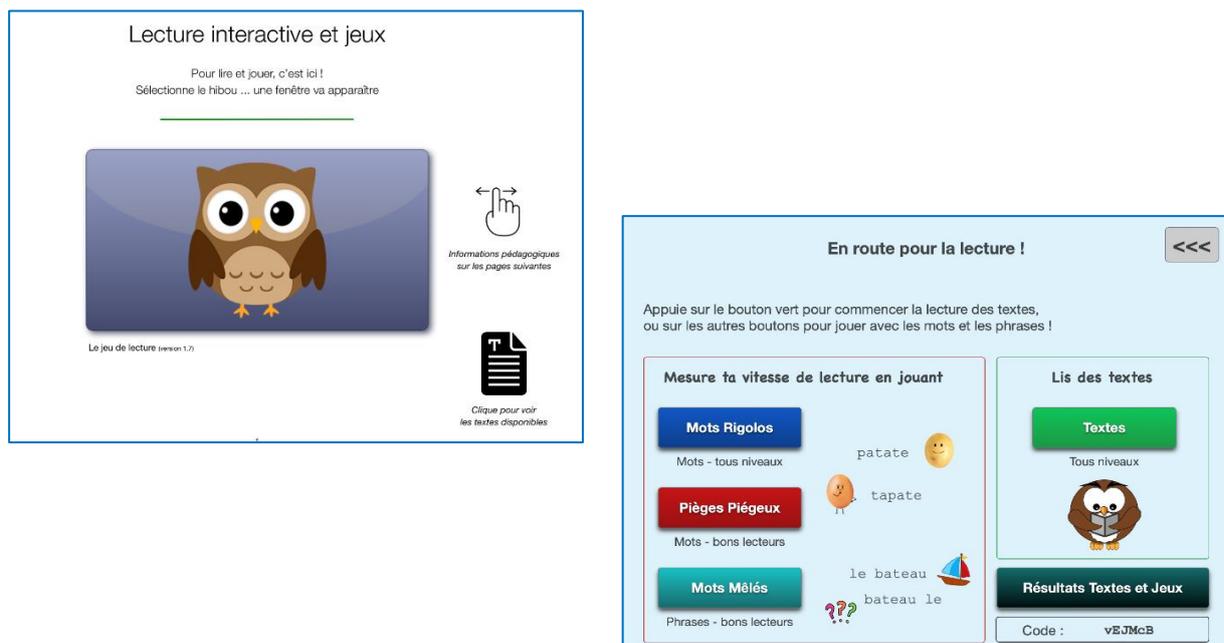
## 12.2 Livre interactif Hibou

Dans la même dynamique, la seconde ressource propose gratuitement au grand public l'ensemble des supports réalisés au cours de ces trois années d'expérimentation. Nous avons continué le développement de l'outil de recherche afin de le proposer sous forme d'application ludique pour les élèves. La totalité de la programmation a été réalisée par Stéphane Dufau, ingénieur de recherche au sein du Laboratoire de Psychologie Cognitive.

L'application intitulée « Hibou » est un livre interactif<sup>64</sup> qui propose également des jeux mais uniquement sous format iOS actuellement, car le terrain de recherche était équipé en iPad ; un développement pour une version ouverte à d'autres systèmes d'exploitation est envisagé. Plusieurs livres sont téléchargeables sur le même iPad ou grand iPhone ou ordinateur dans différentes couleurs afin que, dans une famille de plusieurs enfants, chacun puisse avoir son propre livre avec ses étoiles et ses trophées.

Dans la partie des lectures, sont proposés 69 textes en version originale et version simplifiée sur trois niveaux : lexical, morphosyntaxique et discursif, ainsi que l'ensemble des questions et des résumés. L'application informe l'utilisateur de son contenu et présente dès le premier chapitre (Figure 38), les possibilités de travail de la langue : lectures ou jeux de décision lexicale.

Figure 38 : Vues de l'application, phase des choix possibles dans l'outil.



<sup>64</sup> <https://lpc.univ-amu.fr/fr/hibou-livre-interactif>

L'enfant peut choisir l'espacement de caractères qui lui convient le mieux, une fois ce paramètre fixé, il peut néanmoins être changé à tout moment. L'élève ou l'enseignant peut ensuite choisir le niveau des textes, de « Débutant » correspondant à la classe de CE1 jusqu'à « lecteur confirmé » qui correspond au niveau CM. L'utilisateur choisit la catégorie, « défi » pour les textes en version simplifiée ou « challenge » pour les textes en version originale.

Sur la page affichant les textes, une icône correspondant au thème du texte s'affiche.

Encadrée en vert, elle correspond aux textes littéraires et encadrée en rouge elle correspond aux documentaires scientifiques. Avec les données empiriques, il a été possible de programmer l'apparition d'un message « tu lis trop rapidement » lorsque le temps de lecture était inférieur au seuil minimum en fonction du nombre de mots dans la phrase (Figure 39).

Nous avons conservé les mêmes règles que lors de l'expérimentation. Pour les réponses au questionnaire : les étoiles dorées valorisent les bonnes réponses ; les étoiles argentées valorisent les réponses erronées mais dont un ou plusieurs éléments sont présents dans le texte et si la réponse est incorrecte le lecteur n'obtient pas d'étoile. Pour le résumé (Figure 39), chaque mot ou groupe de mots bien positionné dans le résumé à trous permet d'obtenir une étoile dorée. Par contre, à la différence de l'expérimentation, si l'élève le souhaite, il peut lire plusieurs fois le texte avant de passer au questionnaire, ce qui n'était pas possible lors de la phase expérimentale. Il peut aussi reprendre son évaluation si les performances obtenues ne lui conviennent pas, mais seulement après relecture du texte.

Figure 39 : Vues de l'application, lecture et résumé.

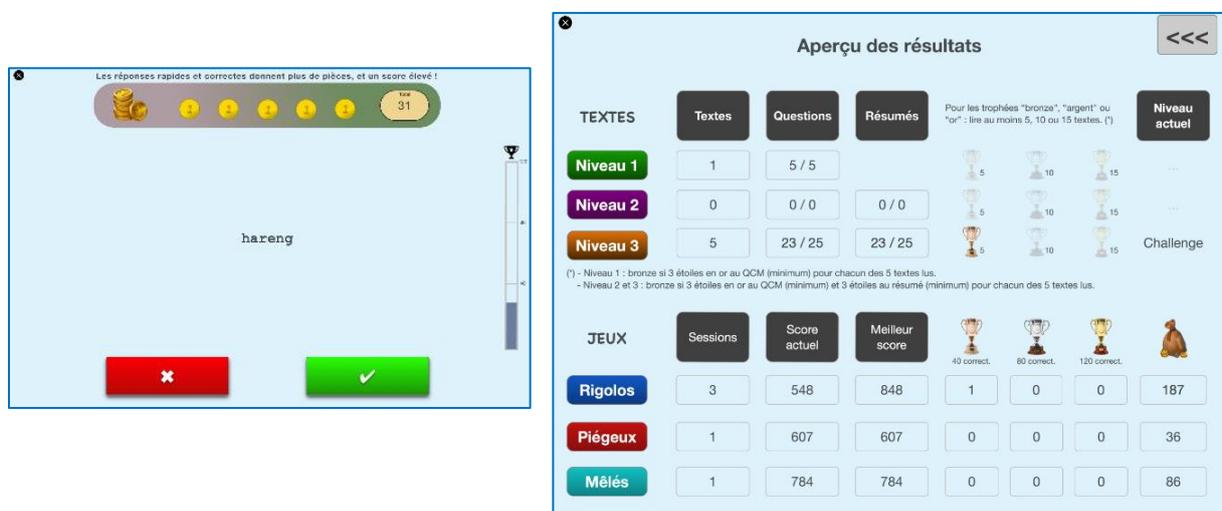
The image shows two screenshots from an application. The left screenshot displays a warning message: "Tu lis trop rapidement !" with a cartoon owl icon and the text "Les Romains et les Gaulois avaient aussi leur savon." Below the text are navigation buttons for back and "Suivant" (Next), and a footer indicating "Phrase : 3 / 25".

The right screenshot shows a "Résumé" (Summary) game interface. At the top, it asks "Peux-tu placer les mots au bon endroit?" (Can you place the words in the right place?). Below this is a table with five columns: "Tout à coup", "L'après-midi", and "devant sa porte". The first row contains the words: "1. dans les bois", "2. toute la nuit", "3.", "4.", and "5.". Below the table is a paragraph of text with several words missing, indicated by underscores and numbers: "C'était une très vieille sorcière qui habitait une maison **dans les bois** . Un jour, un jeune homme passa devant sa fenêtre. Il était très beau. La vieille sorcière en devint plus amoureuse que jamais. Elle chercha des recettes **toute la nuit** dans de vieux livres et prépara un élixir pour rajeunir. 3 , elle avala l'élixir. Devenue jeune et jolie, elle se créa aussi une robe superbe grâce à une formule magique. Enfin, elle accrocha à sa robe une rose blanche trempée dans un philtre d'amour. Elle attendit sur un banc, 4 . Peu après, le beau jeune homme arriva, vêtu d'un riche costume brodé d'or, une fleur blanche à sa veste. Il salua la sorcière et ils parlèrent. Très vite le jeune homme devint amoureux aussi et ils se donnèrent leur premier baiser. 5 , la sorcière le quitta très vite et s'enferma chez elle. Quelques secondes après, elle redevenait une vieille sorcière. Les jours suivants ils continuèrent à se rencontrer, se dire des mots d'amour, échanger des baisers et partir. Le jeune homme la quittait très vite, en souriant, sans se retourner." At the bottom, there is a green button with ">>>>" and a text box that says "Fais glisser les mots sur les numéros du tableau et ils s'ajoutent au texte. Tu peux remettre les mots en haut du tableau si tu t'es trompé(e)." (Drag the words onto the numbers in the table and they will be added to the text. You can put the words back at the top of the table if you were wrong.)

Une page permet de visualiser rapidement le nombre des textes lus, le score obtenu aux questions et aux résumés et de gagner des coupes : trophée « bronze » pour 5 textes lus, trophée « argent » pour 10 textes lus et trophée « or » pour 15 textes lus (Figure 40).

Trois jeux sérieux sont aussi proposés, deux sur la reconnaissance orthographique des mots et un sur la sémantique avec des syntagmes. Les jeux reposent sur une tâche de décision lexicale. Un feed-back sonore permet à l'élève de savoir si sa réponse est juste ou fausse afin de ne pas mémoriser une mauvaise orthographe. Le premier jeu de mots, "Mots Rigolos", est destiné aux lecteurs de tous niveaux. L'ensemble des stimuli a été créé par Stéphane Dufau, en se basant sur les mots de Manulex. Le deuxième jeu de mots, "Pièges Piégeux", est destiné aux bons lecteurs, niveau CM. Les pièges sont construits principalement en modifiant l'orthographe des mots, par exemple "écuoter". Les stimuli ont été proposés par Stanislas Dehaene. Le troisième jeu, "Mots Mêlés", est un jeu basé sur des syntagmes. Ici, les phrases pièges sont des phrases où les mots ont été déplacés. Ces phrases sont correctes ou n'ont pas de sens et ne sont pas correctes d'un point de vue grammatical ou syntaxique. Les stimuli ont été conçus par Stéphane Dufau. Les trois jeux mobilisent les connaissances en langue mais aussi l'attention du joueur.

Figure 40 : Pages d'un jeu de décision lexicale et page générale des scores.



L'ensemble de ces écrans permet d'avoir une vision globale des résultats mais aussi de stimuler les enfants par le biais des étoiles et des trophées.

Au travers de l'application Hibou, l'enfant peut prendre conscience, en autonomie, de ce qu'il sait réaliser, de ce qu'il sait déjà, de ce qu'il découvre et peut-être même peut se projeter dans ses progrès grâce au système d'évaluation par étoiles. Pour autant, autour des activités engagées, la communication, entre pairs et avec l'adulte, est indispensable pour institutionnaliser, renforcer, valoriser les connaissances auxquelles l'application le confronte.

## Chapitre 12

### Ressources disponibles pour donner suite aux études

Le but de ces ressources numériques gratuites est de permettre à l'ensemble de la communauté éducative d'accéder à un panel plus important de supports d'entraînement à la lecture, supports à adapter aux besoins des lecteurs, grâce à l'espacement des caractères et à la proposition des textes en version simplifiée.

## Conclusion

Ces travaux de recherche ont permis d'étudier et d'évaluer si la simplification de textes pouvait améliorer la fluidité et la compréhension de la lecture chez des enfants entre 7 et 9 ans. Nous avons vu que la simplification permettait de gagner en fluidité de lecture quel que soit l'âge du lecteur. La simplification de textes améliore aussi la compréhension en lecture des jeunes élèves, davantage encore pour les plus fragiles d'entre eux, ayant de faibles capacités en vocabulaire et lecture de mots familiers notamment. Nous avons montré l'importance des facteurs lexicaux qui ont permis un gain en fluidité de lecture et en compréhension comme la diminution du taux de mots peu fréquents et complexes, par exemple, ou la réduction de la longueur des mots et des phrases et l'importance de variables syntaxiques comme le remplacement des pronoms par leurs référents. Le corpus parallèle créé pour les études de cette thèse a permis de définir des règles de simplification qui aideront à généraliser cette démarche pour d'autres textes.

Ces travaux ont contribué à la création de deux outils numériques dont l'usage pourrait renforcer les compétences d'apprentis-lecteurs grâce à un entraînement en lecture et à l'exposition des élèves aux textes littéraires et documentaires scientifiques. Il est à souligner que la présence d'acteurs de terrain engagés, ainsi que les outils présents sur le terrain de recherche, notamment l'équipement numérique, ont favorisé la mise en place de ces travaux.

Enfin, ces travaux, qui ont été développés grâce aux compétences de chercheurs qui ont mutualisé leurs connaissances et expertises, ouvrent des perspectives d'études intéressantes dans le champ de l'apprentissage et du développement des compétences de la lecture dans le domaine de l'éducation.

**Références bibliographiques**

- Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, 19(9), 933-958.  
<https://doi.org/10.1007/s11145-006-9024-z>
- Ahr, S., & Butlen, M. (2012). Savoir lire/aimer lire : un couple en évolution ou en voie de séparation ? *Ela. Etudes de linguistique appliquée*, (2), 215-235.  
<https://doi-org.lama.univ-amu.fr/10.3917/ela.166.0215>
- Aluísio, S. M., & Gasperin, C. (2010, June). Fostering digital inclusion and accessibility: the Por Simples project for simplification of Portuguese texts. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas* (pp. 46-53). Association for Computational Linguistics.  
<https://dl.acm.org/doi/pdf/10.5555/1868701.1868708>
- Aluísio, S. M., Specia, L., Pardo, T. A., Maziero, E. G., Caseli, H. M., & Fortes, R. P. (2008, September). A corpus analysis of simple account texts and the proposal of simplification strategies: first steps towards text simplification systems. In *Proceedings of the 26th annual ACM international conference on Design of communication* (pp. 15-22).  
<https://dl.acm.org/doi/abs/10.1145/1456536.1456540>
- Anctil, D. (2015). Un meilleur enseignement lexical pour une plus grande appropriation de la langue. Dans : Service de la langue française et de la politique linguistique éd., *S'approprier le français : Pour une langue conviviale* (pp. 101-117). Louvain-la-Neuve, Belgique : De Boeck Supérieur.  
<https://doi.org/10.3917/dbu.slfcl.2015.01.0101>
- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading research quarterly*, 285-303.  
<https://www.jstor.org/stable/748043>
- Andreu, S., Cioldi, I., Conceicao, P., Eteve, Y., Fabre, M., Le Breton, S., Persem, E., Portelli, T., Rocher, T., Vourc'h, R. & Wuillamier, P. (2019). « Evaluations 2019 Repères CP, CE1 : premiers résultats » Document de travail, -MENJ Depp  
[https://www.education.gouv.fr/sites/default/files/imported\\_files/document/dt2019e03\\_1198056.pdf](https://www.education.gouv.fr/sites/default/files/imported_files/document/dt2019e03_1198056.pdf)
- Andreu, S, Dalibard. E & Étève Y., juillet 2016. « Maitrise de la langue en fin d'école : l'écart se creuse entre filles et garçons ». *Note d'information n° 20\_Cèdre 2003-2009-2015, MEN-DEPP*  
[http://cache.media.education.gouv.fr/file/2016/90/1/depp-ni-2016-20-cedre-2015-ecole\\_604901.pdf](http://cache.media.education.gouv.fr/file/2016/90/1/depp-ni-2016-20-cedre-2015-ecole_604901.pdf)

- Augé V., Etève, Y., Ninnin, L-M, DEPP B2, septembre 2019. *Cedre 2007-2013-2018 – Sciences en fin d'école : des résultats stables depuis 11 ans et un niveau plus homogène. Note d'information N° 19.32 MENJ-DEPP*  
<https://www.education.gouv.fr/cedre-2007-2013-2018-sciences-en-fin-d-ecole-des-resultats-stables-depuis-11-ans-et-un-niveau-plus-8828>
- Baeza-Yates, R., Rello, L., & Dembowski, J. (2015). Cassa: A context-aware synonym simplification algorithm. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1380-1385).  
<https://www.aclweb.org/anthology/N15-1156.pdf>
- Barbu, E., Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2015). Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, 42(12), 5076-5086.  
<https://doi.org/10.1016/j.eswa.2015.02.044>
- Barlacchi, G., & Tonelli, S. (2013, March). ERNESTA: A sentence simplification tool for children's stories in Italian. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 476-487). Springer, Berlin, Heidelberg.  
[https://doi-org.lama.univ-amu.fr/10.1007/978-3-642-37256-8\\_39](https://doi-org.lama.univ-amu.fr/10.1007/978-3-642-37256-8_39)
- Bart, D., & Daunay, B. (2015). UN PLAISIR À LIVRE OUVERT ? L'évaluation du plaisir de lire dans le PISA. *Recherches. Revue de didactique et de pédagogie du français*, 63, 147-161.  
[https://revue-recherches.fr/wp-content/uploads/2018/01/147-161\\_R63\\_BartDaunay.pdf](https://revue-recherches.fr/wp-content/uploads/2018/01/147-161_R63_BartDaunay.pdf)
- Bernigole, V., Bret, A., Chabanon, L., Roussel, L., & Verlet, I., (Décembre 2019) *PISA 2018 : culture mathématique, culture scientifique et vie de l'élève - Note d'information N° 19.50 - MENJ-Depp*  
<https://www.education.gouv.fr/pisa-2018-culture-mathematique-culture-scientifique-et-vie-de-l-eleve-6209>
- Baker, L., Dreher, M. J., Shipton, A. K., Beall, L. C., Voelker, A. N., Garrett, A. J., ... & Finger-Elam, M. (2017). Children's comprehension of informational text: Reading, engaging, and learning. *International Electronic Journal of Elementary Education*, 4(1), 197-227.
- Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading research quarterly*, 251-276.  
<https://doi.org/10.2307/747763>

- Bentolila, A. (2019). Conscience syntaxique et apprentissage de la lecture. Dans Bentolila, A & Germain, B (dir.), *L'apprentissage de la lecture*, 211-216, Nathan.
- Bentolila, A. & Richaudeau, F., *Le moniteur de lecture CE2*, Nathan-Retz, 1989  
<https://doi.org/10.3406/spira.1990.1936>
- Benjamin, R. G. (2012). Reconstruction de la lisibilité : Développements récents et recommandations dans l'analyse de la difficulté du texte. *Educational Psychology Review*, 24(1), 63-88.  
 doi : 10.1007/s10648-011-9181-8
- Best, R., Floyd, R. G., & Mcnamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading psychology*, 29(2), 137-164.  
<https://doi.org/10.1080/02702710801963951>
- Best, R., Ozuru, Y., Floyd, R., & McNamara, D. (2006). Children's Text Comprehension: Effects of Genre, Knowledge and Text Cohesion.  
<https://doi.org/10.22318/icls2006.37>
- Bianco, M. (2015). *Du langage oral à la compréhension de l'écrit*. PUG.
- Bianco, M., Bressoux, P., Doyen, A. L., Lambert, E., Lima, L., Pellenq, C., & Zorman, M. (2010). Early training in oral comprehension and phonological skills: Results of a three-year longitudinal study. *Scientific Studies of Reading*, 14(3), 211-246.
- Bianco, M., Coda, M., & Gourgue, D. (2002). *Compréhension, Grande Section*. Grenoble : Editions de la Cigale.
- Bianco, M., Coda, M., & Gourgue, D. (2006). *Compréhension, Moyenne Section*. Grenoble : Editions de la Cigale.
- Bianco, M., & Lima, L. (2017). *Comment enseigner la compréhension en lecture ?* Hatier.
- Bianco, M., Nardy, A. Lima, L., Colé, P. & Megherbi, H. (2014). ANR Devcomp,  
<https://pdessus.fr/devcomp/membres.html>
- Bianco, M., Pellenq, C., Lambert, E., Bressoux, P., Lima, L., & Doyen, A. L. (2012). Impact of early code-skill and oral-comprehension training on reading achievement in first grade. *Journal of research in reading*, 35(4), 427-455.  
<https://doi.org.lama.univ-amu.fr/10.1111/j.1467-9817.2010.01479.x>

- Billami, M. B., François, T., & Gala, N. (2018, August). ReSyf: a French lexicon with ranked synonyms. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2570-2581).  
<https://www.aclweb.org/anthology/C18-1218.pdf>
- Blanc N. & Brouillet D. (2003). *Mémoire et compréhension*. Paris : A. Colin.
- Bormuth, J. R. (1966). Readability: A new approach. *Reading research quarterly*, 79-132.  
<https://doi.org/10.2307/747021>
- Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: the visual attention span deficit hypothesis. *Cognition*, 104(2), 198-230.  
 doi : 0.1016/j.cognition.2006.05.009
- Bott, S., Rello, L., Drndarević, B., & Saggion, H. (2012, December). Can spanish be simpler? lexis: Lexical simplification for spanish. In *Proceedings of COLING 2012* (pp. 357-374).  
<https://www.aclweb.org/anthology/C12-1023.pdf>
- Boyer, J. Y. (1992). La lisibilité. *Revue française de pédagogie*, 5-14.  
<http://www.jstor.org/stable/41200261>
- Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, 49(4), 422-432.  
<https://doi.org/10.1111/j.1469-7610.2007.01849.x>
- Brouwers, L., Bernhard, D., Ligozat, A. L., & François, T. (2014, April). Syntactic sentence simplification for French. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)* (pp. 47-56).  
<https://www.aclweb.org/anthology/W14-1206.pdf>
- Brunato, D., Dell'Orletta, F., Venturi, G., & Montemagni, S. (2014). Defining an annotation scheme with a view to automatic text simplification. *Defining an annotation scheme with a view to automatic text simplification*, 87-92.
- Brunel, A. & Combes, M. : sous la direction de Núria Gala et Johannes Ziegler (2015). *Simplification de textes pour faciliter leur lisibilité et leur compréhension*. [Mémoire de master orthophonie, Aix-Marseille-université].  
<http://www.sudoc.abes.fr/cbs/xslt/DB=2.1//SRCH?IKT=12&TRM=186413904>

- Bulté, B., Sevens, L., & Vandeghinste, V. (2018). Automating lexical simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 8, 24-48.  
<https://clinjournal.org/clinj/article/view/78>
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & cognition*, 29(6), 850-859.  
<https://doi-org.lama.univ-amu.fr/10.3758/BF03196414>
- Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of child language*, 30(3), 681.  
[doi:10.1017/S0305000903005713](https://doi.org/10.1017/S0305000903005713)
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of educational psychology*, 96(4), 671.  
<https://doi.org/10.1037/0022-0663.96.4.671>
- Candido Jr, A., Maziero, E. G., Specia, L., Gasperin, C., Pardo, T., & Aluísio, S. (2009, June). Supporting the adaptation of texts for poor literacy readers: a text simplification editor for Brazilian Portuguese. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 34-42).  
<https://www.aclweb.org/anthology/W09-2105.pdf>
- Cardon, R. (2018). Approche lexicale de la simplification automatique de textes médicaux. In *Actes de la conférence Traitement Automatique de la Langue Naturelle, TALN 2018* (p. 159).  
<https://www.aclweb.org/anthology/2018.jeptalnrecital-recital.12>
- Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998, July). Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology* (pp. 7-10).
- Casalis, S., Colé, P., & Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Annals of dyslexia*, 54(1), 114-138.
- Casini, L., Pech-Georgel, C., & Ziegler, J. C. (2017). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, n/a-n/a.  
[doi:10.1111/desc.12530](https://doi.org/10.1111/desc.12530)
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5-51.  
<https://doi.org/10.1177/1529100618772271>

- Cèbe, S., & Goigoux, R. (2009). *Lector et Lectrix: apprendre à comprendre les textes narratifs*.  
<https://halshs.archives-ouvertes.fr/hal-00974777/>
- Cèbe, S., Goigoux, R., & Thomazet, S. (2014). Enseigner la compréhension. Principes didactiques, exemples de tâches et d'activités.  
<https://hal.archives-ouvertes.fr/hal-00922482/>
- Cellier, M. (2011). Le vocabulaire et son enseignement. Des outils pour structurer l'apprentissage du vocabulaire.  
[http://cache.media.eduscol.education.fr/file/Dossier\\_vocabulaire/57/6/Micheline\\_Cellier\\_111202\\_C\\_201576.pdf](http://cache.media.eduscol.education.fr/file/Dossier_vocabulaire/57/6/Micheline_Cellier_111202_C_201576.pdf)
- Chabanon, L., (Juin 2020) *Journée défense et citoyenneté 2019 - Note d'Information n° 20.20*, Ministère de l'Éducation Nationale-Depp  
<https://www.education.gouv.fr/journee-defense-et-citoyennete-2019-plus-d-un-jeune-francais-sur-dix-en-difficulte-de-lecture-303885>
- Chabanon, L., Durand de Monestrol, H., & Verlet, I., Décembre 2019, *PISA 2018 : stabilité des résultats en compréhension de l'écrit - Note d'information N° 19.49*, MENJ-DEPP  
<https://www.education.gouv.fr/pisa-2018-stabilite-des-resultats-en-comprehension-de-l-ecrit-10976>
- Charolles, M. (2006). Cohérence, pertinence et intégration conceptuelle. Lane, P.(éd.) *Des discours aux textes : modèles et analyses*, Rouen, Presses Universitaires de Rouen, 39-74.  
<https://hal.archives-ouvertes.fr/hal-01404528/document>
- Charpak, G. (1996). *La Main à la pâte : Les sciences à l'école primaire*. Flammarion.
- Chevrie-Muller, C., Maillart, C. Simon, A. M., Fournier, S. (2010). Langage oral, langage écrit, mémoire, 2ème édition : L2MA-2. Paris. *Edition du centre de psychologie appliquée* (ECPA).
- Ciravégna, N. (1995). *Chichois de la rue des Mauvestis*. Pocket.
- Colé, P., Marec-Breton, N., Royer, C., & Gombert, J. E. (2003). Morphologie des mots et apprentissage de la lecture. *Reeducation Orthophonie*, 213(1), 57-60.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.1372&rep=rep1&type=pdf#page=59>
- Colé, P., Royer, C., Leuwens, C., & Casalis, S. (2004). Les connaissances morphologiques dérivationnelles et l'apprentissage de la lecture chez l'apprenti-lecteur français du CP au CE2. *L'année psychologique*, 104(4), 701-750.  
<https://doi.org/10.3406/psy.2004.29686>

- Colmant, M., & Le Cam, M. (2017). PIRLS 2016 : évaluation internationale des élèves de CM1 en compréhension de l'écrit. Évolution des performances sur quinze ans. *Note d'information n° 17.24*, MEN-DEPP  
[http://cache.media.education.gouv.fr/file/2017/73/7/depp-ni-2017-24-pirls-cm1-ecrit\\_860737.pdf](http://cache.media.education.gouv.fr/file/2017/73/7/depp-ni-2017-24-pirls-cm1-ecrit_860737.pdf)
- Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child x instruction interactions. *Journal of educational psychology*, 96(4), 682.  
<https://doi.org/10.1037/0022-0663.96.4.682>
- Content, A., & Peereman, R. (1999). La reconnaissance des mots écrits. *Troubles du langage: Bases théoriques, diagnostic et rééducation*, 257-288.
- Corblin, F. (1995). *Les formes de reprise dans le discours. Anaphores et chaînes de référence*. Presses universitaires de Rennes.  
[https://jeannicod.ccsd.cnrs.fr/ijn\\_00550962/](https://jeannicod.ccsd.cnrs.fr/ijn_00550962/)
- Crinon, J., Legros, D., Marin, B., & Avel, P. (2007). Aides logicielles à la lecture de textes documentaires scientifiques. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 10(1), 51-64.  
<https://doi-org.lama.univ-amu.fr/10.4000/alsic.544>
- Crossley, S. A., Louwse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91(1), 15-30.  
<https://doi-org.lama.univ-amu.fr/10.1111/j.1540-4781.2007.00507.x>
- Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, 137, 57-75.  
<https://doi.org/10.1016/j.jecp.2015.03.005>
- Dascalu, M., Dessus, P., Trausan-Matu, Ș., Bianco, M., & Nardy, A. (2013, July). ReaderBench, an environment for analyzing text complexity and reading strategies. In *International Conference on Artificial Intelligence in Education* (pp. 379-388). Springer, Berlin, Heidelberg.  
[https://doi-org.lama.univ-amu.fr/10.1007/978-3-642-39112-5\\_39](https://doi-org.lama.univ-amu.fr/10.1007/978-3-642-39112-5_39)
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384-398.  
<https://doi.org/10.1016/j.neuron.2007.10.004>
- De Landsheere, G. (1963). Pour une application des tests de lisibilité de Flesch à la langue française. *Le travail humain*, 141-154.  
<http://www.jstor.org/stable/40659295>

- Dell'Orletta, F., Montemagni, S., & Venturi, G. (2011, July). Read-it: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the second workshop on speech and language processing for assistive technologies* (pp. 73-83). Association for Computational Linguistics.  
<https://www.aclweb.org/anthology/W11-2308.pdf>
- Dell'Orletta, F., Montemagni, S., & Venturi, G. (2014). Assessing document and sentence readability in less resourced languages and across textual genres. *ITL-International Journal of Applied Linguistics*, 165(2), 163-193.  
<https://doi.org/10.1075/itl.165.2.03del>
- Deltour, J. J., & Hupkens, D. (1980). *Test de Vocabulaire Actif et Passif (T. V.A.P.)*. Braine le Château, Belgique : Editions Applications des Techniques Modernes.
- Demonet, J. F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *Lancet*, 363(9419), 1451-1460.  
[https://doi.org/10.1016/S0140-6736\(04\)16106-0](https://doi.org/10.1016/S0140-6736(04)16106-0)
- Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, 35(2), 202-224.  
<https://doi.org/10.1598/RRQ.35.2.1>
- Écalle, J. (2006). *Timé-3: test d'identification de mots écrits*. Mot à mot éditions.
- Écalle, J., & Magnan, A. (2002). *L'apprentissage de la lecture : Fonctionnement et développement cognitifs*. Armand Colin.
- Fayol, M., & Jaffré, J. P. (1999). L'acquisition/apprentissage de l'orthographe. *Revue française de pédagogie*, 126(1), 143-170.
- Fellbaum, C. (1998). A semantic network of English: the mother of all WordNets. In *EuroWordNet: A multilingual database with lexical semantic networks* (pp. 137-148). Springer, Dordrecht.  
[https://doi-org.lama.univ-amu.fr/10.1007/978-94-017-1491-4\\_6](https://doi-org.lama.univ-amu.fr/10.1007/978-94-017-1491-4_6)
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3), 221.  
<https://doi.org/10.1037/h0057532>
- François, T. (2011). *Les apports du traitement automatique du langage à la lisibilité du français langue étrangère* (Doctoral dissertation, Ph. D. thesis, Université Catholique de Louvain. Thesis Supervisors: Cédric Fairon and Anne Catherine Simon).

- François, T. (2015). When readability meets computational linguistics: a new paradigm in readability. *Revue française de linguistique appliquée*, 20(2), 79-97.  
<https://doi-org.lama.univ-amu.fr/10.3917/rfla.202.0079>
- François, T., & Bernhard, D. (2014). When text readability meets automatic text simplification. *ITL-International Journal of Applied Linguistics*, 165(2), 89-96.  
<https://doi.org/10.1075/itl.165.2.00int>
- François, T., & Fairon, C. (2012, July). An AI readability formula for French as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 466-477). Association for Computational Linguistics.  
<https://www.aclweb.org/anthology/D12-1043.pdf>
- François, T., Billami, M., Gala, N., & Bernhard, D. (2016, July). Bleu, contusion, ecchymose: tri automatique de synonymes en fonction de leur difficulté de lecture et compréhension.. In JEP-TALN-RECITAL 2016 (Vol. 2, pp. 15-28).  
<https://hal.archives-ouvertes.fr/hal-01346538/file/T37.pdf>
- François, T., Müller, A., Degryse, B., & Fairon, C. *AMesure : une plateforme web d'assistance à la rédaction simple de textes administratifs*. In *Repères DoRiF 16*, DoRiF Università, Roma novembre 2018.  
<http://hdl.handle.net/2078.1/193864>
- Gala, N. & Javourey-Drevet, L. (2020c). Mots ‘faciles’ et mots ‘difficiles’ dans ReSyf : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité. *Lidil. Revue de linguistique et de didactique des langues*, (62).  
<https://journals.openedition.org/lidil/8373>
- Gala, N., & Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)* (pp. 59-66).  
[https://hal-amu.archives-ouvertes.fr/hal-01757941/file/GalaZiegler\\_CL4LC-2016.pdf](https://hal-amu.archives-ouvertes.fr/hal-01757941/file/GalaZiegler_CL4LC-2016.pdf)
- Gala, N., François, T., Javourey-Drevet, L., & Ziegler, J. C. (2018). La simplification de textes, une aide à l'apprentissage de la lecture. *Langue française*, (3), 123-131.  
<https://doi-org.lama.univ-amu.fr/10.3917/lf.199.0123>
- Gala, N., Tack, A., Javourey-Drevet, L., François, T., & Ziegler, J. C. (2020b, May). Alector: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. In *Language Resources and Evaluation for Language Technologies (LREC)*.  
<https://hal.archives-ouvertes.fr/hal-02503986/>

- Gala, N., Todirascu, A., Bernhard, D., Wilkens, R., & Meyer, J. P. (2020a, July). Transformations syntaxiques pour une aide à l'apprentissage de la lecture : typologie, adéquation et corpus adaptés. In *Congrès Mondial de Linguistique Française (CMLF 2020)*.  
<https://hal.archives-ouvertes.fr/hal-02562205/>
- Gala, N., Todirascu A., Javourey-Drevet, L., Bernhard, D., Wilkens, R. & Meyer J.-P. (2020d). Recommandations (*guidelines*) pour des transformations de textes en français afin d'améliorer leur lisibilité et leur compréhension par des enfants dyslexiques et faibles lecteurs.  
[alectorsite.wordpress.com/2020/11/05/recommandations-pour-adapter-des-textes](http://alectorsite.wordpress.com/2020/11/05/recommandations-pour-adapter-des-textes)
- Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The bells test: a quantitative and qualitative test for visual neglect. *International journal of clinical neuropsychology*, 11(2), 49-54.
- Gaussel, M. (2015). Lire pour apprendre, lire pour comprendre. *Dossier de veille de l'IFÉ*, (101).
- Gentaz, E., Sprenger-Charolles, L., Colé, P., Theurel, A., & Gurgan, M. (2013). Évaluation quantitative d'un entraînement à la lecture à grande échelle pour des enfants de CP scolarisés en réseaux d'éducation prioritaire: apports et limites. *Approche Neuropsychologique des Apprentissages chez l'Enfant (ANA E.)*, 123, 172-181.
- Gentaz, E., & Sprenger-Charolles, L. (2014). Bien décoder pour bien comprendre. *Cahiers pédagogiques*, (516)  
<http://www.cahiers-pedagogiques.com/Bien-decoder-pour-bien-comprendre-version-integrale>
- Gentaz E, Sprenger-Charolles L, & Theurel A (2015) Differences in the Predictors of Reading Comprehension in First Graders from Low Socio-Economic Status Families with Either Good or Poor Decoding Skills. *PLoS ONE* 10(3): e0119581.  
<https://doi.org/10.1371/journal.pone.0119581>
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. New Jersey : Hillsdale
- Gernsbacher, Morton A., & Mark E. Faust. "The mechanism of suppression: a component of general comprehension skill." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17.2 (1991): 245.  
<https://doi.org/10.1037/0278-7393.17.2.245>
- Givord, P. & Schwabe, M. *Note par pays, PISA 2018 \_ Direction de l'Éducation et des compétences :*  
[https://www.oecd.org/pisa/publications/PISA2018\\_CN\\_FRA\\_FRE.pdf](https://www.oecd.org/pisa/publications/PISA2018_CN_FRA_FRE.pdf)

- Goigoux, R. (2016). Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages. Rapport de recherche.  
<http://ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire>
- Goigoux, R., & Cèbe, S. (2013). *Lectorino & Lectorinette. Apprendre à comprendre les textes narratifs*. Paris : Retz.
- 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.  
<https://doi-org.lama.univ-amu.fr/10.4000/rfp.5076>
- Gombert, J. E., Martinot, C., & Nocus, I. (1996). Conscience linguistique et traitement de l'information écrite. *Revue Tranel (Travaux neuchâtelois de linguistique)*, 25, 135-153.  
[https://doc.rero.ch/record/19861/files/Gombert\\_Martinot\\_Nocus\\_135-153.pdf](https://doc.rero.ch/record/19861/files/Gombert_Martinot_Nocus_135-153.pdf)
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn Sci*, 15(1), 3-10.  
<https://doi.org/10.1016/j.tics.2010.10.001>
- Gougenheim, G. (1964). *L'élaboration du français fondamental (1er degré) : étude sur l'établissement d'un vocabulaire et d'une grammaire de base* (Vol. 1). Didier.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, 7(1), 6-10.  
<https://doi.org/10.1177/074193258600700104>
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, 101(3), 371.  
<https://doi.org/10.1037/0033-295X.101.3.371>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix : Analyse du texte sur la cohésion et le langage. *Méthodes de recherche sur le comportement, instruments et ordinateurs*, 36(2), 193-202.  
<https://doi-org.lama.univ-amu.fr/10.3758/BF03195564>
- Guillot, R. (2008). *Crin-blanc*. Livre de Poche Jeunesse.
- Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, 32(3), 282-313.  
<https://doi.org/10.1016/j.cedpsych.2006.05.004>

- Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: a cluster randomized trial. *Journal of Child Psychology and Psychiatry*, 58(10), 1132-1140.  
<https://doi.org/10.1111/jcpp.12762>
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, 111(3), 662.  
<https://doi.org/10.1037/0033-295X.111.3.662>
- Hattouti, J., Gil, S. & Laval, V. (2016). Le développement de la compréhension des expressions idiomatiques : une revue de littérature. *L'Année psychologique*, vol. 116(1), 105-136.  
<https://doi-org.lama.univ-amu.fr/10.4074/S0003503316000294>
- Hiebert, E. H., & Cervetti, G. N. (2012). What differences in narrative and informational texts mean for the learning and instruction of vocabulary. *Vocabulary instruction: Research to practice*, 2, 322-344.
- Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International journal of speech-language pathology*, 16(3), 199-207.  
<https://doi.org/10.3109/17549507.2014.904441>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, 2(2), 127-160.  
[http://phonicstrainingonline.com/wp-content/uploads/2015/06/simple\\_view.pdf](http://phonicstrainingonline.com/wp-content/uploads/2015/06/simple_view.pdf)
- Hulme, C., Snowling, M. J., West, G., Lervåg, A., & Melby-Lervåg, M. (2020). Children's language skills can be improved: lessons from psychological science for educational policy. *Current Directions in Psychological Science*, 29(4), 372-377.  
<https://doi.org/10.1177/0963721420923684>
- Inui, K., Fujita, A., Takahashi, T., Iida, R., & Iwakura, T. (2003, July). Text simplification for reading assistance: a project note. In *Proceedings of the second international workshop on Paraphrasing-Volume 16* (pp. 9-16). Association for Computational Linguistics.  
<https://www.aclweb.org/anthology/W03-1602.pdf>
- Jauhar, S. K., & Specia, L. (2012). Uow-shef: Simplex–lexical simplicity ranking based on contextual and psycholinguistic features. In \* *SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 477-481).  
<https://www.aclweb.org/anthology/S12-1066.pdf>

- Javourey-Drevet, L., François, T., Gala, N., Dufau, S., Ginestié, J. & Ziegler, J.-C. (en révision) Simplification of literary and scientific texts to improve fluency and comprehension in beginning readers of French. *Applied psycholinguistics*,
- Jeong, J., Gaffney, J. S., & Choi, J. O. (2010). Availability and use of informational texts in second-, third-, and fourth-grade classrooms. *Research in the Teaching of English*, 435-456. <https://www.jstor.org/stable/25704889>  
<http://www.jstor.org/stable/25704889>
- Johnson-Glenberg, M. C. (2005). Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders. *Reading and Writing*, 18(7-9), 755-786. <https://doi-org.lama.univ-amu.fr/10.1007/s11145-005-0956-5>
- Kandel, L., & Moles, A. (1958). Application de l'indice de flesch à la langue française. *Cahiers Etudes de Radio-Télévision*, 19(1958), 253-274.
- Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning disabilities research & practice*, 29(1), 10-16.  
<https://doi-org.lama.univ-amu.fr/10.1111/ldrp.12025>
- Keshavarz, M. H., Atai, M. R., & Ahmadi, H. (2007). Content schemata, linguistic simplification, and EFL readers' comprehension and recall. *Reading in a Foreign Language*, 19(1), 19-33.  
<https://files.eric.ed.gov/fulltext/EJ759838.pdf>
- Kim, Y. S. G., & Wagner, R. K. (2015). Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading*, 19(3), 224-242.  
<https://doi.org/10.1080/10888438.2015.1007375>
- Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly*, 52(3), 357-382.  
<https://doi.org/10.1002/rrq.171>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, 95(2), 163.  
<https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, W., (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

- Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse processes*, 39(2-3), 125-128.  
<https://doi.org/10.1080/0163853X.2005.9651676>
- Lafourcade, M. (2007, December). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th international symposium on natural language processing* (p. 7).  
<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00200883/>
- Laisney, P., Brandt-Pomares, P., & Ginestié, J. (2011). Influence de l'ordinateur sur l'activité d'enseignement. Le cas d'une situation en Technologie au collège. *Review of science, mathematics and ICT education*, 5(1), 9-26.  
<https://efe.library.upatras.gr/review/article/view/86>
- Lassault, J., & Ziegler, J. C. (2018). Les outils numériques d'aide à l'apprentissage de la lecture. *Langue française*, (3), 111-121.  
<https://doi-org.lama.univ-amu.fr/10.3917/lf.199.0111>
- Lecocq, P. (1996). L'ECOSSE : une épreuve de compréhension syntaxico-sémantique [The Ecosse : A syntactical semantic test]. Lille : Presses Universitaires du Septentrion.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36(1), 156-166.  
<https://doi-org.lama.univ-amu.fr/10.3758/BF03195560>
- Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821-1838.  
<https://doi.org/10.1111/cdev.12861>
- Lima, L., & Bianco, M. (1999). Le problème des références dans la compréhension des textes à l'école primaire : le cas de «il» et de «lui». *Revue française de pédagogie*, 83-95.  
<http://www.jstor.org/stable/41201432>
- Lima, L., Sylvestre, E., & Bianco, M. (2006). Lectures partagées et acquisition de stratégies de compréhension au cycle 3. *Apprentissage et enseignement*, 25-42.  
<http://www.frank-jamet.fr/sites/default/files/Lima%20et%20al%20%282006%29.pdf>
- Lima, L., Bianco, M., Guerin, L., Nanot, M., & Reale-Bruyat, F. (2016), *11 stratégies pour apprendre à comprendre les textes narratifs*. Paris : Hatier.

- Linderholm, T., Everson, M. G., Van Den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more-and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, 18(4), 525-556.  
[https://doi.org/10.1207/S1532690XCI1804\\_4](https://doi.org/10.1207/S1532690XCI1804_4)
- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational administration and supervision*, 9(389-398), 73.
- Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A. M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly (1982-)*, 514-546.  
<http://www.jstor.org/stable/23096205>
- Macchi, L., Casalis, S. & Schelstraete, M. A. (2016). La lecture chez les enfants avec des troubles spécifiques d'articulation, de parole et/ou de langage oral : une revue narrative de littérature. *L'Année psychologique*, vol. 116(4), 547-595.  
<https://doi-org.lama.univ-amu.fr/10.4074/S0003503316000439>
- Macchi, L., Casalis, S., & Schelstraete, M. A. (2019). Phonological and orthographic reading routes in French-speaking children with severe developmental language disorder. *Journal of communication disorders*, 81, 105909.  
<https://doi.org/10.1016/j.jcomdis.2019.05.002>
- Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science*, 15(1), 1-7.  
<https://doi.org/10.1111/j.0963-7214.2004.01501001.x>
- Marin, B. & Legros, D. (2008). Chapitre 4. Psycholinguistique cognitive de la compréhension de textes. Dans : B. Marin & D. Legros (Dir), *Psycholinguistique cognitive : Lecture, compréhension et production de texte* (pp. 61-78). Louvain-la-Neuve, Belgique : De Boeck Supérieur.
- Marin, B., Crinon, J., Legros, D., & Avel, P. (2007). Lire un texte documentaire scientifique : quels obstacles, quelles aides à la compréhension ? *Revue française de pédagogie. Recherches en éducation*, (160), 119-131.  
<https://doi-org.lama.univ-amu.fr/10.4000/rfp.786>
- Martins, D., & Le Bouédec, B. (1998). La production d'inférences lors de la compréhension de textes chez des adultes : une analyse de la littérature. *L'année psychologique*, 98(3), 511-543.  
<https://doi.org/10.3406/psy.1998.28581>

- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological review*, 99(3), 440.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, 51, 297-384.  
[https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- McNamara, D. S., Floyd, R. G., Best, R., & Louwerse, M. (2003). World knowledge driving young readers' comprehension difficulties.
- McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2017). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International electronic journal of elementary education*, 4(1), 229-257.
- Mesnager, J. (2002). Pour une étude de la difficulté des textes. *Le Français aujourd'hui*, (2), 29-40.  
<https://doi-org.lama.univ-amu.fr/10.3917/lfa.137.0029>
- Mioduser, D., Tur-Kaspa, H., & Leitner, I. (2000). The learning value of computer-based instruction of early reading skills. *Journal of computer assisted learning*, 16(1), 54-63.  
<https://doi-org.lama.univ-amu.fr/10.1046/j.1365-2729.2000.00115.x>
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological bulletin*, 137(2), 267.  
<https://doi.org/10.1037/a0021890>
- Morais, J. (1993). Compréhension/décodage et acquisition de la lecture. *Lecture-écriture : acquisition. Les actes de La Villette*, 10-21.
- 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>
- Morel, M., & François, J. (2015). Le Dictionnaire Electronique des Synonymes du CRISCO : un outil de plus en plus interactif. *Revue française de linguistique appliquée*, 20(1), 9-28.  
<https://doi-org.lama.univ-amu.fr/10.3917/rfla.201.0009>
- Morin, M. F., & Montésinos-Gelet, I. (2007). Effet d'un programme d'orthographe approchées en maternelle sur les performances ultérieures en lecture et en écriture d'élèves à risque. *Revue des sciences de l'éducation*, 33(3), 663-683.  
<https://doi.org/10.7202/018963ar>

- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Retrieved from Boston College, TIMSS & PIRLS International Study Center website  
<http://timssandpirls.bc.edu/pirls2016/international-results/>
- Nandiegou, M., Reboul, S. ; [Mémoire orthophonie] sous la direction de Núria Gala et Johannes Ziegler (2018). *La simplification lexicale comme outil pour faciliter la lecture des enfants dyslexiques et faibles lecteurs* ; Aix-Marseille Université.
- National Reading Panel Reports of the Subgroups. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child Health and Human Development
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE™. *L'année Psychologique*, 101(3), 447-462.  
<https://doi.org/10.3406/psy.2001.1341>
- 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.  
<https://doi.org/10.1007/s11145-008-9159-1>
- Ouellette, G., & Shaw, E. (2014). Oral vocabulary and reading comprehension: An intricate affair. *L'année psychologique*, 114(4), 623-645.  
<https://doi-org.lama.univ-amu.fr/10.4074/S0003503314004023>
- Paetzold, G., & Specia, L. (2015, July). Lexenstein. A framework for lexical simplification. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations* (pp. 85-90).
- Paetzold, G., & Specia, L. (2017). Lexical simplification with neural ranking. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (Vol. 2, pp. 34-40).
- Pappas, C. C. (1993). Is narrative “primary”? Some insights from kindergarteners' pretend readings of stories and information books. *Journal of Reading Behavior*, 25(1), 97-129.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, 11(4), 357-383.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86.

- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific studies of Reading*, 18(1), 22-37.
- Perfetti, C., McKeown, M. G., & Kucan, L. (2010). Decoding, vocabulary, and comprehension. *Bringing reading research to life*, 291-303.
- Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510-519
- Pourcin, L., Sprenger-Charolles, L., El Ahmadi, A., & Colé, P. (2016). Reading and related skills in Grades 6, 7, 8 and 9: French normative data from EVALEC. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 66(1), 23-37.
- Quaireau, C., Lavandier, K., De La Haye, F., Noël, Y., Le Bohec, O., & Nogues, J. (2016, September). TACIT-Ortho: un logiciel en ligne pour aider à comprendre l'implicite des textes. In *Entretiens de Bichat 2016-Orthophonie*.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Raven, J. (1998). *Progressive Matrices Couleurs (PM 38), Progressive Matrices Standard (PM 47), Étalonnage Français*. [Ravens Colored Progressive Matrices (PM 38), Standard Progressive Matrices (PM 47), French standardization]. Paris, France: ECPA.
- Rello, L., Baeza-Yates, R., Bott, S., & Saggion, H. (2013, May a). Simplify or help? Text simplification strategies for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (pp. 1-10).
- Rello, L., Baeza-Yates, R., Dempere-Marco, L., & Saggion, H. (2013, September b). Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction* (pp. 203-219). Springer, Berlin, Heidelberg.
- Rémond, G. Richaudeau, F. & Bentolila, A., ( Mai1991) a.r.t.h.u.r un atelier pour maîtriser la lecture, Nathan-Retz, mai 1991
- Richaudeau, F. (1969). La lisibilité. *Bibliothèque Médiations*.
- Rodd, J. M. (2020). Settling Into Semantic Space: An Ambiguity-Focused Account of Word-Meaning Access. *Perspectives on Psychological Science*, 15(2), 411-427.

- Rogde, K., Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review.
- Romero, M. (2020). *Digital games and learning*. Editions JFD.
- Roux-Baron, I., Cèbe, S., & Goigoux, R. (2017). Évaluation des premiers effets d'un enseignement fondé sur l'outil didactique Narramus à l'école maternelle. *Revue française de pédagogie*, (4), 83-104.
- Ruiter, M. B., Beijer, L. J., Cucchiaroni, C., Krahmer, E. J., Rietveld, T. C., & Strik, H. (2012). Human language technology and communicative disabilities: requirements and possibilities for the future. *Language Resources and Evaluation*, 46(1), 143-151.
- Rybing, J., Smith, C., & Silvervarg, A. (2010). Towards a rule-based system for automatic simplification of texts. *Proceedings of SLTC*.
- Safari, M., & Mohaghegh Montazeri, M. (2017). The Effect of Reducing Lexical and Syntactic Complexity of Texts on Reading Comprehension. *Journal of Teaching Language Skills*, 36(3), 59-83.
- Saggion, H. (2017). *Automatic Text Simplification: Synthesis Lectures on Human Language Technologies*, 1 ed. Morgan & Claypool Publishers; 2017.
- Saggion, H., Gómez-Martínez, E., Etayo, E., Anula, A., & Bourg, L. (2011). Text simplification in simplext: Making texts more accessible. *Procesamiento del lenguaje natural*, (47), 341-342.
- Schnedecker, C. (1997). *Nom propre et chaînes de référence*. Université de Lorraine Centre de recherche sur les médiations (CREM) (1 mars 1998)
- Seretan, V. (2012). Acquisition of syntactic simplification rules for French.
- Shardlow, M. (2014). A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1), 58-70.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151-218.
- Sheehan, K. M., Flor, M., & Napolitano, D. (2013, June). A two-stage approach for generating unbiased estimates of text complexity. In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility* (pp. 49-58).

- Siddharthan, A. (2014). A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2), 259-298.
- Simard, D., Foucambert, D., & Labelle, M. (2014). Examining the contribution of metasyntactic ability to reading comprehension among native and non-native speakers of French. *International Journal of Bilingualism*, 18(6), 586-604.
- Smith, C., & Jönsson, A. (2011, May). Automatic summarization as means of simplifying texts, an evaluation for Swedish. In *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)* (pp. 198-205).
- Schmoll, L. (2016). L'emploi des jeux dans l'enseignement des langues étrangères : Du traditionnel au numérique. *Sciences du jeu*, (5).  
<https://doi-org.lama.univ-amu.fr/10.4000/sdj.628>
- Sprenger-Charolles, L. (22-11-2017). *Lien entre décodage et compréhension*. Conférence dans le cadre du projet Lemon, Marseille.  
<http://grapholearn.fr/conferences/comprehension-ecriture-neurosciences/>
- Sprenger-Charolles, L., Colé, P., Béchenec, D., & Kipffer-Piquard, A. (2005). French normative data on reading and related skills from EVALEC, a new computerized battery of tests (end Grade 1, Grade 2, Grade 3, and Grade 4). *Revue européenne de psychologie appliquée/European Review of Applied Psychology*, 55(3), 157-186.
- Sprenger-Charolles, L., Colé, P., Piquard-Kipffer, A., & Leloup, G. (2010). *EVALEC, Batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture*. Ortho-Édition.
- Sprenger-Charolles, L., & Ziegler, J. C. (2019). Apprendre à lire : contrôle, automatismes et auto-apprentissage. Dans Bentolila, A & Germain, B (dir.), *L'apprentissage de la lecture*, 95-109, Nathan.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*, 32-71.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407.
- Stetter, M. E., & Hughes, M. T. (2010). Computer-assisted instruction to enhance the reading comprehension of struggling readers: A review of the literature. *Journal of Special Education Technology*, 25(4), 1-16.

- Swanson, H. L., Howard, C. B., & Sáez, L. (2007). Reading comprehension and working memory in children with learning disabilities in reading. *Children's comprehension problems in oral and written language: A cognitive perspective*, 157-185
- Thorndike, E. L. (1921). Word knowledge in the elementary school. *Teachers College Record*, 22(4), 334-370.
- Tremblay, O., & Polguère, A. (2014). Une ontologie linguistique au service de la didactique du lexique. In *SHS Web of Conferences* (Vol. 8, pp. 1173-1188). EDP Sciences.
- Todirascu, A., François, T., Bernhard, D., Gala, N., Ligozat, A. L., & Khobzi, R. (2017). Chaînes de référence et lisibilité des textes : Le projet ALLuSIF. *Langue française*, (3), 35-52.
- Todirascu, A., François, T., Gala, N., Fairon, C., Ligozat, A. L., & Bernhard, D. (2013). Coherence and cohesion for the assessment of text readability. *Natural Language Processing and Cognitive Science*, 11, 11-19.
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of learning disabilities*, 45(5), 453-466.
- Vajjala, S., & Meurers, D. (2014, April). Assessing the relative reading level of sentence pairs for text simplification. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 288-297).
- Valdois, S. (2010). Evaluation des difficultés d'apprentissage de la lecture. *Revue française de linguistique appliquée*, 15(1), 89-103.
- Van Dijk, T. A., et Kintsch, W. (1983). *Strategies of discourse comprehension*. New York : Academic Press
- Van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. *Developmental spans in event comprehension and representation: Bridging fictional and actual events*, 321-342.
- Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, 15(1), 8-25.
- Snowling, M. J. (2000). *Dyslexia*. Oxford: Blackwell Publishing.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, 101(2), 192.

- Wechsler, D. (2004). *The Wechsler intelligence scale for children-fourth edition*. London: Pearson Assessment.
- Wilkins, R., Dalla Vecchia, A., Boito, M. Z., Padró, M., & Villavicencio, A. (2014, November). Size does not matter. Frequency does. A study of features for measuring lexical complexity. In *Ibero-American Conference on Artificial Intelligence* (pp. 129-140). Springer, Cham.
- Willingham, D. T. (2006). The usefulness of brief instruction in reading comprehension strategies. *American Educator*, 30(4), 39-50.
- Wolfe, M. B. (2005). Memory for narrative and expository text: independent influences of semantic associations and text organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 359.
- Woodsend, K., & Lapata, M. (2011, July). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 409-420). Association for Computational Linguistics.
- Wright, G. A., Reeves, E., Williams, J., Morrison-Love, D., Patrick, F., Ginestié, J., Mammes, I. & Graube, G. (2018). Abridged international perspectives of technology education and its connection to STEM education. *International journal of education*, 10(4), 31. doi:10.5296/ije.v10i4.13704
- Yaneva, V., & Evans, R. (2015, September). Six good predictors of autistic text comprehension. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 697-706).
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3-29. doi:10.1037/0033-2909.131.1.3
- Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120397.
- Ziegler, J. C., Perry, C., & Zorzi, M. (2020). Learning to Read and Dyslexia: From Theory to Intervention Through Personalized Computational Models. *Current Directions in Psychological Science*, 0963721420915873.

- Ziegler, J. C., Pech-Georgel, C., Dufau, S., & Grainger, J. (2010). Rapid processing of letters, digits and symbols: what purely visual-attentional deficit in developmental dyslexia? *Developmental Science*, 13(4), F8-F14.
- Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, 107(1), 151-178.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal?. *Journal of experimental child psychology*, 86(3), 169-193.
- Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and language*, 93(3), 369-373.
- Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., George, F., Pech-Georgel, C., & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, 109(28), 11455-11459.

## Liste des sigles et acronymes

ANLCI : Agence Nationale de Lutte Contre l'Illettrisme

APC : Activités Pédagogiques Complémentaires

CASSA : Context-Aware Synonym Simplification Algorithm

CECRL : Cadre Européen Commun de Référence pour les Langues

CEDRE : Cycle des Evaluations Disciplinaires Réalisées sur Echantillon

C-I : Construction-Intégration

CENTAL : Centre pour le Traitement Automatique du Langage

CRISCO : Centre de Recherches Interlangues sur la Signification en Contexte

Depp : Direction de l'évaluation, de la prospective et de la performance

DES : Dictionnaire Électronique des Synonymes

DGL : Digital Game Learning

DVC : Décodage-Vocabulaire-Compréhension

IA : Intelligence Artificielle

IFé-ACTé : Institut Français d'éducation-Activité Connaissance Transmission éducation (EA 4281)

IREST : International Reading Speed Texts

JDC : Journée Défense et Citoyenneté

MENJ : Ministère de l'Éducation Nationale et de la Jeunesse

OCDE : Organisation de Coopération et de Développement Economiques

PIRLS : Progress in International Reading Literacy Study

PISA : Programme for International Student Assessment

QCM : Questionnaire à Choix Multiples

REP : Réseau d'Éducation Prioritaire

ROLL : Réseau des Observatoires Locaux de la Lecture

SAT : Simplification Automatisée du Texte

STARI : Strategic Adolescent Reading Intervention

STEM : Science, Technology, Engineering, and Mathematics

TIC : Technologie d'Information et Communication

SVR : Simple View of Reading

TOEFL : Test of English as a Foreign Language

**Tableaux**

Tableau 1 : Résultats des évaluations Repères, dans le domaine du français, en début de CE1, rentrée 2019.....	7
Tableau 2 : Variables de lisibilité sélectionnées pour les versions simplifiées et originales des textes, sur les trois années d'expérimentation. Ecart types entre parenthèses. ....	70
Tableau 3 : Statistiques descriptives pour les 11 tâches cognitives et langagières ainsi que les mesures de l'asymétrie et de l'aplatissement en CE1. ....	86
Tableau 4 : Statistiques descriptives des variables dépendantes du test de compréhension et des lectures de textes sous format numérique (moyenne sur l'ensemble des 20 textes) pour chaque année (CE1, CE2, CM1). ....	88
Tableau 5 : Résultats de l'ANOVA Simplification*Type de textes*Niveau en intra sujet pour la fluidité. ....	88
Tableau 6 : Effet de la simplification en fonction des niveaux scolaires pour la fluidité. ....	89
Tableau 7 : Résultats de l'ANOVA Simplification*Type de textes*Niveau en intra sujet pour le QCM.....	91
Tableau 8 : Statistiques descriptives de la variable dépendante « résumé » du test de compréhension sous format numérique en fonction des deux conditions (simplification et type de textes) pour chaque année, CE2 et CM1. ....	93
Tableau 9 : Résultats de l'ANOVA Simplification*Type de textes*Niveau en intra sujet pour les résumés. ....	93
Tableau 10 : Corrélations bivariées entre les variables cognitives et linguistiques individuelles et les gains en simplification relatifs aux scores pour la vitesse de lecture et la compréhension. ....	96
Tableau 11 : Résultats de l'ANOVA Simplification*Type de textes*Niveau en intra sujet* inter-sujets pour la variable fluence. ....	99
Tableau 12 : Résultats de l'ANOVA Simplification*Type de textes*Niveau en intra sujet* inter-sujets pour la variable compréhension.....	101
Tableau 13 : Corrélations entre les variables textuelles [XSIMP - XORIG] et les gains de simplification pour tous les textes en termes de vitesse de lecture et de compréhension. ....	105
Tableau 14 : Description des variables pour chacun des tests et pour chaque groupe.....	113
Tableau 15 : Résultats de l'ANOVA Simplification*Type de textes *Groupe pour la variable nombre de mots lus par seconde. ....	114
Tableau 16 : Résultats de l'ANOVA Simplification*Type de textes *Groupe pour la variable compréhension. ....	116
Tableau 17 : Statistiques descriptives pour les 9 tâches cognitives et langagières ainsi que les mesures de l'asymétrie et de l'aplatissement pour les trois groupes de CE2 (N=188). ....	125
Tableau 18 : Résultats de l'ANOVA trois factorielles (types de texte*gain*groupe) pour la variable nombre de mots lus par seconde.....	126
Tableau 19 : Résultats de l'ANOVA trois factorielles (types de texte*gain*groupe) pour la variable compréhension.....	130

**Figures**

Figure 1 : Structures et mécanismes impliqués dans la compréhension de textes (Bianco & Lima, 2017, p. 21). .....	2
Figure 2 : Modèle "simple" de la compréhension en lecture (Sprenger-Charolles & Ziegler, 2019, p. 96).....	10
Figure 3 : Modèle d'équation structurelle schématisant les effets de la compréhension orale et du décodage de mots sur la croissance des compétences de compréhension en lecture d'après Lervåg, Hulme et Melby-Lervåg (2018, p.1831). .....	12
Figure 4 : Modèle de la compréhension de Gernsbacher d'après Blanc et Brouillet (2003). ..	13
Figure 5 : Modèle de Construction-Intégration. Schéma réalisé d'après Kintsch (2005). .....	15
Figure 6 : Entrées lexicales interconnectées par l'entrée phonologique (Perfetti & Hart, 2002, p.194).....	17
Figure 7 : Triangle DVC. Les capacités de décodage, de vocabulaire et de compréhension se combinent pour produire des compétences générales en lecture (Perfetti, McKeown & Kucan 2010, p. 293).....	18
Figure 8 : Schéma simple du flux d'informations, de la connaissance sur la forme et la signification des mots à la compréhension, par l'identification des mots (Perfetti, McKeown & Kucan, 2010, p. 296). .....	18
Figure 9: Compréhension des mots dans le cadre des systèmes de lecture (Perfetti & Stafura, 2014, p. 33).....	19
Figure 10 : Modèle d'apprentissage de la lecture basé sur le décodage et l'auto-apprentissage des représentations orthographiques (Sprenger-Charolles & Ziegler, 2019, p. 101). .....	21
Figure 11 : « La fluidité de lecture orale est liée, mais distincte, de la fluence de décodage et de lecture de mots isolées. » (Bianco & Lima, 2017, p. 27). .....	23
Figure 12 : Moyenne des valeurs en score Z. Performances dans les sept domaines entre 1 an et 6,5 ans et dans les compétences en langue écrite pour les deux premières années de l'école d'après Lyytinen et al. (2006). .....	32
Figure 13 : Utilisation pédagogique des jeux numériques pour l'apprentissage (Romero, 2020, p. 22).....	42
Figure 14 : Exemple de simplification lexicale d'après le modèle de Shardlow (2014). .....	50
Figure 15 : Typologie des variations syntaxiques en vue d'une simplification de textes (Gala et al., 2020a).....	53
Figure 16 : Exemples de questions à choix multiples avec les propositions de réponses pour chacune. La réponse correcte est présentée en bleu. .....	64
Figure 17: Protocole expérimental conduit en deux phases chaque année. .....	80
Figure 18 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour la fluidité. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. .....	89
Figure 19 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour les textes littéraires pour la fluidité. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. .....	90

Figure 20 : Effet de la simplification par niveau scolaire en nombre de mots lus par seconde pour les documentaires scientifiques. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.....	90
Figure 21: Effet de la simplification par niveau scolaire en pourcentage de réponses correctes au QCM. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. .	91
Figure 22 : Effet du type de texte par niveau scolaire en pourcentage de réponses correctes au QCM. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	92
Figure 23: Corrélation entre les gains en compréhension (gain relatif) pour les textes simplifiés et originaux et un test de lecture à haute voix normalisé (lecture en 1 minute).....	97
Figure 24: Corrélation entre les gains en pourcentage de réponses correctes pour les textes simplifiés et originaux et un test standardisé de vocabulaire (Echelle de vocabulaire TVAP). .....	97
Figure 25 : Taille d'effet en score Z de la simplification pour les gains en fluence en fonction du niveau initial en vocabulaire des élèves. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	100
Figure 26: Interaction non-significative de la simplification en fonction des 3 groupes de simplification (Lexicale, Syntaxique, Lexico-syntaxique) pour la fluidité de lecture. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.....	115
Figure 27 : Effet de la simplification en fonction du type de textes au niveau de la fluidité de lecture. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ..	115
Figure 28 : Interaction non significative de la simplification en fonction des 3 niveaux de simplification : lexical, syntaxique et lexico-syntaxique en compréhension. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	116
Figure 29 : Effet de la simplification en fonction du type de texte au niveau de la compréhension. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	117
Figure 30 : Taille d'effet en score Z de l'entraînement (pré-test-post-test) pour les gains en fluence en fonction des trois groupes. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet.....	127
Figure 31 : Effet de l'entraînement entre pré-test et post-test en fluence pour les trois types d'entraînement en fonction du type de textes. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	127
Figure 32 : Gains en nombre de mots lus par seconde entre les pré et post-tests en fonction de la vitesse moyenne du nombre des mots à haute voix en effet fixe. ....	129
Figure 33 : Gains en nombre de mots lus par seconde entre les pré et post-tests en fonction du test intelligence non verbale en effet fixe.....	129
Figure 34 : Moyenne des réponses correctes au QCM des textes littéraires en fonction des pré et post-tests pour les trois types d'entraînement. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	130
Figure 35 : Moyenne des réponses correctes au QCM des documentaires scientifiques en fonction des pré et post-test pour les trois types d'entraînement. Les barres d'erreurs représentent 95% des intervalles de confiance en intra sujet. ....	131
Figure 36 : Présentation de l'interface permettant une recherche multi-critères des textes disponibles.....	142

## Figures

Figure 37 : Présentation d'un texte parallèle avec la police Open Dyslexic. ....	142
Figure 38 : Vues de l'application, phase des choix possibles dans l'outil.....	143
Figure 39 : Vues de l'application, lecture et résumé. ....	144
Figure 40 : Pages d'un jeu de décision lexicale et page générale des scores.....	145

## Annexe A

## Alector: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers

Núria Gala<sup>1</sup>, Anaïs Tack<sup>2,3,4</sup>, Ludivine Javourey-Drevet<sup>5,6</sup>, Thomas François<sup>2</sup>, Johannes C. Ziegler<sup>5</sup>

<sup>1</sup>Aix Marseille Univ, Laboratoire Parole et Langage, LPL CNRS (UMR 7309), France

<sup>2</sup>CENTAL, UCLouvain <sup>3</sup>ITEC, imec research group at KU Leuven <sup>4</sup>F.R.S.-FNRS Research Fellow, Belgium

<sup>5</sup>Aix Marseille Univ, Laboratoire de Psychologie Cognitive, LPC CNRS (UMR 7290), France

<sup>6</sup>Aix Marseille Univ, Apprentissage, Didactique, Évaluation, Formation (EA 4671), France

{nuria.gala,ludivine.javourey,johannes.ziegler}@univ-amu.fr

{anaïs.tack,thomas.francois}@uclouvain.be

### Abstract

In this paper, we present a new parallel corpus addressed to researchers, teachers, and speech therapists interested in text simplification as a means of alleviating difficulties in children learning to read. The corpus is composed of excerpts drawn from 79 authentic literary (tales, stories) and scientific (documentary) texts commonly used in French schools for children aged between 7 to 9 years old. The excerpts were manually simplified at the lexical, morpho-syntactic, and discourse levels in order to propose a parallel corpus for reading tests and for the development of automatic text simplification tools. A sample of 21 poor-reading and dyslexic children with an average reading delay of 2.5 years read a portion of the corpus. The transcripts of readings errors were integrated into the corpus with the goal of identifying lexical difficulty in the target population. By means of statistical testing, we provide evidence that the manual simplifications significantly reduced reading errors, highlighting that the words targeted for simplification were not only well-chosen but also substituted with substantially easier alternatives. The entire corpus is available for consultation through a web interface and available on demand for research purposes.

**Keywords:** Parallel corpora, text simplification, readability, linguistic complexity, misreading, poor-readers, dyslexia

## 1. Introduction

Reading is a complex cognitive task. Since reading comprehension is necessary for all school learning activities, poor reading and comprehension skills compromise children's academic and professional success. Typical readers also tends to progress quickly in reading because, as the process becomes more and more automatized, they enter a virtuous circle in which good reading comprehension skills boosts word identification and vice-versa (Stanovich et al., 1986; Stanovich, 2009). On the contrary, a child facing difficulties will tend to read less and therefore will not enter this virtuous circle. His/her reading difficulties will increase as the grade level becomes more demanding in terms of reading speed and comprehension (Tunmer and Hoover, 2019).

Given that reading comprehension skills of French-speaking students have decreased over recent years (Mullis et al., 2017), we have decided to address this issue in the framework on the Alector project<sup>1</sup>. Our aim was to develop and to test resources that make it possible to propose simplified texts to children facing problems in reading. For these children, text simplification might be a powerful and possibly the only way to leverage document accessibility. The idea is not to impoverish written language, but to propose simplified versions of a given text that convey the exact same meaning. The main assumption is that the simplification of a text will allow children with reading difficulties to eventually get through a text and thus discover the pleasure of reading through understanding what they actually read. This will allow them to enter the above mentioned

virtuous circle, whereby word recognition and decoding skills are trained through reading more. The promise of this enterprise is that training children on simpler texts will lower their give-up threshold and improve their decoding, word recognition and comprehension skills, which ultimately would allow them to move on to more complex texts.

In order to test our hypothesis on text simplification and readability, we compiled a corpus of 183 texts (including 79 authentic texts), which was tested in schools during a three-year study. In this paper, we describe the corpus, its possibilities of use, and its availability. The resource is mainly addressed to a community of professionals interested in helping French-speaking learners who struggle with learning to read. It could also be of interest for research, i.e. for developing and training automatic text simplification systems.

The paper is organized as follows. In Section 2., we give an overview of related work (currently available simplified corpora and annotated corpora with errors). In Section 3., we specify how the corpus was created and provide quantitative details about it. Section 4. describes how a sub-part of the corpus was annotated with reading errors from poor and dyslexic readers.

## 2. Related work

The use of corpora is essential in many domains for different purposes. For reading, there are a number of standardized reading tests such as the International Reading Tests (IReST) (Vital-Durand, 2011) which exists in a variety of languages. However, standardized or specifically annotated corpora (i.e., with errors) are very costly to build and not al-

<sup>1</sup><https://alectorsite.wordpress.com/>

ways available. In this section, we first report on resources for text simplification that are similar to ours in the sense that they include parallel original and simplified texts. Second, we discuss previous resources having reading errors annotated.

### 2.1. Parallel Corpora for Text Simplification

Researchers in supervised text simplification have used English Wikipedia – Simple English Wikipedia (EW–SEW) sentence-aligned corpora for training the systems (Štajner and Nisioi, 2018). Zhu et al. (2010) were the first to investigate this way of collecting simpler versions of texts. Soon after, Coster and Kauchak (2011) built a corpus of 137K aligned sentence pairs and computed transformations to compare original to simplified sentences (rewordings, deletions, reorders, mergers and splits). This approach received much attention from researchers working on text simplification in English, until it got criticized by Xu et al. (2015). More recently, (Xu et al., 2015) have advocated for the use of the Newsela corpus (Newsela, 2016) in automatic text simplification and demonstrated its value. This corpus was initially developed by editors and intended for learners of English. The data-set (version 2016-01-29.1) is composed of 10,787 news articles in English: 1,911 articles in their original form and in 4 equivalent versions rewritten by humans to suit different reading levels. Having different versions of an original article offers a great potential to study linguistic transformations, which explains why the automatic text simplification community was eager to use this resource. However, the way levels were defined - using the Lexile formula (Stenner, 1996) - should be subject to caution. To the best of our knowledge, there is no equivalent corpus for other languages offering the possibility to align sentences at different levels of reading complexity.

### 2.2. Corpora of Reading Errors

Apart from the use of parallel simplified corpora as a gold standard for text simplification, a number of studies have resorted to empirical measures of cognitive difficulty when reading in a native or foreign language including, but not exhaustively, eye-tracking data on readers suffering from autism spectrum disorder (Yaneva et al., 2016; Štajner et al., 2017) or from dyslexia (Bingel et al., 2018), as well as subjective annotations of difficult words (Paetzold and Specia, 2016; Tack et al., 2016; Yimam et al., 2018).

As far as reading errors are concerned, the situation can be considered unsatisfactory given that we find a limited availability of resources of this type. For dyslexia, some examples of corpora with writing errors have been compiled, for instance that of (Pedler, 2007), a dataset of productions of dyslexic English readers (3,134 words and 363 errors), or Dyslist (Rello et al., 2014), a corpus of 83 texts written in Spanish by dyslexic children (with 887 misspelled words). Such corpora can be used as source of knowledge to study different aspects of dyslexia. They can also be used to develop tools such as spellcheckers and games, and for screening with applications for readers (Rauschenberger et al., 2019), e.g. Dytective (Rello et al., 2016) a web-based game with different stages to detect dyslexia with machine learning prediction models.

Nb ORIG	Nb SIMP	Level	Type of corpus	
10	15	IReST	1 LIT	9 SCI
25	45	CE1	15 LIT	10 SCI
24	24	CE2	14 LIT	10 SCI
20	20	CM1	10 LIT	10 SCI
79	104		40	39

Table 1: Distribution of the original versions (LITerary and SCientific texts) across primary school levels: CE1 (second grade), CE2 (third grade), CM1 (fourth grade).

Nb	Type of Simplification
79	Lexical, morpho-syntactic, discourse
15	Lexical simplification only
10	Syntactic simplification only
104	Total simplified versions

Table 2: Type and number of simplified versions in the corpus.

To the best of our knowledge, there is currently no available data set for French with reading errors of poor readers and dyslexics.

## 3. Alector Corpus: Simplified Texts in French

The Alector corpus is a collection of 79 original literary (tales, stories) and scientific (documentary) texts along with their simplified equivalents (see Table 1). The texts were chosen among a variety of materials available for students in French primary schools<sup>2</sup>. We targeted second to fourth grades, i.e. beginning readers (in the French educational system, this corresponds to *cours élémentaire 1* and 2 (CE1 and CE2), and *cours moyen 1* (CM1), as described in Table 1). We also included the French version of the IReST corpus (Vital-Durand, 2011), a set of 10 standardized texts usually used for assessment of reading performances.

Our focus was on literary and scientific genres. While literary texts reflect the world view and the sensitivity of its author with a language that emphasizes the aesthetics, the rhythm, etc. (they highlight the poetic or expressive function of the language), scientific (documentary) texts aim at explaining or describing a scientific or technological causality, they are descriptive and explanatory with a logical structure based on scientific reasoning.

All the 79 original texts underwent simplifications at four linguistic levels, namely lexical, morphological, syntactic, and discursive (see Section 3.1.). In addition, 5 IReST and 15 CE1 texts were also simplified only at the lexical level, whereas 10 CE1 texts were simplified using only syntactic strategies. As a result, the total amount of simplified texts is 104, as detailed in Table 2.

<sup>2</sup>In more detail, we selected extracts from books in the Antoon Krings' collection "Drôles de petites bêtes", from the "J'aime Lire" collection or from Goscinny Sempé's book "Le Petit Nicolas". Scientific extracts were selected from Wapiti, Bibliothèque de Travail Junior (BTJ) or Images DOC, which are standard scientific magazines in French for young readers.

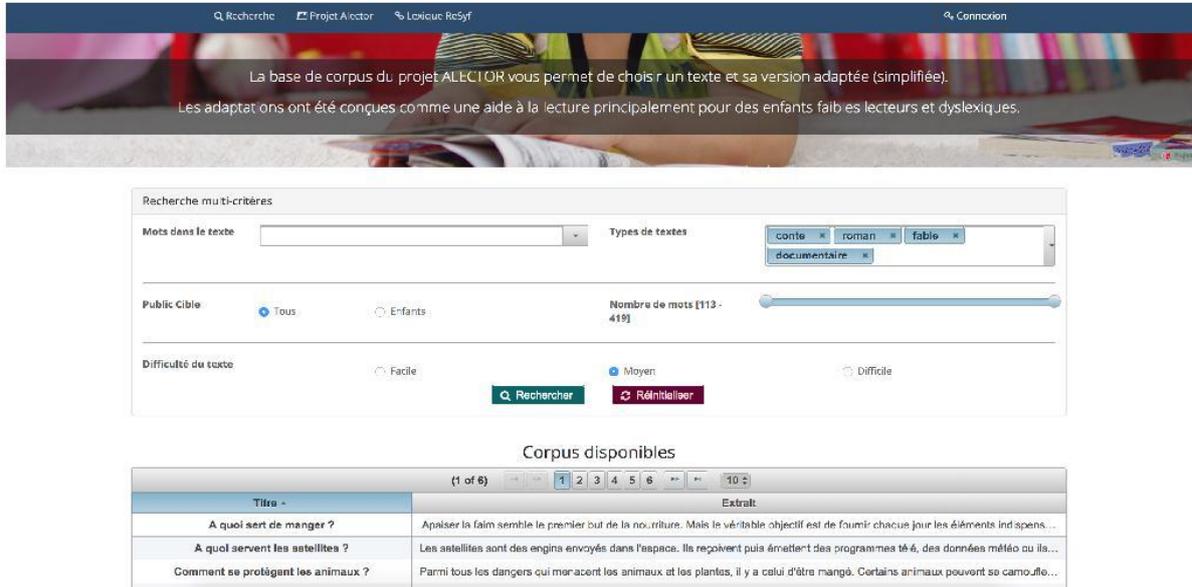


Figure 1: Interface for searching a corpus.

### 3.1. Manual Simplifications

Text simplification is defined as a reduction of the complexity of a text while preserving its original content (Saggion, 2017). By doing this, the text may be more easily read and understood by a target audience (in our work, we focus on the needs of poor-readers and dyslexic children).

From an initial set of 79 original texts, 104 different simplified versions were manually carried out by a group of researchers in educational sciences, cognitive psychology, linguistics and speech therapy. The objectives were twofold: (a) to keep the simplified text as close as possible to the original version; (b) to consider only linguistic transformations that could be later implemented in an automatic text simplification system (e.g. lexical and coreference chains substitutions, paraphrasing). The resulting simplified versions, along with their originals, were tested in five classes during current reading activities (children read the texts without knowing that they were reading adapted versions). On the other hand, simplified versions will be used as reference (gold-standard) to train and to evaluate a text simplification system. To our knowledge, despite Wikidia free adaptations of Wikipedia, our corpus is the first one to propose simplified versions of original texts in French.

Linguistically speaking, manual simplifications were done at four levels or dimensions (Gala et al., 2018): lexical (lexical substitutions), morphological (grammatical changes), syntactic (phrase structure adaptations), discourse (anaphora resolution). Guidelines were established following recommendations in the literature and after studying edited material for children (with and without dyslexia) and for illiterate adults (i.e. belgian collection *La Traversée*<sup>3</sup>).

- LEXICAL REPLACEMENTS were performed using Manulex<sup>4</sup> (Lété et al., 2004), ReSyf<sup>5</sup> (Billami et al., 2018) and Lexique<sup>6</sup> (New et al., 2001), three available lexical resources in French that contain indications of the presence of a word in a school level (Manulex), reading difficulty grades (ReSyf) and word frequencies in standard oral and written French (Lexique 3). We took into account specifications already defined in Gala and Ziegler (2016) and François et al. (2016). For instance, long and less frequent words, with irregular graphemes and complex syllable structures were modified by simpler synonyms: e.g. *volumineux* (unwieldy) by *gros* (big) and *ressemblaient* (seemed) by *étaient* (were).
- MORPHOLOGICAL SIMPLIFICATIONS dealt with complex verb forms replacements: e.g., *elle peut devenir* ('she can become') was replaced by *elle devient* ('she becomes'); replacements in a same morphological family by more frequent equivalents in the same family, e.g., *construction* replaced by the infinitive form *construire* ('to build'), and diminutive morphemes deletion, e.g., *maisonnette* ('little house') became *maison* ('house').
- SYNTACTIC TRANSFORMATIONS included changes on the sentence structures, such as deletion of subordinates and of complex modifiers, transformation of passive voice to active voice or of negative sentences to positive, etc. For instance, the original construction *C'est le vent qui apporte la pluie* ('it is the wind that brings the rain') becomes *Le vent apporte la pluie*

<sup>4</sup><http://www.manulex.org/>

<sup>5</sup><https://cental.uclouvain.be/resyf/>

<sup>6</sup><http://www.lexique.org/>

<sup>3</sup><http://www.lire-et-ecrire.be/latraversee>

The screenshot shows the Alector web interface. At the top, there is a search bar with the text 'Recherche', 'Projet Alector', and 'Lexicoue RéSyl'. To the right, there is a 'Déconnexion' button. Below the search bar, there is a navigation bar with 'Grand interligne' selected. The main content area is titled 'Emilie et le crayon magique' and contains two columns of text. The left column is labeled 'Original (Henriette Bichonnier)' and the right column is labeled 'Simplifié (Alector)'. The text in both columns is displayed in a large font with wide line spacing and medium character spacing. The original text is on the left and the simplified text is on the right.

Figure 2: Interface of visualisation with medium font size, big interlinear spacing and medium intercharacter spacing.

(‘the wind brings the rain’), the later keeps the standard subject-verb-object (SVO) order.

- d. DISCOURSE SIMPLIFICATION mainly dealt with replacements of pronouns with the referenced expression, e.g., *il y plonge la main* (‘he thrusted his hand’) by *il mit la main* (‘he put his hand’) and the pronoun *y* is replaced by the full referent *dans son cartable* (‘in his schoolbag’).

Details on all types of transformations can be found in the simplification guidelines, soon available on the webpage of the Alector project (see Footnote 1).

### 3.2. Corpus Analysis

A statistical analysis of the Alector corpus was carried out. The corpus contains a total of 52,704 tokens, with a global average number of 288 tokens/text, distributed as showed in Table 3. When looking in more details within the different components of the corpus – defined by genre (LIT vs SCI), condition (original vs simplified) and levels (CE1, CE2, CM1) –, several interesting facts appear. First, simplified texts are shorter on average (275 words/text) than original texts (306 words/text). This is the case for every grade level as can be seen when one compares the two total columns in Table 3. Second, it is also obvious that literary texts are on average longer than scientific documents not only for original versions (339 vs 271 words/text), but also for their simplified equivalents (313 vs 239 words/text). Finally, we also observed that texts become longer on average as their grade level increases, both for original and for simplified versions.

In Table 4, we can notice that scientific documents, although shorter, tend to have longer sentences either in the original (16.1 vs 15 words/sentence) or in the simplified

version (13.5 vs 11.2 words/sentence). This difference between genre is more noteworthy as the grade level increases. As regards the grade level in general, original versions of texts do not seem to vary much as regards to sentence length, whereas a slight trend can be seen in simplified versions. However, Table 4 clearly shows that the simplification process has reduced sentence length (15.5 vs 12.4 words/sentence).

We also report the proportion of nouns in the text, which roughly corresponds to the conceptual density of texts. Without surprise, scientific documents include a higher density of nouns, as they refer to various notions and realities. There also seems to occur an increasing proportion of nouns as the grade level rises, with the notable exception of the IReST literary text. Finally, it is interesting to note that the proportion of nouns does not decrease as a result of the simplification process (23% vs 22.6%). Most likely, this can be explained by the attempt to make explicit some of the pronominal anaphora, using the referent instead of the pronoun. In sum, it is clear that the simplified documents have different properties from the original texts.

### 3.3. Availability

The Alector corpus is available through a web interface. A text can be searched from a keyword, or by introducing some predefined labels on the kind of document (narrative, documentary, etc.). The number of words in the text or its difficulty to be read by the students can also be selected (Figure 1).

The “text difficulty” score (easy, medium, difficult) is based on the reading and comprehension tests of 164 students. We have used a normalized average of two factors, namely the reading speed and the comprehension scores obtained from each student.

The result of the search through the web interface is a list

Level	Original texts			Simplified texts		
	LIT	SCI	Total	LIT	SCI	Total
IResT	1 (135)	9 (141)	10 (140)	1 (115)	9 (131)	10 (129)
CE1	15 (282)	10 (243)	25 (266)	25 (276)	25 (211))	50 (244)
CE2	14 (363)	10 (303)	24 (338)	14 (340)	10 (288)	24 (318)
CM1	10 (413)	10 (385)	20 (399)	10 (388)	10 (358)	20 (373)
Total	40 (339)	39 (271)	79 (306)	50 (313)	54 (239)	104 (275)

Table 3: Number of texts per level, with the average number of words for each component.

Level	Original texts			Simplified texts		
	LIT	SCI	Total	LIT	SCI	Total
IResT	16.9 / 28.1%	16.2 / 21.6%	16.3 / 22.3%	14.4 / 27.8%	13.7 / 21.8%	10 / 22.4%
CE1	15.9 / 19.8 %	15 / 24.7%	15.6 / 21.7%	10.2 / 20.2%	12.9 / 24.9 %	11.6 / 22.5 %
CE2	14.9 / 20.2 %	15.5 / 26.4%	15.2 / 22.8%	12.4 / 20.8%	13.8 / 27%	12.9 / 23.4 %
CM1	13.5 / 22%	17.8 / 25.4%	15.6 / 23.7%	11.9 / 22.6%	14.4 / 26.3%	13.2 / 24.4 %
Total	15 / 20.7%	16.1 / 24.6%	15.5 / 22.6%	11.2 / 21 %	13.5 / 25 %	12.4 / 23%

Table 4: Average number of words per sentence and percentage of nouns per level.

of texts corresponding to the selected criteria. When the user clicks on the beginning of the text s/he has access to the parallel corpora (Figure 2).

Following the guidelines developed in the project and based on existing guidelines for other languages, e.g. Spanish (Rello, 2014) or English (British Dyslexia Association<sup>7</sup>), the interface enables different possibilities of visualization in terms of font (Open Dyslexic), interline spacing and intercharacter spacing (Zorzi et al., 2012).

#### 4. A Subcorpus with Alignments of Misreadings

From the parallel corpus, twenty texts (i.e., ten authentic texts with their simplified version) were selected for use in speech therapy interventions with reading-impaired children. Based on the transcripts of the read-aloud interventions (Section 4.1.), we aligned and aggregated all errors made by each subject with the correct word in the base text (Section 4.2.). As a result, we were able to identify the probability of misreading a word in an authentic and a simplified text in the sample of dyslexic children. Using this error probability as a measure of lexical difficulty, we found evidence that the manual lexical simplifications had a significant alleviating effect on reading difficulties (Section 4.3.).

##### 4.1. Tests with Dyslexic Readers

Reading tests were conducted in a sample ( $N=21$ ) of French-speaking children aged between 9 and 12 attending mainstream schools. The subjects had a reading delay of two and a half years on average. During the experiment, the participants were asked to read aloud 10 different texts, including 5 original and 5 simplified texts drawn from both the literary and scientific genres. The texts were relatively

short, counting 296 (literary) and 124 (scientific) tokens on average. All texts were presented on a digital tablet, displaying the sentences one by one. The children had to click to move to the next sentence. After each text, the participants were asked to answer a reading comprehension test in a multiple-choice format (read by the therapist to avoid a reading bias when evaluating comprehension). The experiments were run in private practices of speech-language therapists in Marseille (France), between November 2017 and March 2018. The read-aloud data were recorded and transcribed by students majoring in speech therapy (Nandiegou and Reboul, 2018). They manually encoded reading errors using ad-hoc guidelines.

##### 4.2. Aligning Reading Errors

In order to identify the reading errors in the original and simplified texts, we aggregated the read-aloud transcripts of all participating children. First, because the transcripts had not been encoded with specialized annotation or transcription software and had not been linked to their original text, we needed to find a way to align each one of the 210 transcripts (i.e., ten transcripts per participant) with the original text. The alignment was done on the level of the word, associating each word as it was read aloud with the word as it appeared in the text. To this end, a modified version of the Needleman and Wunsch (1970) sequence alignment algorithm was used. The simplicity of the algorithm did not seem problematic given that the read-aloud transcripts did not contain any major syntactic modifications. The modified version of the algorithm aligned two words by constructing a similarity matrix for each pair of sentences in the original text ( $o$ ) and the read-aloud transcript ( $r$ ), represented as a sequence of tokens. The similarity scores  $s$  between two tokens  $w_o$  and  $w_r$  were computed by the integration of the edit distance and the length of a word (1). The modified algorithm thus privileged an alignment of words and non-words that displayed a higher formal proximity

<sup>7</sup><https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide>

Original version (10 readings)												
TEXT	Voilà	que	tu	t'	agenouilles		devant	ce	tas	de	neige.	
MISREADINGS		<i>qui</i>			<i>agenou</i> <i>agenoui</i> <i>agenouiller</i> <i>agenouillies</i> <i>angenouillies</i> <i>aquenoulés</i>			<i>cette</i> <i>ça</i>				
ERROR COUNT	0	1	0	0	7		0	2	0	0	0	
ERROR PROB.	0.0	0.1	0.0	0.0	0.7		0.0	0.2	0.0	0.0	0.0	
GLOSS	'Now'	'you are'			'kneeling'		'in front of'	'this'	'pile'	'of'	'snow'.	
Simplified version (11 readings)												
TEXT	Voilà	que	tu	te	mets	à	genoux	devant	ce	tas	de	neige.
MISREADINGS				<i>me</i>					<i>ces</i> <i>te</i>			
ERROR COUNT	0	0	0	1	0	0	0	0	4	0	0	0
ERROR PROB.	0.0	0.0	0.0	0.09	0.0	0.0	0.0	0.0	0.36	0.0	0.0	0.0
GLOSS	'Now'	'you are'			'kneeling'			'in front of'	'this'	'pile'	'of'	'snow'.

Table 5: Example of alignment with reading errors on a lexically simplified text with substitutions (literary text 3).

(e.g., *jardin* and *jadi*).

$$s = \begin{cases} +1, & \text{if } w_o = w_r \\ -(1 + \delta_{\text{Levenshtein}}(w_o, w_r)), & \text{if } w_o \neq w_r \\ -(1 + |w|), & \text{as gap penalty} \end{cases} \quad (1)$$

Second, we aggregated all per-participant alignments so as to obtain a list of all different variants read aloud for each word attested in the original text. Table 5 shows an example of sentence in the original corpus in which each word is aligned with observed reading errors. We consider a misreading when a word is either mispronounced by the use of another word or non-word (i.e., substitutions) or skipped (i.e., deletions).<sup>8</sup> For each word, we compute the total number of reading errors  $C_{\text{misreading}}(w)$  and the probability of the word being misread  $C_{\text{misreading}}(w)/C_{\text{reading}}(w)$ .

### 4.3. Analysis

There are two ways in which the aggregated misreadings can be useful to further our understanding of lexical difficulty. On the one hand, the data can be used to identify difficult words in a bottom-up manner, making use of empirical evidence of reading errors as the basis for data-driven lexical simplification. However, we find that further work needs to be done to obtain a valid heuristic of difficulty on this data. From the two excerpts listed in Tables 5 and 6, we see that not all reading errors appear to be equally indicative of lexical difficulty. Indeed, we find many misreadings to be highly idiosyncratic, with only one or two participants mispronouncing the word. This trend is confirmed when we look at the distribution of the probability to misread a word in the original and simplified texts. Figure 3 shows that more than half of the tokens appearing in both versions were correctly read by all subjects, whereas only a small percentage of tokens were mispronounced by all subjects. Also, from the examples in Tables 5 and 6, we see that not

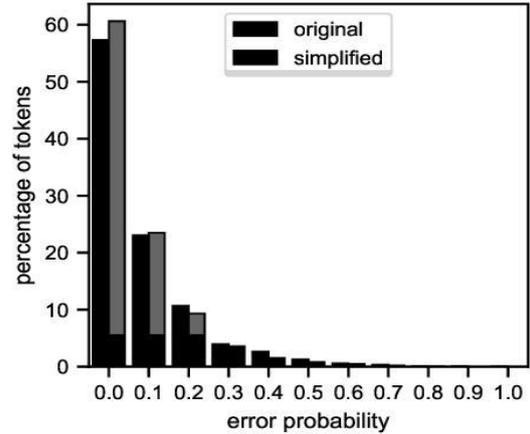


Figure 3: Distribution of the probability to misread a word token in the sample of poor-reading and dyslexic children

all errors might be equally indicative of reading problems. While problems can be unquestionably perceived in errors where either non-existent words (e.g., *aquenoulés*) or existing words with radically different grammatical categories (e.g., *jardin* [noun] instead of *jadis* [adverb]) are produced, a number of errors are concerned with a slight change of form without radically changing the grammatical category (e.g., *qui* instead of *que*). In future work, we therefore aim to further investigate how both this individual and gradual nature of lexical difficulty can be correctly accounted for.

On the other hand, the data can also be used to validate the top-down identification of lexical difficulty as attested in the manually simplified corpus. From the examples in Tables 5 and 6, we see that the only words that were tar-

<sup>8</sup>For now, we do not consider insertions.

Original version (11 readings)												
TEXT	Il	y	avait	<b>jadis</b>	en	Irlande	un	homme	du	nom	de	Jack.
MISREADINGS			<i>avant</i>	<i>jadi</i>		<i>Arlande</i>						
			<i>était</i>	<i>jardin</i>								
			<i>jardis</i>									
ERROR COUNT	0	0	2	<b>7</b>	0	1	0	0	0	0	0	0
ERROR PROBABILITY	0.0	0.0	0.18	<b>0.64</b>	0.0	0.09	0.0	0.0	0.0	0.0	0.0	0.0
GLOSS	'There'	'was'	'was'	' <b>once</b> '	'in'	'Ireland'	'a'	'man'		'named'		'Jack'.
Simplified version (10 readings)												
TEXT	Il	y	avait		en	Irlande	un	homme	du	nom	de	Jack.
MISREADINGS		∅				<i>un</i>	<i>rilande</i>			<i>de</i>		<i>Jean</i>
ERROR COUNT	0	1	0		1	1	0	0	2	0	0	1
ERROR PROBABILITY	0.0	0.1	0.0		0.1	0.1	0.0	0.0	0.2	0.0	0.0	0.1
GLOSS	'There'	'was'	'was'		'in'	'Ireland'	'a'	'man'		'named'		'Jack'.

Table 6: Example of alignment with reading errors on a lexically simplified text with deletions (literary text 4).

geted for simplification (i.e., *agenouiller* ['to kneel'] and *jadis* ['erstwhile']) are also the only words that we can consider to be difficult for the target reader population given the high error probability ( $p = .70$  and  $p = .64$ , resp.). This leads us to believe that the manual simplifications were indeed necessary and that the substitutions were adequate.

In the following two sections, we aim to further substantiate this claim with statistical<sup>9</sup> evidence. Our first hypothesis will be that the lexical simplifications can be considered necessary if the words in the authentic texts that were initially considered to be difficult for the target population (and hence targeted for lexical simplification) will be more often misread than words that were not targeted (Section 4.3.1.). Our second hypothesis will be that lexical simplifications can be considered correctly estimated if there are less misreadings on the substitute words than on the original words (Section 4.3.2.).

#### 4.3.1. Misreadings of words in authentic texts targeted for simplification

To confirm our first hypothesis, we focus on the words attested in the authentic texts only and divide them into into three different categories, viz., words that were targeted for simplification with either substitution or deletion and words that were not targeted. We then examined whether the probability of misreading words belonging to either category was significantly different, which was confirmed by a Kruskal-Wallis test,  $\chi^2(2) = 199$ ,  $p < .001$ ,  $\epsilon^2 = 0.089$ . Post-hoc Dwass-Steel-Critchlow-Fligner comparisons (Figure 7) showed that the largest difference was attested between words that were substituted ( $Mdn = 0.18$ ) and words that were not targeted for simplification ( $Mdn = 0.00$ ),  $W = 19.6$ ,  $p < .001$ . A significant difference was also observed between words that were deleted ( $Mdn = 0.090$ ) and words that were not targeted, but the effect was smaller,  $W = 5.30$ ,  $p < .001$ . We therefore conclude that all words identified for simplification were

<sup>9</sup>From Figure 3, it can be seen that the distribution of error probabilities violates the normality assumption, which is confirmed by a Shapiro-Wilk test,  $W = 0.67$ ,  $p < .001$ . We will therefore use non-parametric tests for all statistical analyses.

Group	N	Mdn	DSCF Comparisons (W)	
			substituted	deleted
substituted	192	0.18		
deleted	80	0.09	6.16 ***	
not targeted	1972	0.00	19.6 ***	5.30 ***

\*\*\*  $p < .001$

Table 7: Post-hoc comparisons of the probability of making reading errors on words in authentic texts that were targeted or not targeted for simplification

well-targeted. However, the need for simplification was not equally strong, as can be evidenced by the significant difference between the number of misreadings on words that were substituted and deleted,  $W = 6.16$ ,  $p < .001$ . While removing superfluous words was also necessary to reduce reading difficulties, substituting core but difficult content words seemed even more crucial. In our final analysis, we will therefore focus on these substituted words and have a look at the effect of the chosen simplifications.

#### 4.3.2. Misreadings on lexical substitutions

To confirm our second hypothesis, we focus on the words that were simplified by means of substitution and compare the expected decrease in number of errors made before and after simplification. A pairwise Friedman rank sum test showed that there were significantly fewer misreadings after simplification ( $Mdn = 0.090$ ) than on the word in the original text ( $Mdn = 0.18$ ),  $X^2_F(1) = 40.6$ ,  $p < .001$ . Moreover, the effect on decreasing the probability of misreading the word in the simplified version was large (Kendall's  $W = .527$ ). Consequently, we find that the words targeted for simplification were not only well-chosen, but also substituted with substantially easier alternatives, hence enhancing the readability of the texts for our targeted readers (i.e., dyslexic children).

## 5. Conclusion

In this paper, we presented the Alector corpus, a resource of parallel texts for French learners struggling with reading. The corpus is made of 79 original literary and scientific

texts that have been manually simplified at different linguistic levels. The corpus also contains a part of texts that have been annotated with reading errors by children with dyslexia. The data are available online and offer different criteria for searching and visualizing the texts. The entire corpus is available on demand for research purposes.

In the future, we plan to add the lexical resource ReSyf (Billami et al., 2018) to the corpus in order to highlight lexical difficulty and provide the reader with a list of graded synonyms, as advised by Rello (2014) for dyslexic learners. Moreover, we will continue to use the corpus in reading tests to extend our analysis of reading errors. Finally, the Alector corpus will be used as a gold standard for evaluating current developments in automatic text simplification for French.

### Acknowledgements

This research was funded by the French *Agence Nationale pour la Recherche* (ANR), through the Alector project (ANR-16-CE28-0005). It also received the financial support of the Belgian National Fund for Scientific Research (F.R.S.-FNRS). We deeply thank Solange Lâm (Université de Bruxelles) and Carlos Ramisch (Aix Marseille Université, Laboratoire d'Informatique Fondamentale) for the current version of the interface. We also thank Aurore Brunel, Mathilde Combes, Marie Nandiegou and Stella Reboul, speech therapists that actively participated in the manual simplifications of parts of the texts and on the reading and comprehension tests with dyslexic children. We finally thank three anonymous reviewers for their thoughtful suggestions for improving the initial version of the paper.

### References

#### Bibliographical References

- Billami, M. B., François, T., and Gala, N. (2018). ReSyf: a French lexicon with ranked synonyms. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 2570–2581.
- Bingel, J., Barrett, M., and Klerke, S. (2018). Predicting misreadings from gaze in children with reading difficulties. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 24–34, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Coster, W. and Kauchak, D. (2011). Simple English Wikipedia: A New Text Simplification task. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA.
- François, T., Billami, M. B., Gala, N., and Bernhard, D. (2016). Bleu, contusion, ecchymose : tri automatique de synonymes en fonction de leur difficulté. In *Actes de la conférence Traitement Automatique des Langues Naturelles*, pages 15–28.
- Gala, N. and Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CLALC), COLING 2014*, pages 59–66.
- Gala, N., François, T., Javourey-Drevet, L., and Ziegler, J.-C. (2018). La simplification de textes, une aide à l'apprentissage de la lecture. *Langue Française «L'apprentissage de la lecture en français langue maternelle et seconde»*, xx:123–131.
- Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). Manulex: a grade level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments and Computers*, 36:156–166.
- Mullis, I. V. S., Martin, M. O., Foy, P., and Hooper, M. (2017). PIRLS 2016 International Results in Reading.
- Nandiegou, M. and Reboul, S. (2018). La simplification lexicale comme outil pour faciliter la lecture des enfants dyslexiques et faibles lecteurs. Master's thesis, Mémoire en vue de l'obtention du Certificat de capacité en Orthophonie, Aix Marseille Université, Marseille, France.
- Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : Lexique<sup>TM</sup>//a lexical database for contemporary French : Lexique. *L'Année psychologique*, 101:447–462.
- Paetzold, G. and Specia, L. (2016). Semeval 2016 task 11: Complex word identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 560–569.
- Pedler, J. (2007). *Computer Correction of Real-word Spelling Errors in Dyslexic Text*. Ph.D. thesis, Birkbeck College, London University.
- Rauschenberger, M., Baeza-Yates, R., and Rello, L. (2019). Technologies for Dyslexia. In Yesilada Y. et al., editors, *Web Accessibility. Human-Computer Interaction Series*. Springer, London.
- Rello, L., Baeza-Yates, R., and Llisterri, J. (2014). Dyslist: An annotated resource of dyslexic errors. In *Conference on Language Resources and Evaluation LREC-2014*, page 1289–1296, Reykjavick, Island.
- Rello, L., Ballesteros, M., Abdullah, A., Serra, M., Sánchez, D., and Bigham, J. (2016). Dytective: diagnosing risk of dyslexia with a game. In *Pervasive health*, Cancun. ACM Press.
- Rello, L. (2014). *DysWebxia : a text accessibility model for people with dyslexia*. Ph.D. thesis, Universitat Pompeu Fabra, Barcelona.
- Saggion, H. (2017). *Automatic text simplification*, volume 10. Morgan & Claypool Publishers.
- Štajner, S. and Nisioi, S. (2018). A detailed evaluation of neural sequence-to-sequence models for in-domain and cross-domain text simplification. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- Štajner, S., Yaneva, V., Mitkov, R., and Ponzetto, S. P. (2017). Effects of Lexical Properties on Viewing Time per Word in Autistic and Neurotypical Readers. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 271–281,

- Copenhagen, Denmark. Association for Computational Linguistics.
- Stanovich, K. E., Nathan, R. G., and Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading research quarterly*, x(x):267–283.
- Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of education*, 189(1-2):23–25.
- Stenner, A. (1996). Measuring reading comprehension with the lexile framework. In *Fourth North American Conference on Adolescent/Adult Literacy*.
- Tack, A., François, T., Ligozat, A.-L., and Fairon, C. (2016). Evaluating Lexical Simplification and Vocabulary Knowledge for Learners of French: Possibilities of Using the FLELex Resource. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 230–236, Portorož, Slovenia, May. European Language Resources Association.
- Tunmer, W. E. and Hoover, W. A. (2019). The cognitive foundations of learning to read: a framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties*, 24(1):75–93.
- Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.
- Yaneva, V., Temnikova, I., and Mitkov, R. (2016). A Corpus of Text Data and Gaze Fixations from Autistic and Non-Autistic Adults. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, May. European Language Resources Association (ELRA).
- Yimam, S. M., Biemann, C., Malmasi, S., Paetzold, G., Specia, L., Štajner, S., Tack, A., and Zampieri, M. (2018). A Report on the Complex Word Identification Shared Task 2018. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 66–78, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Zhu, Z., Bernhard, D., and Gurevych, I. (2010). A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1353–1361. Association for Computational Linguistics.
- Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., and Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. 109(28):11455–11459.

### Language Resource References

- Newsela. (2016). *Newsela Article Corpus 2016-01-29*.
- Vital-Durand, F. (2011). *International Reading Speed Texts IResT (French version)*.

## Annexe B

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



## RECOMMANDATIONS pour des transformations de textes en français afin d'améliorer leur lisibilité et leur compréhension

*Les adaptations de texte proposées dans ce document s'adressent à des enfants dyslexiques et faibles lecteurs, mais elles peuvent faciliter la lecture à d'autres publics en difficulté*

N. GALA<sup>1</sup>, A. TODIRASCU<sup>2</sup>, L. JAVOUREY-DREVET<sup>3,4</sup>, D. BERNHARD<sup>2</sup>, R. WILKENS<sup>2,5</sup>, J.-P. MEYER<sup>2</sup>

<sup>1</sup>Aix Marseille Univ. et Laboratoire Parole et Langage (LPL UMR 7309)

<sup>2</sup>Université de Strasbourg et équipe Linguistique Langues Parole (LiLPa)

<sup>3</sup>Aix Marseille Univ. et Laboratoire de Psychologie Cognitive (LPC UMR 7290)

<sup>4</sup>Aix Marseille Univ. et UR 4671 ADEF Apprentissage, Didactique, Évaluation

<sup>5</sup>University of Essex

§

### TABLE DES MATIÈRES

1	Introduction.....	2
2	Typographie.....	4
2.1	R01.T1. Famille de police.....	4
2.2	R02.T2. Taille de police.....	5
2.3	R03.T3. Espacements entre caractères.....	5
2.4	R04.T4. Espacement interligne.....	5
2.5	R05.T5. Espacement entre colonnes.....	6
2.6	R06.T6. Expressions numériques.....	6
3	Lexique.....	7
3.1	R07.L1. Suppression d'adverbes, adjectifs lorsqu'ils sont peu porteurs d'information.....	7
3.2	R08.L2. Relations graphème-phonème régulières/consistantes.....	7
3.3	R09.L3. Syllabes avec une structure CV ou V.....	8
3.4	R10.L4. Unités lexicales plus courtes et plus fréquentes.....	9
3.5	R11.L5. Explication des sigles et maintien des acronymes.....	9
4	Morphologie.....	11
4.1	R12. M1. Morphèmes diminutifs.....	11
4.2	R13. M2. Fréquence dans une famille morphologique.....	11
4.3	R14. M3. Verbes dans des formes personnelles.....	12
4.4	R15. M4. Verbes dans du temps présent.....	12
4.5	R16. M5. Formes composées ou modales.....	14



5	Syntaxe .....	15
5.1	R17. S1. Ordre SVO .....	15
5.2	R18. S2. Constructions clivées et pseudo-clivées.....	16
5.3	R19. S3. Propositions subordonnées relatives .....	17
5.4	R20. S4. Conjonctions de coordination, de juxtaposition .....	17
5.5	R21. S5. Propositions négatives .....	18
5.6	R22. S6. Constructions passives .....	18
5.7	R23. S7. Constructions attributives .....	19
5.8	R24. S8. Discours rapporté .....	19
5.9	R25. S9. Incises et parenthèses .....	20
6	Discours (cohésion textuelle) .....	21
6.1	R26. D1. Pronoms et antécédents.....	21
6.2	R27. D2. Changement de déterminant.....	22
6.3	R28. D3. Répétition du référent .....	22
6.4	R29. D4. Suppression des expressions référentielles (pronoms, déterminants possessifs) .	23
6.5	R30. D5. Accessibilité des expressions référentielles.....	23
7	Corpus.....	25
7.1	Textes édités en version simplifiée/adaptée .....	25
7.2	Textes manuellement simplifiés.....	26
8	Références bibliographiques.....	29

## §

## 1 Introduction

La présentation et le contenu des textes ont un impact sur leur lisibilité et leur compréhension. Dans le cadre des travaux du projet ANR ALECTOR (16-CE28-0005)<sup>1</sup> nous avons mené plusieurs tests de lecture et analysé différents types de corpus afin de dégager des recommandations pour la transformation des textes. Ces transformations (ou simplifications) permettent de créer des textes équivalents aux textes originaux mais plus lisibles et compréhensibles par des enfants en difficulté de lecture.

À ce jour, en l'état de la technologie actuelle, certaines transformations proposées sont réalisables uniquement de façon manuelle. Dans certains cas, principalement en syntaxe, ces transformations peuvent être implémentées pour obtenir des équivalents de façon (semi-)automatique.

<sup>1</sup> <https://alectorsite.wordpress.com/>

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Dans ce rapport, nous présentons les résultats de nos recommandations. Différentes publications scientifiques permettent d'éclairer ou d'approfondir certains aspects (cf. *Références bibliographiques* de ce rapport ou page *Publications* du site web du projet).

À ce jour, il existe différentes propositions de règles permettant d'adapter les textes pour faciliter leur lecture et leur compréhension. C'est le cas, notamment d'associations en lien avec la dyslexie ou les déficiences cognitives (par exemple, les règles européennes pour une information facile à lire et à comprendre, proposées par l'UNAPEI<sup>2</sup>). Il existe également des propositions en didactique de la lecture, par exemple pour son apprentissage : Proposition de progression pédagogique à travers les difficultés de lecture du français, par L. Sprenger-Charolles et collaborateurs (Dehaene, 2011).

Par ailleurs, le domaine du traitement automatique des langues, qui s'intéresse à la simplification de textes depuis quelques années, compte également avec des publications qui décrivent des **opérations<sup>3</sup> de simplification** (notamment, Siddharthan, 2014 et Saggion, 2017). Néanmoins, les travaux décrits ont été menés principalement pour l'anglais : une typologie détaillée et couvrant différents aspects pour le français n'existe pas encore à notre connaissance (par exemple, Max, 2006, Brouwers et al., 2012 ou Gala et al., 2020b proposent des règles de transformation uniquement pour la syntaxe).

Proposer une telle typologie, détaillée et couvrante, était l'un des objectifs du projet ALECTOR (2017-2020). Pour ce faire, nous avons analysé des corpus en français adressés à des enfants en âge scolaire et à des adultes en difficulté de lecture. Nous avons, ainsi :

- constitué le corpus ALECTOR<sup>4</sup> (Gala, et al. 2020a), un ensemble de textes utilisés du cours élémentaire 1 (CE1) au cours moyen 1 (CM1). Des tests avec différentes cohortes d'enfants ont été menés pour analyser les erreurs de lecture commises dans le cadre de plusieurs dispositifs : dans des cabinets d'orthophonie avec une lecture à voix haute (Brunel et Combes, 2015 ; Nandiegou et Reboul, 2018) et dans des conditions habituelles de classe en milieu scolaire avec une lecture silencieuse (Javourey-Drevet, 2021) ;
- analysé un ensemble de textes publiés destinés à des enfants dyslexiques (corpus CONTES, collection de textes proposés par l'association *Methododys*<sup>5</sup> pour aider les enfants dyslexiques) et aussi à des adultes illettrés (collection *La Traversée*), cf. section 7, Corpus analysés.

À la suite de ces analyses, nous proposons de classer les opérations de simplification en trois types génériques, à savoir : (1) **suppression**, (2) **remplacement** (substitution) et (3) **ajout** d'information<sup>6</sup>.

Pour chaque type, les transformations peuvent opérer sur quatre **niveaux linguistiques** : **lexique, morphologie, syntaxe, discours** (Gala et al. 2018). Certaines transformations au **niveau du document** sont aussi prises en considération au préalable. Elles concernent la **typographie** et les aspects formels et visuels du texte.

<sup>2</sup> <https://www.unapei.org/wp-content/uploads/2018/11/L%E2%80%99information-pour-tous-Re%CC%80gles-europe%CC%81ennes-pour-une-information-facile-a%CC%80lire-et-a%CC%80comprendre.pdf>

<sup>3</sup> « (...) to identify expected simplification operations such as sentence split, sentence deletion, and various types of change operations (syntactic, lexical, etc.). Moreover, additional operations such as insertion and reordering were also documented. » (Saggion 2017, p. 3).

<sup>4</sup> Le corpus est accessible ici : <http://corpuselector.huma-num.fr/faces/index.xhtml>

<sup>5</sup> <https://methododys.ch/>

<sup>6</sup> Dans la suite du rapport, lorsqu'il y aura un exemple de **suppression** ou de **remplacement**, l'unité apparaîtra respectivement en rouge et en bleu.



Dans la suite de ce rapport, un tableau récapitulatif au début de chaque section présente une vue synthétique des transformations proposées, suivi d'une explication détaillée et de quelques exemples obtenus à partir des corpus utilisés pour notre étude.

## 2 Typographie

		Type de transformation	Intitulé	Préconisation
1	T1		Famille de police	<i>Open Dyslexic</i> ou toute autre police qui permet une meilleure visualisation du texte
2	T2		Taille de police	12pt à 14pt à privilégier
3	T3		Espacement entre caractères	
4	T4		Espacement interligne	
5	T5		Espacement entre colonnes	
6	T6	Remplacement	Expressions numériques	Formes numériques à privilégier (au détriment de la forme -plus longue- développée)

### 2.1 R01.T1. Famille de police

Plusieurs travaux dans la littérature indiquent que le type de police a un impact dans la lisibilité (Rello *et al.* 2013 ; Rello 2014, chapitre 9 *Good fonts for Dyslexia*). La police en italique ou Sans Serif devraient être évitées, des familles comme Arial, Calibri, Courier, Helvetica, Roman ou Verdana semblent augmenter la lisibilité des textes (Rauschenberger *et al.* 2019). Il existe, par ailleurs, une police conçue spécialement pour faciliter la lecture chez les dyslexiques (*Open Dyslexic*<sup>7</sup>). Elle se caractérise, entre autres, par une base plus lourde, une ouverture élargie, des bâtons plus longs et un espacement plus grand entre les lettres (mais il n'est pas prouvé qu'elle ait un impact supérieur à d'autres familles de polices comme celles citées précédemment).

Dans le cadre du projet ALECTOR, nous avons fait le choix d'implémenter cette police dans la base de corpus parallèles (cf. note en bas de page numéro 3) :

La cloche de 4 heures et demie vient de sonner. Mme Morot :  
« C'est fini pour aujourd'hui, dit-elle, nous continuerons demain.  
Un murmure de contestation s'élève dans la classe et une fille dit :  
« S'il vous plaît madame ! Finissez les aventures de Robert !  
- Non, ce serait trop long, Émilie. J'ai dit demain. »  
Émilie râle un peu en rangeant ses affaires. L'air grognon, elle dit :  
« Comme le sujet t'intéresse à ce point, Émilie, c'est toi qui n'as rien dit.  
- D'accord ! »  
Émilie court sans se retourner, son cartable balançant sur sa tête.  
C'est un crayon. Elle se baisse, le ramasse. Arrivée chez elle, Émilie va dans sa chambre. La petite fille crie :  
« Oh j'ai une idée ! Je vais dessiner un énorme goûter ! Une brioche !  
Émilie dessine alors une grosse brioche. Puis elle ajoute un peu de sucre.  
Elle pose le crayon sur son bureau. Au bout de quelques secondes, elle a les  
yeux grands ouverts, Émilie se trouve installée devant un bonbon.

Figure 1. Police Open Dyslexic.

<sup>7</sup> <https://www.dafont.com/fr/open-dyslexic.font>



## 2.2 R02.T2. Taille de police

La taille de police préconisée est de 12 à 14 pt (British Dyslexia Association<sup>8</sup> 2018 ; Rello 2014, chapitre 7 *Colors, Sizes and Spacing*), même si une police supérieure peut être nécessaire dans certains cas.

Idéalement, si le document est consultable sur écran, il doit être possible d'augmenter la taille de la police, l'interligne et l'espacement entre caractères. C'est le cas de l'interface de lecture de textes du projet ALECTOR.



**Figure 2.** Ajustements de police, espacement interligne et entre caractères dans ALECTOR.

Par ailleurs, il faudrait éviter les variations de taille, c'est-à-dire limiter les majuscules et les petites majuscules autant que possible (Dyslexia Association of Ireland<sup>9</sup>).

## 2.3 R03.T3. Espacements entre caractères

D'après une étude de Zorzi *et al.* (2014) une augmentation entre l'espace des lettres améliore la lisibilité des textes :

« Extra-large letter spacing helps reading, because dyslexics are abnormally affected by crowding, a perceptual phenomenon with detrimental effects on letter recognition that is modulated by the spacing between letters. Extra-large letter spacing may help to break the vicious circle by rendering the reading material more easily accessible. »

La British Dyslexia Association dans ses guidelines de 2018 préconise un espacement idéal de 35% de la moyenne de la largeur des lettres :

« Larger inter-letter / character spacing (sometimes called tracking) improves readability, ideally around 35% of the average letter width. »

Cependant, un espacement trop important peut nuire à la lisibilité.

## 2.4 R04.T4. Espacement interligne

Un interligne supérieur à l'habituel améliore la lisibilité. Il doit être proportionnel à l'espacement entre mots. La British Dyslexia Association (2018) recommande 1.5/150 %.

<sup>8</sup> <https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide>

<sup>9</sup> <https://www.dyslexia.ie/information/computers-and-technology/making-information-accessible-dyslexia-friendly-style-guide/>

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



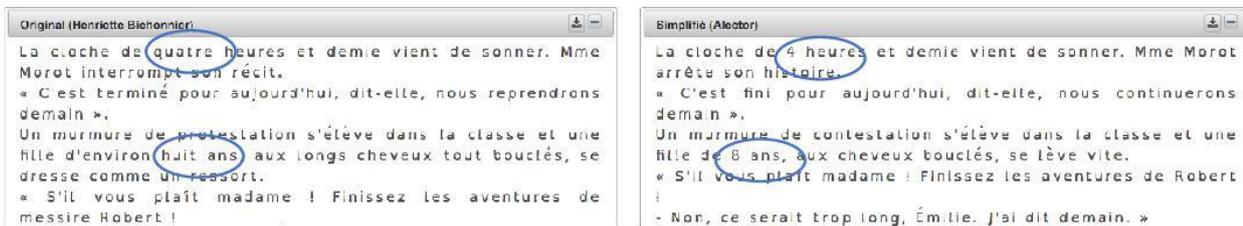
## 2.5 R05.T5. Espacement entre colonnes

Idéalement il faudrait éviter les documents multicolonnés. S'il y en a, elles devraient avoir entre 44 et 66 caractères maximum (Rello, 2014), la British Dyslexia Association recommande 60 à 70 caractères.

## 2.6 R06.T6. Expressions numériques

Différents travaux en psychophysique et en neuropsychologie ont étudié l'impact du contexte et les temps de fixation oculaires au moment de la lecture (Rello 2014, chapitre 11 *Numerical expressions*). Il en ressort que lorsque les expressions numériques apparaissent en chiffres dans le texte, l'accès à leur représentation sémantique est plus rapide que lorsqu'ils apparaissent en lettres.

Le traitement visuel des mots est affecté par des facteurs comme la longueur, la fréquence ou l'ambiguïté. Les expressions numériques étant plus courtes, les temps de lecture sont plus réduits si l'information est présentée en chiffres et non en alphabétiquement (Rello, 2014). La charge cognitive (de déchiffrement) est ainsi réduite.



**Figure 3.** Corpus comparables original et simplifié dans l'interface de corpus ALECTOR, avec remplacement d'expressions numériques en chiffres.

L'utilisation de chiffres au détriment des lettres (lors des expressions numériques simples) est donc à privilégier dans les adaptations de textes (cf. Figure 3).



### 3 Lexique

		Type de transformation	Intitulé	Préconisation
7	L1	Suppression	Adverbes, Adjectifs	Supprimer Adverbes et Adjectifs lorsqu'ils sont peu porteurs d'information
8	L2	Remplacement	Consistance graphème – phonème	Graphèmes avec des relations régulières à privilégier au détriment d'unités lexicales avec des syllabes contenant des relations irrégulières
9	L3	Remplacement	Structure syllabique	Unités lexicales avec des structures simples (CV, V) à privilégier au détriment d'unités lexicales avec des structures plus complexes
10	L4	Remplacement	Fréquence et longueur	Unités lexicales fréquentes et de moins de sept caractères à privilégier, au détriment d'unités lexicales longues et moins fréquentes
11	L5	Ajout	Sigles	Rajouter une explication (si besoin, p. ex. textes documentaires/scientifiques)

#### 3.1 R07.L1. Suppression d'adverbes, adjectifs lorsqu'ils sont peu porteurs d'information

Dans les transformations manuelles de textes pour enfants, nous avons fait le choix de supprimer certains adjectifs et adverbes *peu porteurs d'information*. Les phrases dans les versions simplifiées sont ainsi plus courtes et sont lues avec moins d'erreurs (Brunel et Combes, 2015 ; Nandiegou et Reboul, 2018).

Original. L'âne continuait à examiner la peinture d'un regard **plutôt** **attristé**.

Simplifié. L'âne continuait à regarder la peinture d'un regard **triste**.

**Exemple 1.** Corpus 101\_LIT CE1 Boîtes peinture.

Original. (...) une fille (...) aux **longs** cheveux **tout** bouclés.

Simplifié. (...) une fille (...) aux cheveux bouclés.

**Exemple 2.** Corpus 102\_LIT CE1 Émilie et le crayon magique.

D'un point de vue d'un traitement automatique, il pourra être envisageable de supprimer les adjectifs épithètes et les adverbes précédant les adjectifs (exemple 2).

#### 3.2 R08.L2. Relations graphème-phonème régulières/consistantes

« Le degré de consistance se calcule généralement par le rapport entre le nombre de mots contenant une unité orthographique de prononciation spécifique et le nombre total de mots contenant cette même unité orthographique sans distinction de prononciation (Peereman & Content, 1999). Il s'agit ici de la consistance orthographique prise dans le sens de la lecture, c'est-à-dire de celle qui a pour point de départ le code orthographique et pour point d'arrivée le code phonologique, plus précisément appelée consistance ortho-phonologique » (Pérez 2014 sur Catach et al. (1995)).

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Pour la simplification de textes, il est conseillé de substituer les unités lexicales présentant des syllabes avec consistance irrégulière par des unités lexicales synonymes avec des syllabes avec consistance régulière. Dans tous les cas, il faudra privilégier les unités lexicales avec des graphèmes qui ont une consistance régulière :

Difficulté	Caractéristiques, type de graphème	Graphème	Commentaire
1	Simple, acontextuels, consistants	a, i, u, p, b, t, d, f, v, l, r, m, n	Régularité maximale
2	Complexes, acontextuels, consistants	eu, ou, on, un, oi, oin, ch, gn	
3	Simple, acontextuels, inconsistants	é, è, o, ï, y	Régularité moyenne
4	Complexes, acontextuels, inconsistants	ai, au, eau, an, en, in, il(l), qu	
5	Simple, contextuels, inconsistants	e, g, s, c, ç, z, j, x	Irrégularité maximale
6	Complexes, contextuels, inconsistants	qu, ss, ge	

**Figure 4.** Typologie de graphèmes (Pérez 2014 sur Catach et al. (1995)).

Original. (...) se dresse comme un ressort.

Simplifié. (...) se lève vite.

**Exemple 3.** Corpus 102\_LIT CE1 Émilie et le crayon magique.

Original. Quel que soit le temps, il rapportait toujours au village quantité de beaux et rares poissons.

Simplifié. Par tous les temps, il rapportait au village beaucoup de poissons.

**Exemple 4.** Corpus 176\_LIT CE2 Cinq frères.

### 3.3 R09.L3. Syllabes avec une structure CV ou V

On privilégiera les syllabes simples et fréquentes (consonne-voyelle CV ou voyelle seule V) au détriment de structures plus complexes (CVCC, CCVC, CYVC, CYV, CCV, VCC, CVY, VC, YV, VCCC, CCYC, CCVC).

Une structure complexe est par définition plus longue (par exemple, double consonne suivie de voyelle). Les temps de lecture sont inférieurs et les erreurs moindres dans des phrases ayant des unités lexicales qui comportent des syllabes simples et fréquentes :

Difficulté	Caractéristiques, type de syllabe	Syllabe
1	Simple, fréquente	CV, V
2	Complexe, fréquente	CVC
3	Complexe, moins fréquente	CVCC, CCVC, CYVC, CYV, CCV, VCC, CVY
4	Complexe, rare	VC, YV, VCCC, CCYC, CCVC

**Figure 5.** Typologie de syllabes.

Original. Je suis sûr que tu peux gagner ce concours. (CV-CVC)

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Simplifié. Je suis sûre que tu peux gagner ce défi. (CV-CV)

**Exemple 5.** Corpus 177\_LIT CE2 Les lunes à croquer.

Original. (...) cinq frères chinois qui se ressemblaient comme cinq gouttes d'eau. (CV-CV-CCV)

Simplifié. (...) 5 frères chinois. Ils étaient comme 5 gouttes d'eau. (V-CV)

**Exemple 6.** Corpus 176\_LIT CE2 Cinq frères.

### 3.4 R10.L4. Unités lexicales plus courtes et plus fréquentes

On privilégiera les unités lexicales courtes (avec moins de sept caractères). La longueur étant corrélée à la fréquence les synonymes plus courts et plus fréquents seront préférés (Gala et Ziegler, 2016).

Cela impliquera également de privilégier les graphèmes simples (cf. L2 et L3) et d'éviter les unités lexicales avec des consonnes doubles (ph, ff, ll, pp).

Original. Ses poumons volumineux lui permettent de rester sous l'eau.

Simplifié. Ses gros poumons lui permettent de rester sous l'eau.

**Exemple 7.** Corpus 42\_SCI CE1 Castor.

Original. L'impesanteur n'est pas l'absence de pesanteur mais la combinaison de cette pesanteur avec une certaine vitesse de déplacement.

Simplifié. L'impesanteur n'est pas l'absence de pesanteur. C'est l'union de la pesanteur avec la vitesse de déplacement.

**Exemple 8.** Corpus 84\_SCI CE2 Attraction.

Cependant, dans les corpus scientifiques l'objectif est de faire apprendre du nouveau vocabulaire aux enfants, avec des mots comme *impesanteur*, *atmosphère*, *molécule*, *vitamine*, etc. Même si souvent ces mots sont longs, il n'est pas jugé pertinent de modifier les termes de spécialité de façon systématique.

### 3.5 R11.L5. Explication des sigles et maintien des acronymes

Dans les corpus utilisés pour notre étude (corpus CONTES et corpus ALECTOR, cf. section 7), nous n'avons pas relevé d'abréviations (sigles et/ou acronymes). Ceci est courant dans les matériels pédagogiques pour l'apprentissage de la lecture adressés à des enfants de 6 à 11 ans (textes littéraires et documentaires scientifiques).

Nous avons relevé des acronymes uniquement dans les romans pour adultes dans la collection *La Traversée*, adressée à des adultes faibles lecteurs ou illettrés. Il s'agit d'acronymes fréquents dans la langue (lexicalisés et facilement prononçables comme un mot), par exemple : TV, OK, CD, GSM.

Nous avons regardé dans d'autres corpus simplifiés. Dans le corpus *Newsela* (corpus en anglais, avec quatre niveaux de simplification), les acronymes sont utilisés surtout dans les niveaux où la

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



simplification est la plus importante. Par exemple, NASA apparaît dans les niveaux 3 et 4, au détriment de 'space agency' qui apparaît dans le texte original et dans les deux premiers niveaux de simplification. En revanche, si un sigle peut apparaître dans les niveaux de simplification intermédiaires, toujours antécédé du nom développé, il tend à disparaître dans les niveaux 3 et 4 :

V0 (original) *To help develop an ethical framework for venturing into this unknown territory, the space agency asked the Institute of Medicine to convene a panel of experts to offer some helpful guidelines.*

V1 (simplifié niveau minimal) *The space agency asked the Institute of Medicine (IOM) to offer guidelines to help NASA make such health decisions as it ventures into this unknown territory.*

V2 *The space agency asked the Institute of Medicine (IOM) to help develop a set of guidelines for venturing into this unknown territory.*

V3 *To help NASA make decisions about dealing with the risks, the space agency asked the Institute of Medicine to develop a set of guidelines.*

V4 (simplifié niveau maximal) *To make decisions about the risks, NASA asked a group of doctors to develop guidelines.*

Réf. Newsela DOC561

V0 *The UW researchers also were involved in an earlier experiment, the San Andreas Fault Observatory at Depth (SAFOD), completed in 2007.*

V1 *The UW researchers also were involved in an earlier experiment, the San Andreas Fault Observatory at Depth (SAFOD).*

V2 *The UW researchers also were involved in an earlier experiment, called the San Andreas Fault Observatory at Depth (SAFOD).*

V3 *The university researchers also took part in an earlier experiment at the San Andreas fault in California.*

Réf. Newsela DOC768

Enfin, dans *Vikidia*, encyclopédie francophone pour enfants de 8 à 13 ans, les sigles apparaissent toujours après le nom développé :

« *L'Organisation des Nations unies (ONU)* est une organisation mondiale chargée de faire respecter la paix dans le monde et de calmer les conflits et les guerres. (...) »

L'ONU a reçu le prix Nobel de la paix en 2001. Des organisations dépendant de l'ONU ont également été lauréates du prix. Les casques bleus l'ont reçu en 1988. (...) Le *Haut-commissariat des Nations Unies pour les réfugiés (UNHCR)* l'a reçu deux fois en 1954 et en 1981. »

(Réf. *Vikidia*, entrée ONU).



## 4 Morphologie

		Type de transformation	Intitulé	Préconisation
12	M1	Suppression	Morphèmes diminutifs	Les supprimer pour raccourcir l'unité lexicale sans perte de sens (éventuellement ajouter l'adjectif 'petit')
13	M2	Remplacement	Fréquence dans une famille morphologique	Privilégier les mots les plus fréquents d'une famille morphologique, à substituer sinon
14	M3	Remplacement	Verbes dans des formes personnelles	À privilégier, à substituer sinon
15	M4	Remplacement	Verbes dans des temps du présent	À privilégier, à substituer sinon
16	M5	Remplacement	Formes composées ou modales	À remplacer par des formes simples

### 4.1 R12. M1. Morphèmes diminutifs

Les suffixes diminutifs peuvent être supprimés sans perte majeure de sens. On pourra rajouter (si la nuance sémantique est considérée nécessaire) l'adjectif 'petit'.

**Original.** Une partie est composée d'un nombre incalculable de particules très fines de poussières et de microscopiques gouttelettes d'eau.

**Simplifié.** Une partie est faite d'un nombre très grand de particules très fines de poussières et gouttes d'eau très petites.

**Exemple 9.** Corpus 83\_SCI CE2 Lumière.

**Original.** Elle habitait une maisonnette au fond des bois.

**Simplifié.** Elle habitait une petite maison dans les bois.

**Exemple 10.** Corpus 136\_LIT CM1 Sorcière.

### 4.2 R13. M2. Fréquence dans une famille morphologique

Comme pour les substitutions lexicales, il est préconisé de remplacer une unité lexicale par une autre de la même famille morphologique si celle-ci est plus fréquente.

**Original.** Les grosses branches sont séparées et utilisées comme bois pour la construction de barrages.

**Simplifié.** Il sépare les grosses branches et les utilise comme bois pour construire des barrages.

**Exemple 11.** Corpus 42\_SCI CE1 Castor.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Dans la ressource Lexique 3<sup>10</sup>, la forme 'construction' est renseignée avec une fréquence de 11,3 tandis que 'construire' 26,94 (freqfilms2)<sup>11</sup>. Dans la ressource Manulex<sup>12</sup>, les fréquences sont de 31,31 et 68,73, respectivement.

**Original.** Ces grenouilles (...) profitent de ce moment-là pour leur reproduction.

**Simplifié.** Ces grenouilles (...) profitent de ce moment-là pour se reproduire.

**Exemple 12.** Corpus 43\_SCI CE1 Climat.

Comme pour l'exemple précédent, la forme à l'infinitif est plus fréquente que la forme nominale : dans Lexique 3 (freqfilms2) 4,68 vs 4,08 ; dans Manulex 5,16 vs 3,08. De même pour 'entraînement' et 'entraînent' :

**Original.** Ils doivent aussi subir un entraînement intensif pour supporter les nouvelles conditions de vie.

**Simplifié.** Ils s'entraînent beaucoup pour supporter leurs nouvelles conditions de vie.

**Exemple 13.** Corpus 29\_SCI CE1 Espace.

#### 4.3 R14. M3. Verbes dans des formes personnelles

Lorsque cela est possible, on privilégiera les formes personnelles au détriment des formes non personnelles, principalement infinitifs et participes présents.

**Original.** Le vieil homme est tout heureux et, pour remercier Pierre, il lui donne un pot de lait tout frais.

**Simplifié.** Le vieil homme est heureux. Il remercie Pierre et lui donne un pot de lait frais.

**Exemple 13.** Corpus 177\_LIT CE2 Lunes à croquer.

**Original.** Émilie court sans se retourner, son cartable ballottant sur ses épaules.

**Simplifié.** Émilie court sans se retourner. Son cartable se balance sur ses épaules.

**Exemple 14.** Corpus 102\_LIT CE1 Émilie et le crayon magique.

Dans les deux exemples, les formes personnelles sont plus fréquentes que les formes non personnelles, aussi bien dans Lexique3 (freqfilms2) que dans Manulex : dans Lexique 3 'balance' 9,66 vs 'balançant' 0,71, et dans Manulex 14,22 vs 3,38, respectivement.

#### 4.4 R15. M4. Verbes dans du temps présent

Lors que le texte le permet, il est préconisé d'utiliser les formes verbales au présent (« présent de narration » dans les textes littéraires). Cela permet de rapporter des actions passées en les rendant

<sup>10</sup> <http://www.lexique.org>

<sup>11</sup> La différence dans les corpus littéraires est moindre, 18.24 vs 18.38 respectivement (freqlivres).

<sup>12</sup> <http://www.manulex.org>

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



plus vivantes, en donnant une impression plus directe et proche du lecteur. Les formes au présent sont aussi des formes plus courtes et par conséquent plus familières et fréquentes (par exemple, dans Lexique 3, 'sort' 82,41, 'sortit' 2,31, 'ressort' 2,48, 'ressortit' 0,26).

**Original.** À un moment, le grand-père plongea sa main toute entière sous un amas de cartes postales, fouilla le fond de la boîte, puis en ressortit son poing serré. Il tourna sa paume au ciel et réouvrit doucement ses doigts, dévoilant aux yeux de la petite fille une sorte de pépite couleur or qui scintillait de mille feux. « Ma petite Sophie, ceci est la graine d'un arbre magique (...). »

**Simplifié.** À un moment, le papi met sa main dans un tas de cartes postales, cherche au fond de la boîte et sort sa main fermée. Il ouvre sa main et montre à Sophie une pépite dorée qui brille. « Ma petite Sophie, cette pépite est la graine d'un arbre magique (...). »

**Exemple 15.** Corpus 04\_LIT CE1 Sophie et l'arbre magique.

Par ailleurs, les textes documentaires (scientifiques) utilisent généralement le présent dit « de vérité générale ».

La lumière du Soleil émet toutes les couleurs. L'air terrestre est un filtre. Le ciel de notre planète est bleu.

**Exemple 16.** Corpus 83\_SCI CE2 Lumière.

Les romans de la collection *La Traversée* sont aussi, majoritairement, écrits au présent.

« On lui dit souvent aussi : - Tu as des cheveux magnifiques ! Ils sont roux, presque rouges ! Isabelle ne voit pas les couleurs non plus. Elle les devine. Elle les imagine. Elle devine ce que cela veut dire des cheveux roux. C'est presque rouge. Il y a des cheveux blonds, bruns et noirs. Et puis, des cheveux roux. On lui dit : - Le roux, c'est la couleur des arbres en automne. Alors elle sait que les feuilles des arbres ont la même couleur que ses cheveux. »

**Exemple 17.** Corpus Les chapeaux rouges (La Traversée).

Toutefois, il est important de signaler que dans les textes littéraires pour enfants (contes, histoires, comptines) l'imparfait et le passé simple sont majoritaires et font partie du style de ce type de textes (narration au passé). Par conséquent, il est difficile de les modifier et nous n'avons pas jugé pertinent de transformer ces formes au présent de façon systématique. Seuls les changements lexicaux ont été appliqués le cas échéant.

Un jour, un jeune homme passa devant sa fenêtre. Il était beau.

**Exemple 18.** Corpus 136\_LIT CM1 La sorcière amoureuse.

**Original.** Juché sur la cime d'un arbre mort, le corbeau contemplait pensivement (...) Il songeait (...) Dans un instant, il bouleverserait définitivement l'existence de l'un d'entre eux.

**Simplifié.** Posé sur la cime d'un arbre mort, le corbeau regardait (...) Il pensait (...) Dans un instant, il changerait pour toujours l'existence de l'un d'entre eux.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



**Exemple 19.** Corpus 143\_LIT CM1 La dragonne de minuit.

#### 4.5 R16. M5. Formes composées ou modales

On privilégiera les syntagmes verbaux simples (un seul verbe ou un verbe avec l’auxiliaire selon les temps). Les constructions modales ou les locutions verbales seront à éviter (si le sens le permet).

**Original.** Mesurer un volume, c'est être capable de répondre à cette question : « Quelle quantité de liquide contient cette bouteille ? »

**Simplifié.** Mesurer un volume, c'est pouvoir répondre à la question : « Quelle quantité de liquide contient cette bouteille ? »

**Exemple 20.** Corpus 28\_SCI CE1 Mesure.

**Original.** La température est très basse à l'ombre (...), elle peut devenir extrêmement élevée.

**Simplifié.** La température est très basse à l'ombre (...). Elle devient très élevée.

**Exemple 21.** Corpus 29\_SCI CE1 Mesure.

**Original.** Un éléphant vint à passer.

**Simplifié.** Un éléphant passa.

**Exemple 22.** Corpus 178\_LIT CM1 L'éléphant et l'oiseau.



## 5 Syntaxe

		Type de transformation	Intitulé	Préconisation
17	S1	Suppression	Maintien de l'ordre SVO, suppression de participes passés ou présent en position initiale	Cela peut entraîner une réorganisation de la phrase, parfois avec une suppression (ex. L1)
18	S2	Suppression	Maintien de l'ordre SVO, suppression des constructions clivées et pseudo-clivées	Privilégier les suppressions des clivées et pseudo-clivées
19	S3	Suppression	Propositions subordonnées relatives, appositions, modifieurs nominaux, modifieurs verbaux	Suppression de structures qui apportent des informations non essentielles à la compréhension
20	S4	Suppression, remplacement	Conjonctions de coordination, de subordination, de juxtaposition	Suppression des conjonctions à privilégier, découpage en plusieurs phrases (ce qui entraîne une réduction de la longueur des phrases)
21	S5	Remplacement	Propositions négatives	Privilégier les propositions positives (et transformer les propositions négatives en positives)
22	S6	Remplacement	Constructions passives	Privilégier la transformation en phrases actives
23	S7	Remplacement	Constructions attributives	Les attributives sont transformées en groupes nominaux ou adjectivaux dans les énumérations.
24	S8	Suppression	Discours rapporté (propositions complétives) et interrogatives indirectes	Privilégier le discours direct (mais le discours rapporté peut aussi être présent)
25	S9	Suppression ou remplacement	Incises et parenthèses	À supprimer ou à substituer.

### 5.1 R17. S1. Ordre SVO

Dans les cas où il est possible de le faire, il est préférable de supprimer les modifieurs de phrase initiaux, de façon à préserver l'ordre SVO, avec le syntagme nominal sujet en première place. Les romans de la collection *La Traversée* contiennent majoritairement des phrases avec cette structure, par exemple le fragment de l'exemple 17 plus haut, et aussi l'exemple suivant (où seulement une phrase commence avec des modifieurs de phrase « un jour, en Amérique ») :

Il porte un pull bleu, un veston gris et une casquette. François lève sa main pour lui faire un grand signe et il dit à Sarah : - C'est Simon, Maman. Il est devenu mon meilleur ami. Je suis sûr que tu l'aimeras beaucoup. Un jour, en Amérique, il m'a sauvé la vie. Sarah regarde Simon qui descend l'escalier à son tour. Simon a environ trente ans. Il n'est pas très beau, il a un visage étrange.

**Exemple 23.** Corpus Sans dire un mot (*La Traversée*).



Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020

Déplacement ou suppression d'adverbes :

Original. **Jamais** elle ne rate son but, même quand elle lance le torchon à travers la cuisine.

Simplifié. Elle ne rate **jamais** son but, même quand elle lance le torchon à travers la cuisine.

**Exemple 24.** Corpus 141\_LIT CM1 Chichois.

Suppression de participes passés et leurs arguments :

Original. **Pris de panique**, chacun des trois sots se cache rapidement.

Simplifié. Chacun des trois sots se cache rapidement.

**Exemple 25.** Corpus DYS (PC56\_Les\_trois\_sots\_orig ; PC57\_Les\_trois\_sots\_dys).

Suppression de participes présents :

Original. **Se sentant rejetés**, ils étaient toujours ensemble.

Simplifié. Ils étaient toujours ensemble.

**Exemple 26.** Corpus DYS (PC56\_Les\_trois\_sots\_orig ; PC57\_Les\_trois\_sots\_dys).

## 5.2 R18. S2. Constructions clivées et pseudo-clivées

Les constructions clivées et pseudo-clivées, fréquentes dans les textes documentaires, permettent de porter un focus sur un élément de la phrase. Comme elles impliquent une complexification avec le rajout de la proposition subordonnée, il est préconisé de les modifier en une proposition simple avec l'ordre SVO.

Original. C'est le vent qui apporte la pluie.

Simplifié. Le vent apporte la pluie.

**Exemple 27.** Corpus 25\_SCI CE1 Vent.

Original. Ce sont les océans qui sont les plus sensibles à cette attraction.

Simplifié. Les océans sont les plus sensibles à cette attraction.

**Exemple 28.** Corpus 89\_SCI CE2 Marées.

Original. C'est à ce moment que le chocolat prend sa finesse, son arôme.

Simplifié. À ce moment-là le chocolat prend sa finesse, son arôme.

**Exemple 29.** Corpus 158\_SCI CM1 Chocolat.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



**Original.** C'est en comparant des squelettes d'oiseaux à ceux d'autres théropodes exposés dans les musées que les scientifiques ont observé de nombreux points communs et ont compris qu'il existait un lien de parenté.

**Simplifié.** Les scientifiques ont comparé des squelettes d'oiseaux à ceux d'autres théropodes exposés dans les musées. Ils ont observé de nombreux points communs et ont compris qu'il y avait un lien de parenté.

**Exemple 30.** Corpus 87\_SCI CE2 Dinosaures.

### 5.3 R19. S3. Propositions subordonnées relatives

La suppression des phrases relatives (lorsque cela est possible) implique une réduction de la longueur de la phrase principale. La phrase subordonnée relative devient alors une deuxième proposition indépendante (on obtient alors deux phrases indépendantes).

**Original.** L'homme qui a acheté la vache lui a donné un sachet de graines.

**Simplifié.** L'homme a acheté la vache. Il a donné à Jacques un sachet de graines.

**Exemple 31.** Corpus DYS (PC25 Jacques\_haricot\_magique\_dys ; PC26 Jacques\_haricot\_magique\_orig)

Certaines propositions relatives peuvent être transformées en modifieur nominal :

**Original :** Son mari n'a jamais osé lui dire que sa soupe avait une odeur qui le gênait.

**Simplifié.** Monsieur Dupond ose, pour la première fois, dire à sa femme qu'il n'aime pas l'odeur de sa soupe.

**Exemple 32.** Corpus DYS (PC62\_Soupe\_de\_la\_discorde\_orig ; PC63\_Soupe\_de\_la\_discorde\_dys).

Dans certains cas, la phrase subordonnée peut être modifiée en un participe passé (la subordonnée relative devient alors un modifieur nominal) :

**Original.** Il prit son cartable brun qui traînait sous le lit.

**Simplifié.** Il prit son cartable brun jeté sous le lit.

**Exemple 33.** Corpus 140\_LIT CM1 Les cartables.

### 5.4 R20. S4. Conjonctions de coordination, de juxtaposition

Dans les cas de listes et énumérations, il est préconisé de rajouter une conjonction finale si elle est manquante.

**Original.** De tout temps, l'homme a utilisé son corps pour mesurer : la main, l'œil donnent des informations.

**Simplifié.** L'homme a toujours utilisé son corps pour mesurer. La main et l'œil donnent des informations.

**Exemple 34.** Corpus 28\_SCI Mesure.



### 5.5 R21. S5. Propositions négatives

La compréhension d'une phrase négative est plus complexe que celle d'une phrase affirmative, ce qui implique que le traitement des phrases affirmatives est plus rapide que le traitement des phrases négatives. Nous préconisons de transformer, lorsque c'est possible, une phrase négative en affirmative afin d'alléger la charge cognitive et diminuer, ainsi, les problèmes d'interprétation qui se posent dans certains cas.

**Original.** On croyait auparavant que les araignées étaient protégées de la colle de leurs propres toiles, alors que les mouches et les autres insectes n'ont pas cette protection. *Mais ce n'est pas vrai.*

**Simplifié.** On pensait que les araignées étaient protégées de la colle de leurs toiles, alors que les mouches et les autres insectes non. *Mais c'est faux.*

**Exemple 35.** Corpus 44\_SCI IREST Proies.

**Original.** Son mari n'a jamais osé lui dire...

**Simplifié.** Monsieur Dupond ose, pour la première fois, dire...

**Exemple 36.** Corpus DYS (PC62\_Soupe\_de\_la\_discorde\_orig ; PC63\_Soupe\_de\_la\_discorde\_dys).

**Original.** Il n'y a que mamie Marie-Louise qui doit y penser et c'est ça qui la rend triste.

**Simplifié.** Mais mamie Marie Louise doit penser à lui et ça la rend triste.

**Exemple 37.** Corpus 141\_LIT CM1 Chichois.

### 5.6 R22. S6. Constructions passives

Comme pour la négation, la lecture des phrases passives peut entraîner un majeur temps de lecture et nuire à la compréhension. Il est préconisé de transformer les structures passives en phrases actives (respect de l'ordre SVO également).

**Original.** Des microdoses de vitamines et de minéraux sont demandées un peu partout.

**Simplifié.** Le corps demande des petites doses de vitamines et de minéraux.

**Exemple 38.** Corpus 81\_SCI CE2 Manger.

**Original.** Ces fèves sont entassées au soleil, sur des feuilles de bananier.

**Simplifié.** On met ces fèves au soleil, sur des feuilles de bananier.

**Exemple 39.** Corpus 158\_SCI CM1 Chocolat.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Pour un traitement automatique, cette transformation est relativement facile à mettre en place, sauf lorsque le sujet manque dans la même phrase (il est plus loin dans le texte).

**Original.** Il fut attrapé et mis en pâté par Madame McGregor.

**Simplifié.** Votre pauvre père a eu un accident (...) et Mme McGregor en a fait du pâté.

**Exemple 40.** Corpus DYS (PC55 Pierre\_lapin\_orig ; PC54 Pierre\_lapin\_dys).

### 5.7 R23. S7. Constructions attributives

Dans le cas des énumérations, les constructions attributives peuvent être transformées en modifieurs nominaux ou adjectivaux (si cela est possible), ou en syntagmes nominaux.

**Original.** Se sentant rejetés, ils se promenaient toujours ensemble. L'un des garçons est naïf. Le deuxième garçon est peureux. Le troisième est bavard.

**Simplifié.** Les trois garçons se promenaient toujours ensemble : le naïf, le peureux, le bavard.

**Exemple 41.** Corpus DYS (PC56\_Les\_trois\_sots\_orig ; PC57\_Les\_trois\_sots\_dys).

### 5.8 R24. S8. Discours rapporté

Afin de garder une proximité avec le lecteur, il est recommandé de privilégier le discours direct (et éviter, ainsi le discours rapporté et les interrogatives indirectes). C'est déjà le cas dans les corpus pour enfants que nous avons analysé : les personnages des récits s'expriment grâce au discours direct (dialogues avec utilisation des pronoms à la 2<sup>e</sup> personne du singulier ou du pluriel) :

**Original.** S'approchant des enfants, et s'efforçant de donner à sa voix un ton impérieux : - *Quel mal vous a donc fait, dit-il, cette innocente créature, pour la tourmenter de la sorte ? Ignorez-vous que les dieux punissent les enfants qui maltraitent les animaux ?* - *Mêlez-vous* donc de ce qui vous regarde, répond insolemment le plus âgé de la troupe. Cette tortue n'appartient à personne.

**Simplifié.** Taro s'approcha des enfants, et parla avec une forte voix : - *Quel mal vous a fait, dit-il, cette bête, pour la tourmenter ainsi ? Les dieux punissent les enfants qui tourmentent les animaux.* - *Mêlez-vous* de ce qui vous regarde, répond l'enfant le plus âgé de la troupe. Cette tortue n'est à personne.

**Exemple 42.** Corpus Corpus 62\_LIT CE2 Taro et la déesse des océans (pêcheur-tortue).

Les dialogues sont également des structures très fréquentes dans les corpus des romans pour adultes illettrés, ainsi que dans les corpus littéraires pour enfants de CE1 à CM1. Par exemple :

Irina se retourne avec le bébé dans ses bras. La voisine crie : - *Mon Dieu, qu'est-ce que c'est ?* Elle appelle Irina : - *Viens m'aider, Irina. Viens m'aider, vite.* Irina ne comprend pas bien ce qui se passe. Elle ne bouge plus. La voisine l'appelle encore : - *Irina, viens vite.* La voisine crie encore : - *Pousse, Katarina, pousse encore, c'est presque fini.*

**Exemple 43.** Corpus Le Duel (*La Traversée*).

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



**Original.** Un jour, un petit diable (...) s'en vint trouver, afin de le tourmenter, un pauvre paysan qui travaillait sa terre. – *Que fais-tu donc là, paysan ? – Comme vous voyez, je laboure mon champ... - Je vois (...)*

**Simplifié.** Un jour, un petit diable (...) rencontra, afin de l'ennuyer, un pauvre paysan qui travaillait sa terre. – *Que fais-tu donc là, paysan ? – Comme vous voyez, je laboure mon champ... - Je vois (...)*

**Exemple 44.** Corpus 137\_LIT CM1 Le diable et le paysan.

## 5.9 R25. S9. Incises et parenthèses

Les incises et les parenthèses sont supprimées dans les corpus simplifiés afin d'éviter une surcharge cognitive (souvent avec des éléments moins informatifs).

**Original.** La vie dans l'espace, *en dehors de l'atmosphère*, présente des conditions très particulières.

**Simplifié.** La vie dans l'espace présente des conditions particulières.

**Exemple 45.** Corpus 29\_SCI CE1 Espace.

Dans certains cas, ces constructions peuvent être conservées mais remplacées par des unités lexicales plus simples :

**Original.** Pourtant, un jeune homme va prendre le risque, *courageux et héroïque*, pour aider ses amis à traverser cette période difficile.

**Simplifié.** Pourtant, un jeune homme *brave et fort* va prendre le risque pour aider ses amis.

**Exemple 46.** Corpus 138\_LIT CM1 Arbre à deux branches.



## 6 Discours (cohésion textuelle)

		Type de transformation	Intitulé	Préconisation
26	D1	Remplacement	Pronoms et antécédents	Remplacer les pronoms par un antécédent clairement identifié, déjà présent dans le texte (nom propre, groupe nominal défini, groupe nominal indéfini)
27	D2	Remplacement	Changement de déterminant	Privilégier les déterminants plus génériques aux plus spécifiques (son=>le, le=> un) ou parfois les plus génériques deviendront plus spécifiques (un=> le)
28	D3	Remplacement	Répétition du référent	Répéter le référent (groupe nominal défini ou indéfini ou nom propre)
29	D4	Suppression	Suppression des expressions référentielles (pronoms, déterminants possessifs)	Supprimer les pronoms indéfinis, pronoms personnels, déterminants possessifs
30	D5	Remplacement	Accessibilité des expressions référentielles (préférence pour une accessibilité plus basse)	Remplacer un déterminant possessif par le nom propre ou par un groupe nominal plus générique
	D6	Ajout	Ajout de connecteurs logiques	Rajout de connecteurs logiques pour expliciter le schéma de raisonnement (textes scientifiques)

### 6.1 R26. D1. Pronoms et antécédents

L'objectif de cette règle est de réduire le nombre de pronoms, potentiellement ambigus. La résolution des pronoms pose des problèmes aux faibles lecteurs, par les inférences supplémentaires qui sont demandées pour identifier le référent. Un pronom sera remplacé par un antécédent clairement identifié, déjà présent dans le texte (le nom propre ou un groupe nominal défini ou un groupe nominal indéfini), si le pronom est répété et s'il est ambigu (plusieurs référents possibles avec les mêmes propriétés : nombre, genre).

**Original.** « La deuxième amie dit que la soupe a une odeur agréable. Madame Dupont est en colère contre elle. Elle la trouve hypocrite. »

**Simplifié** « La deuxième amie dit que la soupe a une odeur agréable. Madame Dupont est en colère contre elle. Madame Dupont la trouve hypocrite. »

**Exemple 47.** Corpus DYS (PC62\_Soupe\_de\_la\_discorde\_orig ; PC63\_Soupe\_de\_la\_discorde\_dys).

Le deuxième pronom « elle » est remplacé par le référent « Madame Dupont » pour éviter la confusion de référents (même genre et même nombre).

**Original.** En passant dans un bois, elle rencontra compère le Loup, qui eut bien envie de la manger ; mais il n'osa pas, à cause de quelques bûcherons qui étaient dans la forêt. Il lui demanda où elle allait.

**Simplifié.** En passant dans un bois, elle rencontra le Loup. Il avait envie de la manger ; mais il n'osa pas, parce qu'il y avait quelques hommes dans la forêt. Le loup lui demanda où elle allait.

**Exemple 48.** Corpus 02\_LIT CE1 Chaperon rouge.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



**Original.** Qu'allait-il encore trouver dans son cartable à la place des crayons de couleur ? Du poil à gratter ? Des cocottes en papier ? Il y plongea la main en retenant son souffle.

**Simplifié.** Qu'allait-il encore trouver dans son cartable à la place des crayons de couleur ? Du poil à gratter ? Des cocottes en papier ? Il mit la main dans son cartable en retenant son souffle.

**Exemple 49.** Corpus 140\_LIT CM1 Les cartables.

## 6.2 R27. D2. Changement de déterminant

Le changement d'un déterminant du groupe nominal a des implications sur son accessibilité (Ariel, 1990). Ariel (1990) classe les expressions référentielles depuis ceux ayant une accessibilité basse, qui vient d'être introduit dans le discours (nom propre complet) vers ceux qui sont hautement accessibles, qui font référence à une entité déjà introduite dans le discours (Figure 1). Ainsi, le déterminant indéfini indique que le référent vient d'être introduit dans le discours, alors que le déterminant défini est plutôt utilisé pour indiquer un référent accessible (Figure 1).

On remplace les déterminants moins accessibles par des déterminants plus accessibles, indiquant une accessibilité plus grande du référent (un => le, le => ce). Dans d'autres contextes, les déterminants plus accessibles deviendront moins accessibles (cette => la).

**Original.** Un jour de grande chaleur, dans la savane, un renard avait très soif.

**Simplifié.** Un jour de grande chaleur, le renard a très soif.

**Exemple 50.** PC60 Le\_renard\_hyène\_dys, PC61 Le\_renard\_hyène\_orig.

**Original.** Elle l'apporta à sa marraine, sans imaginer comment cette citrouille pourrait la faire aller au bal.

**Simplifié.** À quoi servirait la citrouille ? Cendrillon l'ignorait.

**Exemple 51.** PC28 Cendrillon\_orig ; PC29 Cendrillon\_dys.

## 6.3 R28. D3. Répétition du référent

Pour éliminer toute ambiguïté et limiter les inférences, on remplace une série de pronoms référentiels par leur antécédent répété (groupe nominal ou nom propre), s'ils sont dans la position de sujet de la phrase principale et s'ils se trouvent dans des phrases différentes.

**Original.** « Il aperçut un puits. Sur la poulie, il y avait une corde, et, à chaque bout de la corde, il y avait un seau. Il s'assit dans un des seaux et fut entraîné au fond. Heureux, il but pendant de longues minutes »

**Simplifié.** « Le renard aperçut un puits. Sur la poulie, il y avait une corde, et, à chaque bout de la corde, il y avait un seau. Le renard s'assit dans un des seaux et fut entraîné au fond. Heureux, le renard but pendant de longues minutes »

**Exemple 52.** Corpus dys (PC60 Le\_renard\_hyène\_dys, PC61 Le\_renard\_hyène\_orig)



Dans ce texte, la répétition du sujet « le renard » renforce la présence du référent et la chaîne de référence est composée de trois groupes nominaux identiques. Attention : il ne faut pas remplacer le pronom s'il est dans une phrase coordonnée ou subordonnée :

**Original.** « Il mange et il dort ».

**Simplifié.** « Le renard mange et il dort »

**Exemple 53.** Corpus dys (PC60 *Le\_renard\_hyène\_dys*, PC61 *Le\_renard\_hyène\_orig*)

#### 6.4 R29. D4. Suppression des expressions référentielles (pronoms, déterminants possessifs)

Les transformations syntaxiques (suppression d'une phrase, d'une subordonnée ou d'une relative) peuvent impliquer la disparition de certaines expressions référentielles. En général, on supprime des pronoms avec une accessibilité basse (*chacun, quelqu'un*) ou des déterminants possessifs (*leur*).

La disparition de la phrase « Ils avaient **chacun leur** particularité » entraîne la disparition du pronom indéfini (l'un, l'autre) et du déterminant possessif (leur).

**Original.** « Se sentant rejetés, ils se promenaient toujours ensemble. Ils avaient **chacun leur** particularité : un était plutôt naïf, l'autre plutôt peureux et le dernier plutôt bavard. »

**Simplifié.** « Les trois garçons se promenaient toujours ensemble : le naïf, le peureux, le bavard. »

**Exemple 54.** Corpus dys (PC56 *Les\_trois\_sots\_orig* ; PC57 *Les\_trois\_sots\_dys*).

#### 6.5 R30. D5. Accessibilité des expressions référentielles

On peut remplacer une expression référentielle de haute accessibilité par une expression moins accessible selon Ariel (figure 1).

1) le groupe nominal est remplacé par un groupe plus générique : on remplace le nom par un hyponyme et le déterminant très accessible par un déterminant moins accessible

« ces haricots » => « les graines »

2) le cas des groupes nominaux possessifs (DETPOS N1) : on peut remplacer les groupes nominaux possessifs par une structure plus explicite.

Par exemple, le groupe nominal possessif, composé d'un déterminant possessif et d'un nom, indique deux référents différents dans l'expression « Mme Dupont et **son mari** » : Mme Dupont et « mari » qui est le nouveau référent (M. Dupont).

Nous proposons deux adaptations possibles :

a) remplacement du groupe possessif par le référent :

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



**Original.** « Son mari dit, pour la première fois, qu'il n'aime pas sa soupe. »

**Simplifié.** « M. Dupont dit, pour la première fois, qu'il n'aime pas sa soupe. »

**Exemple 55.** Corpus DYS (PC62\_Soupe\_de\_la\_discorde\_orig ; PC63\_Soupe\_de\_la\_discorde\_dys)

b) le groupe nominal possessif sera remplacé par une structure « N1 de N2 » où N1 est le nom original dans la structure DETPOS N1, N2 est le référent indiqué par le déterminant possessif (« son mari » => « le mari de Mme Dupont ») :

**Original.** « Mme Dupont fait une soupe dont l'odeur est insupportable. Son mari n'a jamais avoué qu'il déteste sa soupe. »

**Simplifié.** « Mme Dupont fait une soupe dont l'odeur est insupportable. Le mari de Mme Dupont n'a jamais avoué qu'il déteste sa soupe. »

**Exemple 56.** Corpus DYS (PC62\_Soupe\_de\_la\_discorde\_orig ; PC63\_Soupe\_de\_la\_discorde\_dys)



## 7 Corpus

### 7.1 Textes édités en version simplifiée/adaptée

Les textes publiés que nous avons utilisé pour notre étude et pour l'élaboration des recommandations présentées dans ce document sont, d'une part une collection de contes pour enfants, niveau CE1-CM2, et d'autre part dix-huit romans des éditions Weyrich destinés à des adultes illettrés.

#### Corpus édité, recueil de CONTES

Le corpus que nous appelons 'CONTES' est constitué de neuf paires de textes originaux et simplifiés (sauf le Petit Poucet pour lequel nous avons eu accès à deux versions simplifiées) :

PC25	Jacques et le Haricot Magique		Jacques_haricot_dys	SIMP
PC26	Jacques et le Haricot Magique		Jacques_haricot_orig	ORIG
PC28	Cendrillon	Charles Perrault	Cendrillon_orig	ORIG
PC29	Cendrillon		Cendrillon_dys	SIMP
PC40	Le Petit Poucet		Poucet_simp_enfant	SIMP
PC41	Le Petit Poucet	Charles Perrault	Poucet_orig	ORIG
PC42	Le Petit Poucet		Poucet_dys	SIMP
PC54	Pierre Lapin		Pierre_lapin_simp	SIMP
PC55	Pierre Lapin	Beatrix Potter	Pierre_lapin_orig	ORIG
PC56	Les trois sots	Conte africain	Les_trois_sots_orig	ORIG
PC57	Les trois sots		Les_trois_sots_dys	SIMP
PC58	Le chêne et le roseau	Jean de La Fontaine	Le_chêne_le_roseau_orig	ORIG
PC59	Le chêne et le roseau		Le_chêne_le_roseau_dys	SIMP
PC60	Le renard et la hyène	Conte (Methodolodys)	Le_renard_hyène_dys	SIMP
PC61	Le renard et la hyène		Le_renard_hyène_orig	ORIG
PC62	La soupe de la discorde	Conte (Methodolodys)	Soupe_discorde_dys	SIMP
PC63	La soupe de la discorde		Soupe_discorde_orig	ORIG
PC64	Le hérisson et le loup	Conte (Methodolodys)	Hérisson_loup_dys	SIMP
PC65	Le hérisson et le loup		Hérisson_loup_orig	ORIG

L'ensemble de textes étudiés totalise 12.223 mots et 770 phrases en version originale, les versions simplifiées contiennent 9.792 mots et 684 phrases au total.



### Corpus édité, collection *La Traversée*

Bien que nos travaux dans le cadre du projet ANR Alector soient destinés à un public d'enfants dyslexiques et faibles lecteurs, nous avons voulu analyser un corpus édité adressé à des adultes en difficulté de lecture afin de caractériser linguistiquement ses propriétés. Pour cela, nous avons eu accès à dix-huit romans de Collection *La Traversée*, éditions Weyrich<sup>13</sup>. Ces romans ont été écrits par des auteurs belges en tenant en compte d'un cahier de charges précis qui contraignait leur style habituel mais qui permettait d'écrire en adéquation avec les besoins du public cible.

Anna	C. Nys-Mazure	2013	Le monde de Nestor	C. Van Acker	2014
Après ta mort	J. Daussain	2012	Le silence de Cologne	M. Cornélis	2014
Du sang sur le couteau	C. Ruwet	2014	Le voleur de lunettes	C. Collins	2015
Gros	J.-P. Echterbille	2013	Les cerises de Salomon	C. Raucy	2012
Histoires ordinaires	V. De Raeve	2015	Les chapeaux rouges	J. Jauniaux	2015
L'attente	A. Fairon	2012	Rue du Chêne	V. Mabardi	2012
La femme cougar	E. Kosma	2015	Sans dire un mot	X. Deutsch	2012
La fille de la poésie	T. Robberecht	2015	Toute une vie	P. Delperdange	2012
Le duel	S. Mulongo	2015	Un autre choix	F. Dolphijn	2013

L'ensemble des dix-huit romans totalise environ 216.000 mots, avec 11.996 mots par roman. On y dénombre un total de 25.245 phrases, avec 1.402 phrases par roman en moyenne (il s'agit de phrases courtes, avec une moyenne de 8,75 mots par phrase).

## 7.2 Textes manuellement simplifiés

Le corpus ALECTOR est un ensemble de textes issus du matériel pédagogique habituellement proposé dans les classes de CE1 à CM1, dans les genres littéraire (contes, fables, histoires) et documentaire scientifique (sciences de la vie et de la terre, majoritairement). L'ensemble de ces 79 textes originaux et leurs équivalents simplifiés est constitué de 183 textes différents, avec 52.704 mots au total et une moyenne de 288 mots par texte (Gala et al., 2020a). Les textes originaux ont une moyenne de 15,5 mots par phrase, les textes adaptés 12,4 ; dans les deux types de textes, original et simplifié, le genre scientifique contient des phrases plus longues. Les adaptations ont été réalisées manuellement par une équipe de linguistes, orthophonistes et enseignants, à l'appui de ressources lexicales comme Manulex et ReSyf.

Une partie de ces textes a été utilisée dans des tests de lecture dans des cabinets d'orthophonistes. Des enfants dyslexiques et faibles lecteurs ont lu à voix haute aléatoirement des textes originaux ou simplifiés. La lecture se faisait à partir d'une tablette, une phrase par écran. L'enfant lisait accompagné de l'orthophoniste, les temps de lecture et la voix de l'enfant étaient enregistrés (afin d'identifier et analyser les erreurs). À la fin de la lecture, l'orthophoniste soumettait oralement un test de

<sup>13</sup> <https://www.lire-et-ecrire.be/latraversee>

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



compréhension sur le texte, trois questions à choix multiple, identiques pour les versions originales ou simplifiées ; l'enfant cochoit la bonne réponse sur la tablette. Les résultats de ces tests ont montré que la simplification de textes permet d'augmenter la vitesse de lecture, en particulier pour les textes scientifiques, et elle permet également de réduire significativement le nombre d'erreurs de lecture (Brunel et Combes, 2015 ; Nandiegou et Reboul, 2018).

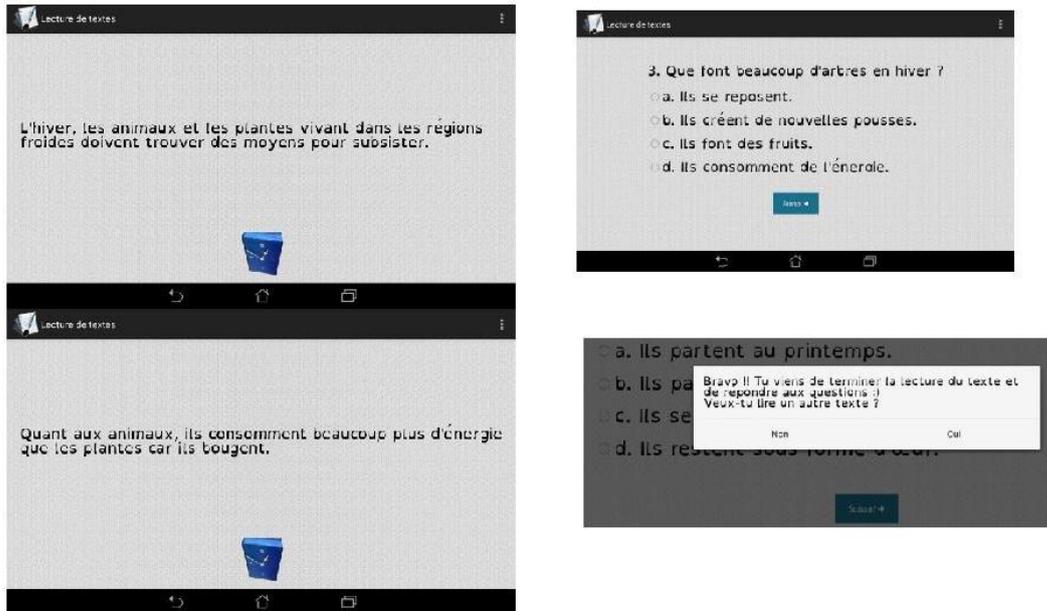


Figure 6. Tablette pour lecture de textes, réalisée par M. B. Billami (2018).

La majorité des textes d'ALECTOR ont également été utilisés dans une étude longitudinale dans six écoles du Var entre 2017 et 2019. Près d'un millier d'élèves ont participé, en individuel et en autonomie, à la lecture silencieuse de textes dans des versions originales ou simplifiées, présentées dans un ordre aléatoire sur des tablettes (cette fois-ci dans le cadre d'une application ludique où l'enfant pouvait gagner des étoiles et suivre la progression de ses résultats). Comme dans les expériences précédentes, les textes ont aussi été présentés phrase par phrase et les enfants devaient répondre à un questionnaire de compréhension à la fin. Le temps de lecture était également enregistré afin d'être analysé par la suite (Javourey-Drevet, 2021).

**En route pour la lecture !**

Appuie sur le bouton vert pour commencer la lecture des textes, ou sur les autres boutons pour jouer avec les mots et les phrases !

**Mesure ta vitesse de lecture en jouant**

**Mots Rigolos**  
Mots - tous niveaux

**Pièges Piégeux**  
Mots - bons lecteurs

**Mots Mêlés**  
Phrases - bons lecteurs

**Lis des textes**

**Textes**  
Tous niveaux

**Résultats Textes et Jeux**  
Code : 5cD1vk

**Aperçu des résultats**

TEXTES	Textes	Questions	Résumés	Pour les niveaux "originaux", "simplifiés" ou "mixtes" (au moins 5, 10 ou 15 textes)			Niveau actuel
				5	10	15	
Niveau 1	1	5 / 5		0	0	0	
Niveau 2	0	0 / 0	0 / 0	0	0	0	
Niveau 3	5	23 / 25	23 / 25	0	0	0	Challenge

(\*) Niveau 1 : lire et 5 questions et un QCM (pour 5 textes) ou 5 textes (5).  
(\*\*) Niveau 2 et 3 : lire et 5 questions et un QCM (lire et 5 questions) pour 10 ou 15 textes (10).

JEUX	Sessions	Score actuel	Meilleur score	10 étoiles	100 points	1000 points	Challenge
Rigolos	3	548	848	1	0	0	187
Piégeux	1	507	507	0	0	0	36
Mêlés	1	784	784	0	0	0	86



**Figure 7.** Tablette avec application ludique pour la lecture de textes et tests de compréhension, réalisée par S. Dufau (Javourey-Drevet (2021).

L'ensemble du corpus ALECTOR est consultable sur le site de la plateforme HumaNum<sup>14</sup>, avec la possibilité de faire des recherches selon différents critères (taille de corpus, difficulté de lecture du corpus, genre littéraire ou documentaire scientifique). Le résultat de la recherche donne accès aux corpus et aux tests de compréhension :

Recherche multi-critères

Mots dans le texte:

Types de textes:  conte  roman  fable  documentaire

Public:  Tous  Enfants

Nombre de mots [113 - 419]:

Difficulté de texte:  Facile  Moyen  Difficile

Corpus disponibles

Titre	Extrait
A quoi sert de manger ?	Apaiser la faim semble le premier but de la nourriture. Mais le véritable objectif est de fournir chaque jour les éléments indispensables...
A quoi servent les satellites ?	Les satellites sont des engins envoyés dans l'espace. Ils reçoivent puis émettent des programmes télévisés, des données météo ou ils...
Comment se protègent les animaux ?	Parmi tous les dangers qui menacent les animaux et les plantes, il y a celui d'être mangé. Certains animaux (propre) se commu...

**Figure 8.** Corpus ALECTOR en ligne.

<sup>14</sup> <http://corpusalector.huma-num.fr/faces/index.xhtml>



## 8 Références bibliographiques

- British Dyslexia Association. *Dyslexia style guide* (2018) <http://www.bdadyslexia.org.uk/>
- Ariel, M. (2001). Accessibility theory: An overview. In Ted Sanders, Joost Schliperoord and Wilbert Spooren eds. *Text representation*. John Benjamins (Human cognitive processing series). (pp. 29-87)
- Ariel, M. (1990). *Assessing noun-phrase antecedents*. London: Routledge.
- Brouwers, L., Bernhard, D., Ligozat, A. L. et François, T. (2012) Simplification syntaxique de phrases pour le français. *Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2012)*, Grenoble.
- Brunel, A. et Combes, M. (2015) Simplification de textes pour faciliter leur lisibilité et compréhension. Mémoire en vue de l'obtention du Certificat de Capacité en Orthophonie, Aix Marseille Université.
- Deahene, S. (2011) *Apprendre à lire. Des sciences cognitives à la salle de classe*. Paris : éd. Odile Jacob.
- Gala, N. et Ziegler, J. C. (2016) Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. *Proceedings of the workshop Computational Linguistics for Linguistic Complexity (CL4LC) at the 26th International Conference on Computational Linguistics (COLING-2016)*. Osaka, Japon.
- Gala, N., Todirascu, A., Bernhard, D., Wilkens, R. et Meyer, J.-P. (2020b) Transformations syntaxiques pour une aide à l'apprentissage de la lecture : typologie, adéquation et corpus adaptés. *Congrès Mondial de Linguistique Française (CMLF 2020)*. Montpellier, France.
- Gala, N., Tack, A., Javourey-Drevet, L., François, T. & Ziegler, J.-C. (2020a) Alector: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. *International conference on Language Resources and Evaluation (LREC 2020)*, pp. 1353-1361. Marseille, France.
- Gala, N., François, T., Javourey-Drevet, L. et Ziegler, J.-C. (2018) La simplification de textes, une aide à l'apprentissage de la lecture. Dans *Langue Française « Lire – écrire : Des savoirs scientifiques aux savoirs pratiques »*, 199 (3). Éd. Liliane Sprenger-Charolles et Alain Desrochers. Armand Colin, pp. 123-131.
- Javourey-Drevet, L., (2021) La simplification de textes comme outil pour améliorer la fluidité et la compréhension de lecture chez les enfants à l'école primaire. Une étude en longitudinal avec des textes littéraires et scientifiques chez des enfants entre 7 et 9 ans. Thèse de doctorat, école doctorale Cognition, Langage, Éducation. Aix Marseille Université.
- Javourey-Drevet, L., François, T., Gala, N., Dufau, S. Ziegler, J.-C. (accepté pour publication 2021) Simplification of literary and scientific texts to improve fluency and comprehension in seven years old learners of French. *Applied Psycholinguistics*.
- Koptient, A. Cardon, R. and Grabar, N. (2019) Simplification-induced transformations: typology and some characteristics. *BioNLP workshop at Association for Computational Linguistics (ACL)*. Florence, Italy, pp. 309-318
- Max, A. (2006) Writing for Language-Impaired Readers. In: Gelbukh A. (eds) *Computational Linguistics and Intelligent Text Processing. CICLing 2006. Lecture Notes in Computer Science*, vol 3878. Springer, Berlin, Heidelberg.
- Nandiegou, M. et Reboul, S. (2018) La simplification lexicale comme outil pour faciliter la lecture des enfants dyslexiques et faibles lecteurs. Mémoire en vue de l'obtention du Certificat de Capacité en Orthophonie, Aix Marseille Université.
- Rauschenberger, M., Baeza-Yates R., Rello, L. (2019) Technologies for Dyslexia. In: Yesilada Y., Harper S. (eds) *Web Accessibility. Human-Computer Interaction Series*. Springer, London.

Projet ALECTOR - ANR-16-CE28-0005  
Novembre 2020



Rello, L. (2014) *DysWebxia : a text accessibility model for people with dyslexia*. PhD Universitat Pompeu Fabra, Barcelona.

Rello, L., R Baeza-Yates, R., Dempere, L. and Saggion, H. (2013) Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. *Proceedings of INTERACT*.

Rello, L. and Baeza-Yates, R. (2013) Good fonts for dyslexia. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

Rello, L., Kanvinde, G. and Baeza-Yates, R. (2012) Layout Guidelines for Web Text and a Web Service to Improve Accessibility for Dyslexics. In *Proceedings of the 9th International Cross-Disciplinary Conference W4A*.

Saggion, H. (2017) *Automatic Text Simplification*. Synthesis lectures on Human Language Technologies. Graeme Hirst, Series Editor.

Seretan, V. (2012) Acquisition of Syntactic Simplification Rules for French. In: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis. *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

Siddhartan, A. (2014) A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*.

Zorzi, M., Barbiero, C., Facchetti, A., Lonciari, I., Carrozzi, M., Montico, M. Bravar, L., George, F., Pech-Georgel, C. and Ziegler, J. C. (2014) Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences (PNAS)* volume 109(28), pages 11455 – 11459.

## Mes publications

### Publications en lien avec cette thèse

- Gala, N., Tack, A., Javourey-Drevet, L., François, T., & Ziegler, J. C. (2020b, May). Alector: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. In *Language Resources and Evaluation for Language Technologies (LREC)*. <https://hal.archives-ouvertes.fr/hal-02503986/>
- Gala, N. & Javourey-Drevet, L. (2020c). Mots ‘faciles’ et mots ‘difficiles’ dans ReSyf : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité. *Lidil. Revue de linguistique et de didactique des langues*, (62). <https://doi-org.lama.univ-amu.fr/10.4000/lidil.8373>
- Gala, N. Todirascu A., Javourey-Drevet, L., Bernhard, D., Wilkens, R. & Meyer J. P. (2020d). Recommandations (*guidelines*) pour des transformations de textes en français afin d’améliorer leur lisibilité et leur compréhension par des enfants dyslexiques et faibles lecteurs. <https://alectorsite.wordpress.com/blog/>
- Gala, N., François, T., Javourey-Drevet, L., & Ziegler, J. C. (2018). La simplification de textes, une aide à l’apprentissage de la lecture. *Langue française*, (3), 123-131. <https://www-cairn-info.lama.univ-amu.fr/revue-langue-francaise-2018-3-page-123.htm>
- Javourey-Drevet, L., François, T., Gala, N., Dufau, S., Ginestié, J. & Ziegler, J. C. (en révision) Simplification of literary and scientific texts to improve fluency and comprehension in beginning readers of French. *Applied psycholinguistics*

### Les articles qui relèvent de travaux avec la même cohorte d'élèves

- Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., & Grainger, J. (2020). Morphological processing across modalities and languages. *Scientific Studies of Reading*, 1-20. <https://doi.org/10.1080/10888438.2020.1730847>
- Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., Grainger, J. & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, e12952. doi: 10.1111/desc.12952 <https://onlinelibrary.wiley.com/doi/full/10.1111/desc.12952>
- Beyersmann, E., Mousikou, P., Schroeder, S., Javourey-Drevet, L., Ziegler, J. C., & Grainger, J. (accepté). The dynamics of learning to read complex words. *Journal of Experimental Child Psychology*