



**HAL**  
open science

## Rapport final du projet ANR ALECTOR (Aide à la LECTure pour amélioRer l'accès aux documents pour enfants dyslexiques)

Núria Gala, Amalia Todirascu, Thomas Francois, Ludivine Javourey-Drevet

### ► To cite this version:

Núria Gala, Amalia Todirascu, Thomas Francois, Ludivine Javourey-Drevet. Rapport final du projet ANR ALECTOR (Aide à la LECTure pour amélioRer l'accès aux documents pour enfants dyslexiques). [Rapport de recherche] Agence Nationale de la Recherche. 2021. hal-03361468

**HAL Id: hal-03361468**

**<https://hal.science/hal-03361468v1>**

Submitted on 1 Oct 2021

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

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

Projet ANR-16-CE28-0005

## ALECTOR

Programme DS0804 2016

A	IDENTIFICATION.....	2
B	RÉSUMÉ CONSOLIDÉ PUBLIC .....	3
B.1	Résumé en français .....	3
B.2	Résumé en anglais.....	3
C	MÉMOIRE SCIENTIFIQUE.....	4
C.1	Enjeux et problématique, état de l'art .....	4
C.2	Approche scientifique et technique .....	4
C.3	Résultats obtenus .....	6
C.4	Conclusion.....	9
C.5	Références.....	10
D	LISTE DES LIVRABLES.....	13
D.1	Ressources .....	13
D.2	Prototypes .....	13
D.3	Guidelines .....	13
D.4	Publications et communications .....	14
D.5	Autres éléments de valorisation.....	17
E	ANNEXES.....	20
E.1	Personnels recrutés .....	20
E.1.1	Post-Doctorants .....	20
E.1.2	Stagiaires .....	20
E.1.3	Doctorants hors financement ANR .....	21
E.2	Site Web public du projet .....	21

## A IDENTIFICATION

Acronyme du projet	<b>ALECTOR</b>
Titre du projet	<b>Reading Aids to leverage Document Accessibility for Children with Dyslexia</b> <i>Aide à la <b>LECT</b>ure pour amé<b>liOR</b>er l'accès aux documents pour enfants dyslexiques</i>
Coordinateur du projet (société/organisme)	Núria GALA, Laboratoire Parole et Langage (LPL UMR 7309), Aix Marseille Université
Date de début du projet Date de fin du projet	<b>01/01/17</b> – 30/06/20 prolongé suite à la situation sanitaire jusqu'au 30/03/2021 (+9 mois) puis jusqu'au <b>30/09/2021</b> (+6 mois)
Labels et correspondants des pôles de compétitivité (pôle, nom et courriel du corresp.)	
Financement demandé	395 k€
Financement ANR	<b>245 k€</b> dont : 83.716 € (LPL AMU) 40.620 € (LPC AMU) 62.645 € (LiLPa Strasbourg) 64.645 € (LIMSI Orsay)
Site web du projet	<a href="https://alectorsite.wordpress.com/">https://alectorsite.wordpress.com/</a>

Rédacteur de ce rapport	
Civilité, prénom, nom	Núria Gala (LPL), Amalia Todirascu (LiLPa), Thomas François (CENTAL), Ludivine Javourey-Drevet (LPC)
Téléphone	
Courriel	<a href="mailto:nuria.gala@univ-amu.fr">nuria.gala@univ-amu.fr</a>
Date de rédaction et envoi	27/09/21
Période faisant l'objet du rapport d'activité	01/01/17 – 30/09/21

## B RÉSUMÉ CONSOLIDÉ PUBLIC

### B.1 RÉSUMÉ EN FRANÇAIS

Le projet **ALECTOR** est un projet multidisciplinaire (linguistique, psycholinguistique, traitement automatique des langues, informatique, sciences de l'éducation) qui vise à proposer des outils d'aide à la lecture pour les enfants dyslexiques et faibles lecteurs. Avec 10% des enfants en difficulté de lecture à l'entrée du collège, la France se situant également en bas de l'échelle au niveau de la compréhension des textes (référence), le problème de l'accès à la lecture est un défi majeur.

Dans ce projet, nous avons étudié ce qui rend difficiles à la lecture les textes utilisés au niveau de l'enseignement en primaire. Pour cela, nous avons : (i) réalisé des tests de lecture à grande échelle dans des écoles, (ii) et dans des cabinets d'orthophonistes libérales ciblés à des enfants avec des retards de lecture importants. Nous avons produit des recommandations et des outils de transformation de textes qui intègrent des nouvelles approches pour la simplification (semi-) automatique de textes en français. Les outils de simplification s'appliquent au niveau lexical, syntaxique et discursif. La simplification lexicale remplace les mots complexes avec des équivalents plus simples. La simplification syntaxique modifie les phrases complexes avec des phrases plus simples : suppression ou modification de subordonnées, des formes passives, des parenthèses. Enfin, la simplification discursive propose le remplacement des pronoms par leur antécédents.

### B.2 RÉSUMÉ EN ANGLAIS

The **ALECTOR** project is a multidisciplinary project (psycholinguistics, natural language processing, education, and language sciences). Its goal is to develop and to test resources and a web-based application that makes it possible to propose simplified texts for children who face major problems in reading and understanding written texts. For these children, text simplification might be a powerful 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 core hypothesis is that the simplification of a text allows children with reading difficulties to get through a text and thus discover the pleasure of reading through understanding what they read. This allows them to enter a 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.

The **ALECTOR** project addresses scientific issues including (a) readability assessment (identifying the complex elements of a text that make it hard to understand and read at an optimal speed), (b) lexical simplification (replacing the complex lexical units detected with simpler equivalents, while dealing with polysemy and multiword expressions), (c) syntactic simplification (replacing complex sentences with shorter ones, avoiding subordinations, passive voices, parenthetical explanations, etc.), and (d) discourse transformations (replacing pronouns with their antecedents).

## C MÉMOIRE SCIENTIFIQUE

### C.1 ENJEUX ET PROBLÉMATIQUE, ÉTAT DE L'ART

La simplification de textes est l'adaptation d'un texte afin de le rendre plus accessible à un public ciblé, tout en conservant le contenu sémantique (Gala et al., 2018). Elle est particulièrement utile dans le cas de publics confrontés à des difficultés de lecture et/ou de compréhension. Depuis plusieurs années, ce domaine a intéressé les professionnels de l'enseignement des langues (enseignants, éditeurs, etc.). Les défis les plus importants peuvent être résumés en deux questions : (1) quelles transformations doivent être opérées pour simplifier un texte tout en conservant son contenu ? (2) lesquelles de ces transformations sont les plus propices à améliorer la compréhension des textes par les lecteurs ?

Par ailleurs, le domaine du traitement automatique des langues s'intéresse depuis quelques années à la problématique de la simplification automatique de textes (SAT), vue comme une alternative à la simplification manuelle (pour une synthèse, voir Siddharthan, 2014, ou Saggion, 2017). Il s'agit d'intervenir sur un texte à l'aide de routines informatiques afin (1) de détecter les éléments linguistiques complexes et (2) de les transformer en des équivalents plus simples à l'aide d'opérations de simplification (du moins celles qui peuvent être automatisées). Pour différentes raisons liées à l'état des technologies (cf. section C.2), au début du projet il n'existait aucun système de SAT complet pour le français. Pendant la période 2016-2021, différentes contributions ont vu le jour et constituent des apports non négligeables (par exemple, Martin et al. 2020). Du côté du projet ALECTOR, nous avons rendu disponible un système de SAT hybride complet pour le français, qui propose des simplifications syntaxiques, discursives et lexicales. Au moment de la rédaction de ce rapport, le système est en cours d'évaluation finale, une publication dans une conférence internationale est prévue pour 2022.

### C.2 APPROCHE SCIENTIFIQUE ET TECHNIQUE

Il est possible de structurer le domaine de la SAT à l'aide de deux **grands axes** :

(A) **type de simplification** : lexicale, syntaxique, discursive ;

(B) **type d'approche** : par règles *vs* approches statistiques (*data-driven*, *machine learning*, *deep learning*) *vs* approches hybrides.

**A.** La simplification **lexicale** consiste à remplacer les mots détectés comme complexes par des équivalents plus simples et adaptés au contexte d'origine. Cela soulève plusieurs difficultés : (1) la détection des mots complexes ; (2) la génération de synonymes adéquats au contexte ; (3) la sélection du synonyme le plus simple. Différentes approches existent basées sur des ressources ou par apprentissage (p. ex. plongements de mots, cf. B).

La simplification **syntactique** consiste à simplifier les structures syntaxiques considérées comme complexes. Généralement, il s'agit de découpages de phrases longues contenant des conjonctions de subordination ou coordination, substitutions de formes passives en formes actives, suppression d'informations superflues (par exemple des appositions). À certains égards, la simplification syntaxique est en lien avec des techniques de résumé automatique. Elle a été soit abordée à l'aide de règles linguistiques appliquées au résultat d'un analyseur syntaxique automatique (ex. Siddharthan & Angrosh, 2014), soit via des techniques d'apprentissage automatique (modèles statistiques)

appliquées à des corpus parallèles (version originale et simplifiée) comme dans les travaux de Woodsend et Lapata (2011).

Enfin, la simplification **discursive** est encore un domaine peu exploité dans le domaine de la simplification automatique mais des travaux commencent à voir le jour (Wilkins et Todirascu, 2020). Le travail d'adaptation se fait ici sur le texte et non pas sur la phrase individuelle (d'où des difficultés plus importantes au niveau des traitements automatiques). Il s'agit d'opérer des modifications dans la structure textuelle, tout en préservant la cohésion textuelle. Ainsi, les anaphores (Quiniou et Daille, 2018) et les chaînes de référence (Schneidecker, 1997) sont des marqueurs de cohésion textuelle que l'on peut modifier pour rendre le contenu textuel plus accessible. Pour faciliter la compréhension des textes et pour éviter des inférences supplémentaires par les lecteurs, nous proposons de remplacer certaines expressions référentielles (pronoms, groupes nominaux démonstratifs etc) par leur antécédents.

**B.** Les **systèmes à base de règles** ont été historiquement les premiers à voir le jour, par exemple (Inui et al., 2003, Brouwers et al., 2014). Il s'agit de systèmes où des experts modélisent des phénomènes linguistiques proposés sur la base d'analyses de textes (notre système s'appuie sur cette technique pour la partie syntaxique et discursive). Cette approche est intéressante dans le cas d'absence de corpus annotés (ce qui était le cas au démarrage du projet ALECTOR) et surtout dans le cas de langues peu dotées. Néanmoins elle présente des inconvénients liés au coût de la mise en place du système (écriture des règles) et de la portabilité (adaptation à d'autres langues).

Les **approches statistiques** se sont développées ces dernières années du fait de la plus grande accessibilité de corpus annotés (surtout pour l'anglais). Différentes techniques existent, l'évolution étant très rapide dans ce domaine. Les premiers travaux se basent sur **la comparaison avec une ressource lexicale**, par exemple WordNet pour l'anglais (Shardlow, 2013) ou ReSyf pour le français (François et al., 2016). D'autres travaux extraient les unités lexicales candidates à partir de **corpus** parallèles (Biran et al., 2011 ; Horn et al., 2014). Cependant, comme ce type de corpus est difficile à obtenir, des approches par **plongement de mots** (*word embeddings*) ont vu le jour (Glavas et Stajner, 2015). Il s'agit de projeter les mots dans un espace de représentation vectorielle. Chaque mot est représenté par un vecteur à  $n$  dimensions qui correspond à une projection du mot dans un espace où les distances modélisent les relations inter-mots (distances sémantiques). L'inconvénient de ces approches est que les plongements lexicaux peuvent produire des candidats à la substitution lexicale qui ne correspondent pas forcément à des synonymes (Hmida et al., 2018).

Les méthodes plus récentes utilisent des **méthodes neuronales** (Zhang & Lapata, 2017; Nisioi et al., 2017; Sulem et al., 2018 ; Abdul Rauf et al., 2020). Les transformations sont directement apprises par le modèle, qui reste une 'boîte noire'. Les traitements lexicaux et syntaxiques sont réalisés conjointement dans le processus de simplification, qui est vu comme une tâche de traduction automatique (TA) (MT pour *Machine Translation*). La simplification de textes est alors considérée comme un problème de traduction monolingue (Specia, 2010). Dans cette approche, les algorithmes de simplification sont appris automatiquement à partir de corpus parallèles constitués de paires de phrases (originales-simplifiées), par exemple (Zhu et al., 2010 ; Xu et al., 2016 ; Abdul Rauf et al., 2020). Ces systèmes nécessitent des corpus originaux et simplifiés de taille conséquente que nous n'avons pas eu à notre disposition pendant le projet.

Récemment des **approches hybrides** se développent afin de tirer profit des avantages des deux approches précédentes (règles et statistiques) en combinant la simplification syntaxique et discursive par règles, et la simplification lexicale par approches statistiques (ex. Siddharthan and Angrosh, 2014). **Les travaux d'ALECTOR se situent dans cette approche.**

Enfin, un des points cruciaux du domaine de la SAT est **l'évaluation des outils**. La tâche est encore difficile, spécialement pour le français, à cause de l'absence de corpus annotés de taille significative et des coûts importants nécessaires pour créer des corpus simplifiés manuellement. Il y a aussi des controverses au niveau de l'utilisation des métriques d'évaluation, qui doivent être corrélées avec le jugement de simplicité des humains (Cardon, 2018). La plus adaptée est SAMSA (Sulem et al., 2018), qui nécessite une pré-annotation sémantique des états, des événements et des participants, indisponible actuellement en français. Alternativement, BLEU compte le nombre de *n-grammes* communs entre le texte original et simplifié : par conséquent, plus le texte original est proche, meilleur est le score. SARI, une autre mesure utilisée dans des travaux de simplification, plus proche des jugements humains, évalue les simplifications par leur simplicité (Xu et al., 2016) et prend en compte plusieurs références.

### C.3 RÉSULTATS OBTENUS

À la fin du projet ALECTOR nous pouvons faire état de plusieurs résultats prometteurs à différents niveaux.

Au niveau des **publics** cibles :

- la simplification de textes améliore les résultats de lecture quel que soit le profil du lecteur (thèse de L. Javourey-Drevet, 2021).
- les enfants dyslexiques et faibles lecteurs font moins d'erreurs de lecture sur des textes adaptés (simplifiés) (article LREC Gala et al., 2020).

Au niveau des **ressources** :

- le corpus **ALECTOR** (Gala et al., 2020) est devenu un corpus de référence pour l'évaluation des systèmes de simplification en français, qui a été cité dans les états de l'art les plus récents sur le domaine (Al-Thanyyan and Azmi, 2021 ; Stajner, 2021).

- la ressource **ReSyF** (Billami et al., 2018) est une ressource de synonymes utilisée et citée dans le domaine du TAL (Al-Thanyyan and Azmi, 2021). Des travaux sont en cours pour le diffuser dans la communauté de la didactique du français (Gala et Javourey-Drevet, 2020).

- le livre de lecture **HIBOU** est un dispositif numérique (tablettes avec tests de lecture) qui intègre le corpus ALECTOR, des questions de compréhension de texte et quelques jeux pédagogiques autour de la lecture. Le dispositif a été testé dans des écoles du Var par 26 enseignants et plus de 600 élèves (Javourey et al. 2019)<sup>1</sup>. Les résultats obtenus à l'issue des tests de lecture ont permis de valider le fait que la simplification des textes aide les enfants faibles lecteurs à mieux comprendre les textes et à augmenter leur vitesse de lecture (Javourey et al., *soumis, en révision*).

Au niveau **technique** :

---

<sup>1</sup>Cf. page 20 du rapport de l'Institut Carnot Cognition 2019 :

[http://www.institut-cognition.com/wp-content/uploads/2020/07/InstitutCognition\\_Yearbook2019\\_A4\\_WEB.pdf](http://www.institut-cognition.com/wp-content/uploads/2020/07/InstitutCognition_Yearbook2019_A4_WEB.pdf)

Un premier système de simplification syntaxique et discursive a été construit, suivant une approche symbolique (Wilkens et Todirascu, 2020). Ces systèmes sont opérationnels et développés en Java et Python. Parallèlement, un système de substitution lexicale a également été développé sous la forme d'une API en Python (Rolin et al., 2021).

Notre système a été construit sur la base des observations réalisées sur un corpus comparable, composé des textes narratifs et informatifs originaux et simplifiés à destination des enfants dyslexiques et faibles lecteurs (le corpus ALECTOR est composé de 79 paires de textes). Nous avons étudié les textes originaux et simplifiés pour identifier les transformations les plus fréquentes, aussi bien au niveau lexical, syntaxique que discursif. Sur la base de ces observations, nous avons défini des recommandations pour la simplification au niveau lexical, morphologique, syntaxique et discursif (Gala et al., 2020) et nous avons choisi d'implémenter plusieurs catégories de règles.

Un système de **substitution lexicale** a été développé (Rolin et al., 2021) en s'inspirant de l'existant pour l'anglais (Paetzold et Specia (2015) ou pour le chinois (Qiang et al., 2021). Ce système sera disponible sous la forme d'un service web API<sup>2</sup>. Le système propose diverses approches classiques ou innovantes pour réaliser trois des quatre principales étapes de la simplification lexicale. L'étape d'identification des mots complexes n'a pas été intégré dans l'API, car elle devrait, à notre avis, faire l'objet d'un système dédié (cf. Tack, 2021). Pour l'étape de génération de synonymes, nous avons implémenté une approche basée sur la ressource ReSyf (Billami et al., 2018), mais également deux méthodes utilisant des *embeddings* (FastText et CamemBERT). Pour la phase de sélection des synonymes – qui consiste à exclure les candidats à la substitution incompatible avec le contexte de la phrase – nous avons implémenté un simple filtre basé sur les catégories du discours, mais aussi deux filtres basés sur FastText. Enfin, pour l'évaluation de la difficulté des différents synonymes, trois méthodes ont été proposées : (1) une basée sur les fréquences lexicales de ces synonymes dans la ressource Lexique<sup>3</sup> ; (2) une qui adapte pour le français le *ranker* Frenetic (Elhadad, 2006) et (3) la dernière qui intègre le modèle SVM de François et al. (2016), qui repose sur 21 caractéristiques lexicales. Les résultats obtenus pour ces différentes méthodes apparaissent comparables à l'état de l'art en anglais et en chinois, en ce qui concerne les scores de rappel, de précision et de F1.

Au niveau de la **simplification syntaxique**, le système met en application neuf types de transformations :

- changement de la voix passive en voix active
- suppression des participiales initiales pour garder l'ordre SVO
- découpage des phrases complexes contenant des conjonctions
- découpage des phrases lorsqu'il y a une subordonnée relative (pronoms dont, où, qui, que, où)
- suppression des parenthèses ou éléments parenthétiques
- transformation des structures clivées

Ces règles syntaxiques ont été définies dans un langage propre, explicitant les conditions d'application de ces règles et les transformations associés. Elles appliquent le formalisme Semgrex<sup>3</sup> pour la description des conditions d'applications : on exploite les propriétés morpho-syntaxiques, lemmes ou les dépendances syntaxiques. Une fois les conditions vérifiées, des opérations de transformations sur les arbres de dépendances s'appliquent (ces transformations s'appuient sur Tsurgeon et ont été adaptées pour modifier les dépendances). Les règles nécessitent un prétraitement à l'aide d'un analyseur syntaxique (StanfordNLP et sa version plus récente, appelée

---

<sup>2</sup> Lien bientôt disponible sur le site du projet ALECTOR).

<sup>3</sup> <https://nlp.stanford.edu/software/tregex.html>

*Stanza*<sup>4</sup>). Les règles sont définies manuellement, après l'analyse détaillée des sorties de l'analyseur syntaxique et aussi l'identification de plusieurs configurations syntaxiques. Pour le développement des règles une interface graphique a été développée en Java : elle permet de mettre en place un suivi des règles (et leur développement). Pour la mise en place des règles nous avons eu recours à plusieurs stagiaires qui ont pris en charge : la mise en place d'une première version fonctionnant avec StanfordNLP et le modèle fr-sequoia, l'extension de ces règles, l'adaptation pour *Stanza* et le modèle GSD.

L'écriture des règles a soulevé plusieurs problèmes : si les conditions sont trop restrictives, les règles s'appliquent uniquement pour des cas très précis. Si elles sont trop floues, les transformations ne peuvent pas être bien contrôlées. Parfois l'application des règles est influencée par les erreurs de l'analyse syntaxique en dépendances. Par ailleurs, l'évaluation de l'ordonnement d'application des règles de transformation syntaxique a aussi posé problème. Pour l'application des règles dans un ordre donné, il faut prendre en compte les modifications qui empêcheront leur application. On doit donc vérifier plusieurs configurations possibles pour proposer la meilleure combinaison. Si la version initiale du système permettait l'application de chaque règle une seule fois dans la phrase, actuellement il est possible d'appliquer la même règle plusieurs fois si les conditions sont satisfaites.

Enfin, la **simplification discursive** se concentre sur la modification de la composition des chaînes de référence et les marques de cohésion textuelle. Les chaînes de référence sont composées d'expressions référentielles (noms propres, groupes nominaux définis, démonstratifs, pronoms personnels et relatifs). Dans le cadre de cette partie du système, nous avons développé un premier module de prétraitement qui annote les chaînes de référence (annote les mentions et les relie dans une seule chaîne). Il s'agit d'un premier module de détection automatique de la coréférence *end-to-end*, disponible pour le français<sup>5</sup>, qui utilise des plongements lexicaux (word2vec) et le corpus français DEMOCRAT pour l'entraînement (corpus annoté en chaînes de coréférence). Ce système applique une technique d'apprentissage à base de réseau de neurones. Les résultats obtenus (85 % de précision) sont comparables avec les systèmes disponibles pour l'anglais. Les expressions référentielles de la même chaîne sont identifiées par ce module. Pour modifier la structure des chaînes, le système remplace les expressions accessibles (pronoms) par des expressions référentielles moins accessibles qui indiquent sans ambiguïté le référent (noms propres, groupes nominaux définis), selon la théorie de l'Accessibilité (Ariel, 1990). Les expressions référentielles n'ont pas la même accessibilité : plus la forme est explicite, plus on a des expressions de faible accessibilité. La simplification discursive applique les transformations suivantes : le remplacement d'un pronom ou d'une série de pronoms par leur(s) antécédent(s), le changement de déterminant.

Pour l'**évaluation du système**, nous avons mis en place une campagne d'annotation manuelle. Chaque sortie générée automatiquement a été annoté par plusieurs annotateurs (4-6 annotateurs) en termes de sens préservé (la simplification garde le sens initial ou pas), de grammaticalité (la phrase résultante est grammaticale ou pas), de simplicité (le résultat est plus simple que la phrase originale), selon les critères explicités dans le guide d'annotation manuelle.

Dans un deuxième temps, nous avons comparé la sortie automatique du système avec un corpus de simplification manuelle (réalisé à l'aide de la plateforme Psytoolkit<sup>6</sup>). Plusieurs étudiants en master Technologies de Langues et Sciences du Langage (Université de Strasbourg) ont réalisé des

---

<sup>4</sup> <https://github.com/stanfordnlp/stanza>

<sup>5</sup> Réalisé en collaboration avec le projet ANR DEMOCRAT(ANR-15-CE38-0008).

<sup>6</sup> <https://www.psytoolkit.org/cgi-bin/psy2.5.1/>

simplifications manuelles selon les recommandations au niveau discursif. Un deuxième corpus a été créé avec les simplifications syntaxiques et manuelles. Nous évaluons les résultats à l'aide du module EASSE, qui propose l'implémentation des métriques SARI et BLEU (métriques utilisées par la communauté de SAT).

## C.4 CONCLUSION

Les objectifs initiaux du projet ont été atteints : nous avons pu identifier des difficultés de lecture auprès d'un échantillon d'enfants dyslexiques et auprès d'une cohorte importante d'enfants (tous profils confondus) d'entre 7 et 9 ans. Nous avons également analysé des textes publiés adressés à des enfants dyslexiques : tout ce travail nous a permis de proposer un ensemble de recommandations d'adaptations de textes à différents niveaux linguistiques (lexical, syntaxique, discursif, cf. section D.3).

Nous avons par ailleurs développé un système qui s'appuie sur ces recommandations pour implémenter des règles de simplification automatique. Ainsi, nous avons développé des modules de simplification automatique pour le français adoptant une méthode symbolique pour le niveau syntaxique et discursif et une approche par apprentissage pour le niveau lexical (cf. section D.2). Le système de simplification syntaxique et discursive, ainsi que l'interface graphique permettant de définir les règles de transformation sont disponibles librement<sup>7</sup>. Nous avons aussi contribué au développement du CoFR, module de détection automatique de la coréférence, conjointement avec le projet ANR DEMOCRAT<sup>8</sup>. Une version est installée sur un serveur hébergé par la TGIR HumNum (qui sera disponible prochainement). En plus de ce système, nous avons constitué un corpus contenant plusieurs simplifications manuelles pour l'évaluation automatique des résultats (à l'aide des métriques telles que SARI ou BLEU). Nous avons aussi mis en place une campagne d'évaluation manuelle selon un guide d'annotation disponible sur le site du projet<sup>9</sup>. Nous avons également testé un système de simplification fondé sur des approches neuronales état de l'art, mais bien que les métriques d'évaluation automatique aient été au niveau de l'état de l'art, les évaluations manuelles ont montré qu'il n'était pas possible d'utiliser ces résultats en contexte réel. Enfin, l'ensemble de ressources développées pour le projet (base de données lexicale ReSyf, corpus ALECTOR et livre de lecture interactif HIBOU) sont tous accessibles en ligne (cf. section D.1).

Plusieurs perspectives de recherche aux travaux menés peuvent être envisagées. Notamment, nous comptons continuer à travailler à l'amélioration de l'outil de simplification automatique hybride par l'ajout de nouvelles règles permettant de hiérarchiser les règles appliquées et par l'implémentation de nouvelles méthodes d'apprentissage dans une perspective d'adaptation des textes à des profils d'apprenants ciblés (*adaptive learning*).

En parallèle, et grâce à l'obtention d'un financement ANRT, nous allons rendre disponible les corpus au travers du livre interactif HIBOU qui sera accessible en ligne. En nous appuyant sur les recommandations linguistiques, nous comptons mettre en place une nouvelle campagne d'adaptation de textes menée par des enseignants de l'éducation nationale que nous allons former.

---

<sup>7</sup> <https://github.com/rswikens/text-rewrite>

<sup>8</sup> <https://github.com/boberle/cofr>

<sup>9</sup> <https://alectorsite.wordpress.com/2020/11/05/recommandations-pour-adapter-des-textes-nov-2020/>

Le résultat de cette campagne permettra d'enrichir la base de corpus ALECTOR, les textes seront accessibles via l'interface dédiée ou via l'application HIBOU.

## C.5 RÉFÉRENCES

Liste complète des publications liées au projet ALECTOR, cf. section D.4 et site web du projet<sup>10</sup> :

**Abdul Rauf, S., Ligozat, A.-L., Yvon, F., Illouz, G. & Hamon, T.** (2020). Simplification automatique de texte dans un contexte de faibles ressources. In C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla & S. Schneider, Eds., *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles*, pp. 332–341, Nancy, France.

Al-Thanyyan, S. S., & Azmi, A. M. (2021). Automated text simplification: A survey. *ACM Computing Surveys (CSUR)*, 54(2), 1-36.

Biran, O., Brody, S. & Elhadad, N. (2011). Putting it simply : a contextaware approach to lexical simplification. In Annual Meeting of the Association for Computational Linguistics, pp. 496-501.

Brouwers, L., **Bernhard, D., Ligozat, A.-L. & Francois, T.** (2014). 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.

Cardon, R. (2018). Approche lexicale de la simplification automatique de textes lexicaux. Actes de la conférence pour le Traitement Automatique des Langues (TALN), pp. 175-189, Rennes, France.

Elhadad, N. (2006). Comprehending technical texts: Predicting and defining unfamiliar terms. In *AMIA annual symposium proceedings* (Vol. 2006, p. 239). American Medical Informatics Association.

**François, T., Billami, M. B., Gala, N. & Bernhard, D.** (2016). Bleu, contusion, ecchymose : tri automatique de synonymes en fonction de leur difficulté. Actes de la conférence *Traitement Automatique des Langues Naturelles* (TALN 2016), Paris.

**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.

Glavaš, G. et Štajner, S. (2015). Simplifying Lexical Simplification: do we need simplified corpora ? *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, pages 63–68, Beijing, China.

**Hmida, F., Billami, M. B., François, T. & Gala, N.** (2018). Assisted Lexical Simplification for French Native Children with Reading Difficulties. *11th International Conference on Language Generation. Workshop on Automatic Text Adaptation*. Tilburg University, The Netherlands.

Horn, C., Manduca, C. & Kauchak, D. (2014). Learning a lexical simplifier using Wikipedia. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, Volume 2, pages 458–463, Baltimore, Maryland, US.

---

<sup>10</sup> <https://alectorsite.wordpress.com/bibliographie/>

- Inui, K., Fujita, A., Takahashi, T. Iida, R. & Iwakura, T. (2003). Text simplification for reading assistance: a project note. In *Proceedings of the 2nd International Workshop on Paraphrasing – Volume 16*, pp. 9–16.
- Javourey-Drevet, L., François, T. & Ziegler, J-C.** (soumis à une revue internationale, en cours de révision). Simplification of Literary and Scientific Texts to Improve Reading Fluency and Comprehension in Beginning Readers of French.
- 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., Dufau, S. & Ziegler, J-C.** (2019). Hibou, un livre électronique et interactif pour améliorer la lecture et la compréhension de textes. Rapport annuel de l'Institut Carnot Cognition.
- Martin, L., Fan, A., de la Clergerie, E., Bordes, A. & Sagot, B. (2020). *Multilingual unsupervised sentence simplification*. Pré-publication (document de travail) ([hal-03109299](https://hal.archives-ouvertes.fr/hal-03109299))
- Paetzold, G., & Specia, L. (2015, July). Lexenstein: A framework for lexical simplification. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations* (pp. 85-90).
- Qiang, J., Lu, X., Li, Y., Yuan, Y. H., & Wu, X. (2021). Chinese Lexical Simplification. In *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1819-1828.
- Quiniou, S. and Daille, B. (2018). Towards a diagnosis of textual difficulties for children with dyslexia. In *11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan.
- Rolin, E.** , Langlois, Q. , Watrin, P. and **François, T.** (2021). FrenLyS: a tool for the automatic simplification of French general language texts. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2021 online)*, session Poster, pp. 1200-1209.
- Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P Dinu. (2017). Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 85–91.
- Saggion (2017). Automatic text simplification. *Synthesis Lectures on Human Language Technologies* 10, 1, pp. 1–137.
- Shardlow, M. (2013). The CW Corpus: A new resource for evaluating the identification of complex words. In *Proceedings of 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations*. 69–77.
- Siddharthan, A. & Angrosh (2014). Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, pp. 722–731.
- Siddharthan, A. (2014). A survey of research on text simplification. *International Journal of Applied Linguistics* 165, 2, pp. 259–298.
- Stajner, S. (2021). Automatic Text Simplification for Social Good: Progress and Challenges. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp.2637-2652.
- Specia, L. (2010). Translating from complex to simplified sentences. In *International Conference on Computational Processing of the Portuguese Language*, pp. 30–39.
- Sulem, E., Abend, O. et Rappoport, A. (2018). Simple and Effective Text Simplification Using Semantic and Neural Methods. *Proceedings of Association for Computational Linguistics (ACL 2018)*. arXiv:1810.05104v1.

- Woodsend, K. et Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the Conference on Empirical Methods in NLP*, pp. 409–420.
- Xu, W., Napoles, C., Pavlick, E., Chen, Q. and Callison-Burch, Ch. (2016). Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics* 4, pp. 401–415.
- Zhang, X. & Lapata, M. (2017). Sentence simplification with deep reinforcement learning. *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2017)*. arXiv:1703.10931v2.
- 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*, pp. 1353–1361.

## D LISTE DES LIVRABLES

### D.1 RESSOURCES

Durant le projet, nous avons créé les ressources suivantes :

- **RESYF**, une ressource lexicale avec des synonymes gradués en fonction de leur difficulté :

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

La ressource est accessible librement et les données distribuables sur demande.

- **ALECTOR**, un corpus de textes comparables originaux/adaptés avec leurs questionnaires de compréhension :

<http://corpusalector.huma-num.fr/faces/index.xhtml>

La plateforme est actuellement utilisée par une centaine de professionnels enregistrés (enseignants, orthophonistes). Les textes et questionnaires sont accessibles en format pdf, distribuables en format .txt sur demande.

- **HIBOU**, un livre électronique de lecture interactive et de jeux développé pour iPad et lisible sur Mac :

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

Nous avons obtenu un financement ANRT (juillet 2021) pour développer l’outil en collaboration avec la société ISI. Le livre interactif sera mis sur une plateforme en ligne afin de la rendre accessible au plus grand nombre (la plateforme **ISIaccess** est accessible pour tous les élèves, dans les écoles et collèges ; elle est agréé GAR par le MENJ).

### D.2 PROTOTYPES

- **Outil de simplification automatique**

Le lien vers le serveur hébergeant le système de simplification syntaxique et discursive sera disponible sur un serveur de la TGIR Huma-Num<sup>11</sup>.

### D.3 GUIDELINES

- **Guidelines** de simplification/adaptation de textes (typologie linguistique)

L’ensemble de recommandations pour adapter les textes afin d’améliorer leur lisibilité et leur compréhension est disponible sur la page du projet<sup>12</sup>.

---

<sup>11</sup> <https://www.huma-num.fr>

<sup>12</sup> [https://alectorsite.files.wordpress.com/2020/11/guidelines-linguistiques\\_alector\\_final.pdf](https://alectorsite.files.wordpress.com/2020/11/guidelines-linguistiques_alector_final.pdf)

- **Guidelines** d'annotation

Le guide d'analyse manuelle des erreurs de simplification du système de simplification est disponible sur la page du projet.

#### D.4 PUBLICATIONS ET COMMUNICATIONS

Liste des publications <b>multipartenaires</b> (résultant d'un travail mené en commun)		
International	Revue à comité de lecture	<ol style="list-style-type: none"> <li>1. <b>Javourey-Drevet, L., François, T., Gala, N., Dufau, S., Ginestié, J. &amp; Ziegler, J.C.</b> (<i>en cours de soumission</i>) Simplification of literary and scientific texts to improve fluency and comprehension in beginning readers of French. <i>Applied Psycholinguistics</i>.</li> <li>2. <b>Gala, N. et Javourey-Drevet, L.</b> (2020) Mots 'faciles' et mots 'difficiles' dans ReSyf : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité. <i>Revue de linguistique et de didactique des langues (LiDiL) n°62</i> « <i>Recherches actuelles en didactique du lexique : avancées, réflexions, méthodes</i> ». Éd.s. Ophélie Tremblay et Dominic Anctil.</li> </ol>
	Ouvrages ou chapitres d'ouvrage	<ol style="list-style-type: none"> <li>1.</li> </ol>
	Communications (conférence)	<ol style="list-style-type: none"> <li>1. <b>Gala, N., Todirascu, A., Bernhard, D., Wilkens, R. et Meyer, J.-P.</b> (2020) Transformations syntaxiques pour une aide à l'apprentissage de la lecture : typologie, adéquation et corpus adaptés. <i>Congrès Mondial de Linguistique Française (CMLF 2020)</i>. SHS Web of Conferences <b>78</b>, 14006.</li> <li>2. <b>Wilkens, R. &amp; Todirascu, A.</b> (2020) Un corpus d'évaluation pour un système de simplification discursive. Actes de la 6e conférence conjointe <i>Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)</i>. Volume 2 : Traitement Automatique des Langues Naturelles. ATALA et AFCP. Nancy, France. pp. 361-369.</li> <li>3. <b>Abdul Rauf, S., Ligozat, A.-L., Yvon, F., Illouz, G &amp; Hamon, T.</b> (2020) Simplification automatique de texte dans un contexte de faibles ressources. Actes de la 6e conférence conjointe <i>Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)</i>. Volume 2 : Traitement Automatique des Langues Naturelles. ATALA et AFCP. Nancy, France. pp. 332-341.</li> <li>4. <b>Wilkens, R. &amp; Todirascu, A.</b> (2020) Simplifying Coreference Chains for Dyslexic Children. <i>International conference on Language Resources and Evaluation (LREC 2020)</i>, pp. 1142-1151. Marseille, France.</li> <li>5. <b>Wilkens, R., Oberle, B., Landragin, F. &amp; Todirascu, A.</b> (2020). French Coreference for Spoken and Written Language. <i>International conference on</i></li> </ol>

		<p><i>Language Resources and Evaluation</i> (LREC 2020), pp. 80-89. Marseille, France.</p> <p>6. <b>Gala, N., Tack, A., Javourey-Drevet, L., François, T. &amp; Ziegler, J.-C.</b> (2020) ALECTOR: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. <i>International conference on Language Resources and Evaluation</i> (LREC 2020), pp. 1353-1361. Marseille, France.</p> <p>7. Goriachun, D. &amp; <b>Gala, N.</b> (2020) Identifying Abstract and Concrete Words in French to Better Address Reading Difficulties. <i>Workshop Tools and Resources to Empower People with Reading Difficulties (READI) at International conference on Language Resources and Evaluation</i> (LREC 2020), poster session, pp. 33-40. Marseille, France.</p> <p>8. <b>Hmida, F., Billami, M. B., François, T. &amp; Gala, N.</b> (2018) Assisted Lexical Simplification for French Native Children with Reading Difficulties. <i>Proceedings of the Workshop Automatic Text Adaptation, 11th International Conference on Natural Language Generation (INLG 2018)</i>, Tilburg.</p> <p>9. <b>Billami, M. B., François, T., Gala, N.</b> (2018) ReSyf: a French lexicon with ranked synonyms. <i>Proceedings of the 27th International Conference on Computational Linguistics (COLING-2018)</i>, Santa Fe, New Mexico.</p> <p>10. <b>Gala, N. &amp; Ziegler, J.</b> (2016) Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. <i>Proceedings of the workshop Computational Linguistics for Linguistic Complexity (LC4LC), workshop at the 26th International Conference on Computational Linguistics (COLING-2016)</i>. Osaka, Japan. <a href="#">〈hal-01757941〉</a></p> <p>11. <b>Todirascu, A., François, T., Bernhard, D., Gala, N. &amp; Ligozat, A.-L.</b> (2016) Are Cohesive Features Relevant for Text Readability Evaluation? <i>Proceedings of the 26th International Conference on Computational Linguistics (COLING-2016)</i>. Osaka, Japan. pp. 987 – 997. <a href="#">〈hal-01430554〉</a></p>
France	Reuves à comité de lecture	<p>1. <b>Gala, N., François, T., Javourey-Drevet, L. &amp; Ziegler, J.-C.</b> (2018) Vers une simplification automatique de textes pour une meilleure compréhension. Dans <i>Langue Française « L'apprentissage de la lecture en français langue maternelle et seconde »</i> (éds. L. Sprenger-Charolles et A. Desrochers), Armand Colin.</p> <p>2. <b>Todirascu, A., François, T., Bernhard, D., Gala, N., Ligozat, A.L. &amp; Khobzi, R.</b> (2017) Chaînes de référence et lisibilité des textes : le projet ALLuSIF. Dans <i>Langue Française</i>, Armand Colin, <i>Les chaînes de référence en corpus</i> (éds. Catherine Schnedecker, Julie Glikman, Frédéric Landragin), 195 (3), pp.35-52.</p>
	Ouvrages ou chapitres d'ouvrage	
	Communications (conférence)	
Actions de diffusion	Articles de vulgarisation	
	Conférences de vulgarisation	

	Autres	<p>Présentation du dispositif numérique (tablettes avec tests de lecture) réalisé au sein de la Communauté d'Agglomération Sud Sainte Baume dans le cadre de l'apprentissage de la lecture au sein de plusieurs classes de la circonscription de Sanary :</p> <p><a href="https://www.facebook.com/watch/?v=2356347051325311">https://www.facebook.com/watch/?v=2356347051325311</a></p> <p><b>Réseau CANOPE, recommandations de ressources pédagogiques,</b> communication du Conseil Scientifique de l'Éducation Nationale sur des logiciels conçus sur des principes scientifiques et testés par de nombreux élèves : Hibou, du CE1 au CM2.</p> <p><a href="https://www.reseau-canope.fr/conseil-scientifique-de-leducation-nationale-site-officiel/outils-pedagogiques/recommandations-de-ressources-pedagogiques.html/">https://www.reseau-canope.fr/conseil-scientifique-de-leducation-nationale-site-officiel/outils-pedagogiques/recommandations-de-ressources-pedagogiques.html/</a></p> <p><b>Javourey-Drevet, L., Dufau, S. &amp; Ziegler, J-C.</b> (2019). Hibou, un livre électronique et interactif pour améliorer la lecture et la compréhension de textes. Institut Carnot cognition - YEARBOOK / FAITS MARQUANTS. 2019.</p> <p><a href="http://www.institut-cognition.com/wp-content/uploads/2020/07/InstitutCognition_Yearbook2019_A4_WEB.pdf">http://www.institut-cognition.com/wp-content/uploads/2020/07/InstitutCognition_Yearbook2019_A4_WEB.pdf</a></p>
--	--------	---

Liste des publications <b>monopartenaires</b> (impliquant un seul partenaire)		
International	Revue à comité de lecture	
	Ouvrages ou chapitres d'ouvrage	
	Communications (conférence)	<ol style="list-style-type: none"> <li data-bbox="683 1153 1444 1377">1. <b>Rolin, E.</b> and Langlois, Q. and Watrin, P. and <b>François, T.</b> (2021). FrenLyS: a tool for the automatic simplification of French general language texts In Proceedings of Recent Advances in Natural Language Processing (RANLP 2021 online), session <b>Poster</b>, pp. 1200-1209.</li> <li data-bbox="683 1388 1444 1612">2. <b>Gala, N.</b> (2019) Mots 'faciles' et mots 'difficiles' : un outil pour la didactique du lexique mobilisant polysémie, synonymie et complexité. <i>87<sup>e</sup> Congrès de l'ACFAS, Recherche en didactique du lexique : avancées, réflexions, méthodes.</i> Université du Québec en Outaouais, 29 et 30 mai 2019.</li> <li data-bbox="683 1624 1444 1780">3. <b>Ziegler, J. C.</b> (2018) Field Experiments in Education and Cognitive Science. <i>Institut d'études avancées de Paris.</i> Organized by Elizabeth Spelke, Esther Duflo, Stanislas Dehaene. Paris, 7 Juin 2018.</li> </ol>
France	Revue à comité de lecture	
	Ouvrages ou chapitres d'ouvrage	
	Communications (conférence)	
Actions de diffusion	Articles de vulgarisation	<ol style="list-style-type: none"> <li data-bbox="683 1993 1444 2072">1. <b>Ziegler, J. C.</b> (2018). Le temps d'apprendre à lire. <i>Cerveau &amp; Psycho</i>, Numéro 100, pp 62-67. Juin 2018.</li> </ol>

	Conférences de vulgarisation	<ol style="list-style-type: none"> <li>1. <b>Gala, N.</b> (2021) Vaincre les difficultés de lecture et compréhension de textes. <i>Université Populaire d'Aubagne et du Pays de l'Etoile (UPOP)</i>. Marseille, janvier 2021.</li> <li>2. <b>Ziegler, J. C.</b> (2019) Dyslexie et troubles de la lecture : des causes aux interventions à l'heure du numérique. IXème <i>Journée Scientifique CERTA/ RESODYS</i>. Marseille, 24 Mai 2019.</li> <li>3. <b>Ziegler, J. C.</b> (2018) La neuroéducation : une nouvelle science pour l'école ? 6ème <i>Congrès de l'Association pour la Recherche en Neuroéducation</i>. Sorbonne, Paris, 6 Juin 2018 .</li> <li>4. <b>Gala, N.</b> (2019) Des machines et des langues : l'ordinateur peut-il s'approprier le langage humain ? <i>Université Populaire d'Aubagne et du Pays de l'Etoile (UPOP)</i>. Marseille, mars 2019.</li> </ol>
	Autres	<ol style="list-style-type: none"> <li>1. <b>Ziegler, J. C.</b> (2018) L'art de lire en enseigner la lecture. Journée de formation. Académie Aix-Marseille, Collège Jean Izzo. 18 Octobre 2018.</li> <li>2. <b>Ziegler, J. C.</b> (2018) <b>Table Ronde</b> : L'échec scolaire. Colloque SFERE-Provence « Apprentissage et Éducation » Conditions, contextes et innovations pour la réussite scolaire, universitaire et professionnelle Marseille 11, 12 et 13 avril 2018.</li> </ol>

## D.5 AUTRES ÉLÉMENTS DE VALORISATION

Liste des éléments. Préciser les titres, années et commentaires	
Brevets internationaux obtenus	
Brevet internationaux en cours d'obtention	
Brevets nationaux obtenus	
Brevet nationaux en cours d'obtention	
Licences d'exploitation (obtention / cession)	
Créations d'entreprises ou essaimage	
Mémoires de Master 2 Sciences du Langage	<ol style="list-style-type: none"> <li>1. Brogialdi, Albin (2021) <i>Amélioration d'un outil de détection et de simplification de la complexité syntaxique des textes en vue de l'apprentissage des enfants dyslexiques</i>. Sous la direction d'A. Todirascu, master Technologies des Langues, Université de Strasbourg.</li> <li>2. Garcia Holgado, Cristina (2021) <i>Analyse thématique fine pour l'analyse de la lisibilité et la simplification automatique de textes en français</i>. Sous la direction d'A. Todirascu, master Technologies des Langues, Université de Strasbourg.</li> </ol>

	<p>3. Goriachun, Daria (2020) <i>Construction d'une base de données lexicale pour les mots français abstraits et concrets. Contribution au projet ANR ALECTOR</i>. Sous la direction de N. Gala, master Sciences du Langage, Aix Marseille Université.</p>
Nouveaux projets collaboratifs	<p>1. Projet ANRT avec la société ISI (V. Wartelle) autour du développement et déploiement d'HIBOU (tests de lecture et compréhension auprès d'enfants du primaire). Participants du projet : L. Javourey-Drevet, JC. Ziegler et N. Gala.</p>
Colloques scientifiques	<p>1. Javourey-Drevet, L., Gala, N., François, T., Dufau, S., Ginestié, J., &amp; Ziegler, JC. (2021). <b>Présentation orale</b> Text simplification of literary and scientific texts to improve reading fluency and comprehension in children learning to read in French. Colloque SFERE-Provence / AMPIRIC « Apprentissages, stratégies et politiques éducatives. Quelles interdisciplinarités, méthodologies et perspectives internationales ? », Marseille. <a href="https://sfere2020.sciencesconf.org/browse/session?sessionid=50290">https://sfere2020.sciencesconf.org/browse/session?sessionid=50290</a></p> <p>2. Javourey-Drevet, L. (2019) (Avec S. Dufau, N. Gala et JC. Ziegler). <b>Poster</b> Workshop de l'Institut Langage Communication Cerveau (ILCB), Porquerolles, mai 2019.</p> <p>3. Gala, N. (2018) <a href="#">Outils d'aide à la lecture et à l'apprentissage du vocabulaire en français langue maternelle</a>. (Avec Javourey-Drevet, L., François, T., Billami, M., Ziegler, JC.) <b>Présentation orale</b> Colloque <i>SFERE-Provence</i> « Apprentissage et Éducation ». Aix-en-Provence, septembre 2018.</p> <p>4. Javourey-Drevet, L. (2018) Outil de recherche numérique en insertion classe : Simplification du texte pour améliorer la fluidité de lecture (Avec Dufau, S., Gala, N., Ginestié, J. et Ziegler, JC.) <b>Présentation orale</b> Colloque <i>SFERE-Provence</i> « Apprentissage et Éducation ». Aix-en-Provence, septembre 2018.</p> <p>5. Javourey-Drevet, L. (2018) Simplification de textes pour améliorer la fluence et la compréhension. (Avec S. Dufau, N. Gala, J. Ginestié et JC. Ziegler). <b>Présentation orale</b> Colloque <i>SFERE-Provence</i> « Apprentissage et Éducation » Conditions, contextes et innovations pour la réussite scolaire, universitaire et professionnelle. Marseille, avril 2018.</p> <p>6. Javourey-Drevet, L. (2017) <a href="#">Le numérique comme outil pédagogique</a> : l'impact de la simplification et de la nature du texte sur la compréhension en lecture, en fonction du profil cognitif d'élèves entre 7 et 9 ans. (Avec S. Dufau, N. Gala, J. Ginestié et J. Ziegler). <b>Poster</b> lors de la <i>Journée Scientifique SFERE</i>, ESPE Aix-en-Provence, septembre 2017.</p> <p>7. Gala, N. (2017) (<b>Séminaire</b> aux Journées d'Orthophonie) Ressources lexicales pour l'aide à la lecture : quelles informations pour quels lecteurs? <i>Le lexique dans tous ses mots</i>, <a href="#">Campus Orthophonique Méditerranéen</a>, La Grande Motte, juin 2017.</p>
Autres (préciser)	<p>1. <b>Organisation du Workshop READI</b> (actes édités) : Gala, N. et Wilkens, R. (2020) <i>Proceedings of the First Workshop on Tools and</i></p>

*Resources to Empower People with REAding Difficulties (READI) at the 12th edition of the Language Resources and Evaluation Conference (LREC2020). Copyright by the European Language Resources Association (ELRA), licensed under CC-BY-NC. ISBN: 979-10-95546-44-3.*

2. **Wilkins, R.** Séminaire FDT/LiLPa, 31 janvier 2020. Transformations des chaînes de coréférence pour l'aide à la lecture.
3. **Gala, N.** Séminaire du CENTAL, 9 avril 2019. [Outils et méthodes pour l'aide à la lecture : vers une simplification automatique de textes pour un meilleur décodage et compréhension](#), Université catholique de Louvain. Louvain-la-Neuve, 5 avril 2019.
4. **Gala, N.** Séminaire du Laboratoire Parole et Langage déc. 2017 <https://amupod.univ-amu.fr/video/1318-seminaire-du-laboratoire-parole-et-langage-cnrs-amu-15-decembre-2017-equipe-de-recherche-ready-representations-et-dynamiques/>
5. **Bernhard, D., Gérard, Ch., Meyer, J.-P. et Todirascu, A.** Séminaire LiLPa, 17 novembre 2017. Le projet ALECTOR Aide à la LECTure pour améliORer l'accès aux documents pour enfants dyslexiques".

## E ANNEXES

### E.1 PERSONNELS RECRUTÉS

#### E.1.1 POST-DOCTORANTS

Avant le recrutement sur le projet								
Nom et prénom	Sexe H/F	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. antérieure (ans)	Partenaire ayant embauché la personne	Date de recrutement	Durée missions (mois) (3)
HMIDA, Firas	H	01/04/19	Doctorat	France	0	LPL	01/02/18	13
WILKENS, Rodrigo	H	24/09/21	Doctorat	Hors UE	2	LILPA	14/03/19	12
ABDUL RAUF, Sadaf	F	30/06/20	Doctorat	France	11	LIMSI	01/03/19	12

#### E.1.2 STAGIAIRES

Nom et prénom	Sexe H/F	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. antérieure (ans)	Partenaire ayant embauché la personne	Date de recrutement	Durée missions (mois) (3)
LAM, Solange	F	Licence 3	UE	0,5	LPL - LIS	15/02/2018	2,5
HUBER, Laurine	F	Master 2	France	0	LiLPa	01/02/2018	5
ROLIN, Eva	F	Master 2	Belgique	1	CENTAL	30/09/2018	3
LAUZIER, Carla	F	Master 1	France	0	LPC	30/05/2019	2
GIRARD, Nathalie	F	Master 2	France	1	LPC	30/05/2019	2
GORYACHUN, Daria	F	Master 1	France	0	LPL	15/05/2019	2,5
MARTIN, Célia	F	Licence 3	France	0	LPL	15/05/2019	2,5
BALLEROY, Ambre	F	Licence 3	France	0	LPL	02/05/2021	2
BOUCHWALDER Léonie	F	Master 2	France	0	LiLPa	25/02/2021	3
EL HAFF Karim	M	Master 2	France	0	LILPa	11/01/2021	2,5
SY Yaya	F	Master 1	France	0	LiLPa	01/06/2021	3

DUNUAN Lara	F	Master 2	France	1	LiLPa	07/06/2021	3,5
BROGIALDI Albin	M	Master 2	France	0	LILPa	10/02/2021	1,5

### E.1.3 DOCTORANTS HORS FINANCEMENT ANR

Trois doctorants ont activement participé aux travaux d'ALECTOR sans avoir été directement financés par le projet :

Nom et prénom	Sexe H/F	Dates	Directeur(s) de thèse	Université	Partenaire	Contributions	Situation professionnelle après la thèse
BILLAMI, Mokhtar	H	2014 - 2018	N. Gala et J.C. Ziegler	Aix Marseille Univ.	LPL	Ressource lexicale ReSyf, désambiguïisation sémantique	CDI
JAVOUREY -DREVET, Ludivine	F	2016 - 2021	J.C. Ziegler et J. Ginestié	Aix Marseille Univ.	LPC	Simplifications manuelles et questionnaires de compréhension, tests dans les écoles en longitudinal (2017, 2018, 2019)	Maître de conférences
TACK, Anaïs	F	2016 - 2021	T. François, C. Fairon et P. Desmet	Univ. catholique de Louvain	CEN TAL	Analyse des erreurs de lecture (tests avec enfants dyslexiques)	Post-Doc

### E.2 SITE WEB PUBLIC DU PROJET

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