



HAL
open science

Recommandations pour des transformations de textes français afin d'améliorer leur lisibilité et leur compréhension

Núria Gala, Amalia Todirascu, Ludivine Javourey-Drevet, Delphine Bernhard,
Rodrigo Wilkens, Jean-Paul Meyer

► To cite this version:

Núria Gala, Amalia Todirascu, Ludivine Javourey-Drevet, Delphine Bernhard, Rodrigo Wilkens, et al.. Recommandations pour des transformations de textes français afin d'améliorer leur lisibilité et leur compréhension. [Rapport de recherche] ANR. 2020. hal-03198905

HAL Id: hal-03198905

<https://hal.science/hal-03198905>

Submitted on 15 Apr 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.

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. GALA¹, A. TODIRASCU², L. JAVOUREY-DREVET^{3,4}, D. BERNHARD², R. WILKENS^{2,5}, J.-P. MEYER²

¹Aix Marseille Univ. et Laboratoire Parole et Langage (LPL UMR 7309)

²Université de Strasbourg et UR Linguistique Langues Parole (LiLPa UR 1339)

³Aix Marseille Univ. et Laboratoire de Psychologie Cognitive (LPC UMR 7290)

⁴Aix Marseille Univ. et UR Apprentissage, Didactique, Évaluation (ADEF UR 4671)

⁵University 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 des temps du 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)¹ 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.

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



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'UNAPEI²). 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 **opérations³ 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 ALECTOR⁴ (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*⁵ 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'information⁶.

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.

² <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%80-lire-et-a%CC%80-comprendre.pdf>

³ « (...) 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).

⁴ Le corpus est accessible ici : <http://corpusalector.huma-num.fr/faces/index.xhtml>

⁵ <https://methododys.ch/>

⁶ 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*). Par exemple, la police en italique devrait être évitée ; autrement, 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*⁷). 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 4) :

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
« 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
« Comme le sujet t'intéresse à ce point, Émilie, c'est toi qui n
- D'accord ! »
Émilie court sans se retourner, son cartable balançant sur sa
chose de rond et dur. C'est un crayon. Elle se baisse, le ramasse
Arrivée chez elle, Émilie va dans sa chambre. La petite fille c
« Oh j'ai une idée ! Je vais dessiner un énorme goûter ! Une brioche
Émilie dessine alors une grosse brioche. Puis elle ajoute un peu
pose le crayon sur son bureau. Au bout de quelques secondes
yeux grands ouverts, Émilie se trouve installée devant un bo

Figure 1. Police Open Dyslexic.

⁷ <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 Association⁸ 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.

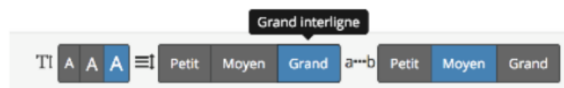


Figure 2. Ajustements de police, espacement interligne et entre caractères dans ALECTOR.

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 Ireland⁹).

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 %.

⁸ <https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide>

⁹ <https://www.dyslexia.ie/information/computers-and-technology/making-information-accessible-dyslexia-friendly-style-guide/>

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 lors qu'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.

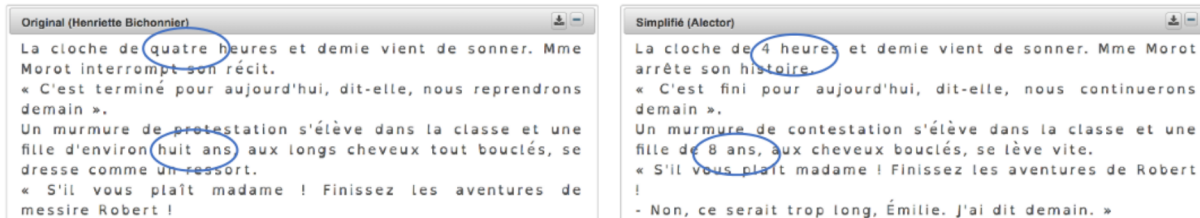


Figure 3. Corpus comparables original et simplifié dans l'interface de corpus ALECTOR, avec remplacement d'expressions numériques en chiffres.

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)).

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 **dress**e comme un **ress**ort.

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 **concou**rs. (CV-CVC)

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.

Par ailleurs, nous avons analysé d'autres corpus manuellement simplifiés. Dans le corpus *Newsele* (corpus en anglais, avec quatre niveaux de simplification), les acronymes sont utilisés surtout dans les

niveaux où la 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 généralement 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.

Dans la ressource Lexique 3¹⁰, la forme 'construction' est renseignée avec une fréquence de 11,3 tandis que 'construire' 26,94 (freqfilms2)¹¹. Dans la ressource Manulex¹², 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 des temps du 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

¹⁰ <http://www.lexique.org>

¹¹ La différence dans les corpus littéraires est moindre, 18,24 vs 18,38 respectivement (freqlivres).

¹² <http://www.manulex.org>

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.

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*).

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.

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.

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 modificateurs 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^e 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*).

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.

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 dans une gradation : de celles ayant une accessibilité basse, qui viennent d'être introduites dans le discours (nom propre complet) vers celles qui sont hautement accessibles, qui font référence à une entité déjà introduite dans le discours. 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.

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.

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, par exemple « 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 :

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 les versions simplifiées contiennent 9.792 mots et 684 phrases au total (originale (respectivement, 15,8 mots et 14,3 mots par phrase en moyenne).

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 Weyrich¹³. 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 (*ibid.*). 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

¹³ <https://www.lire-et-ecrire.be/latraversee>

compréhension sur le texte, trois questions à choix multiple, identiques pour les versions originales ou simplifiées ; l'enfant cochant 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).

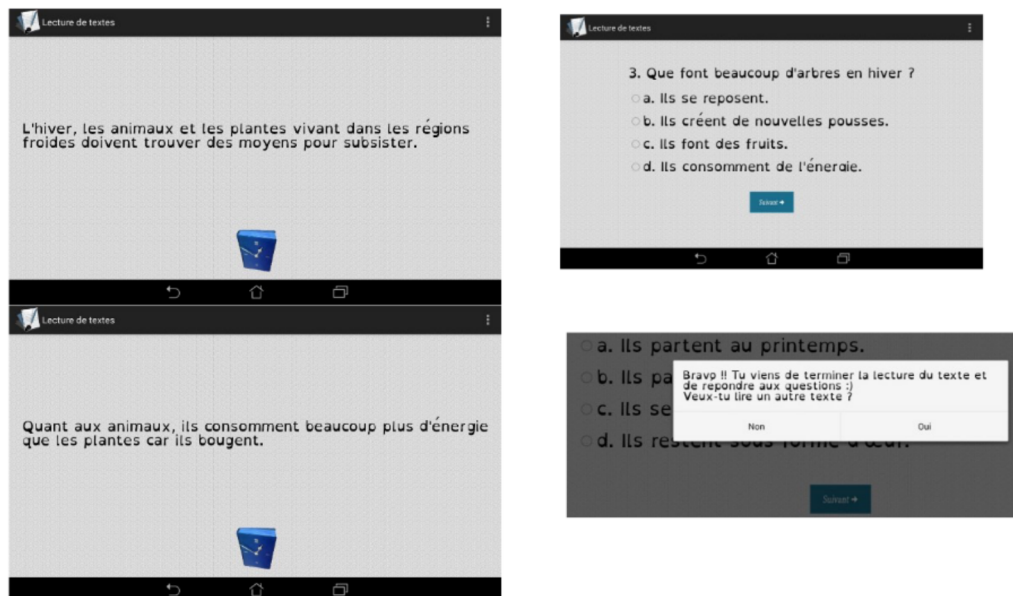
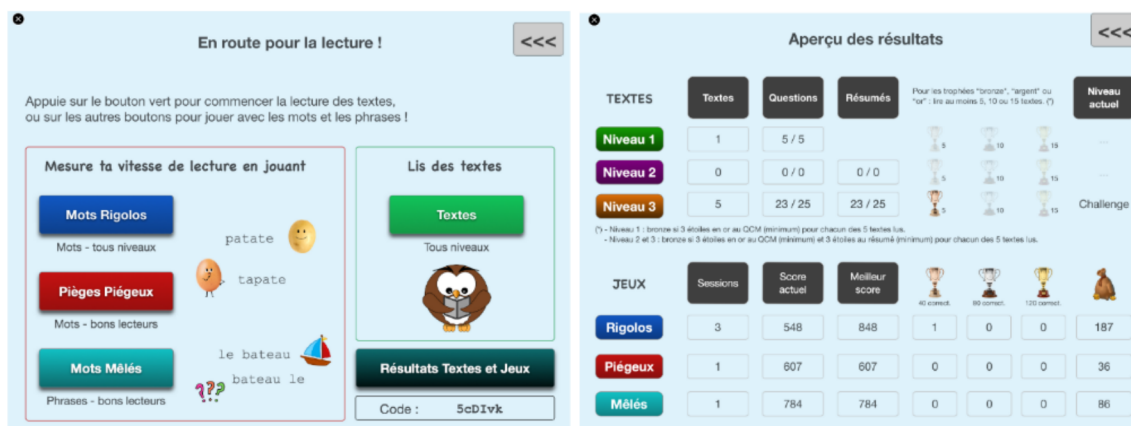


Figure 6. Tablette pour lecture de textes, réalisée par M. B. Billami (2018).

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).



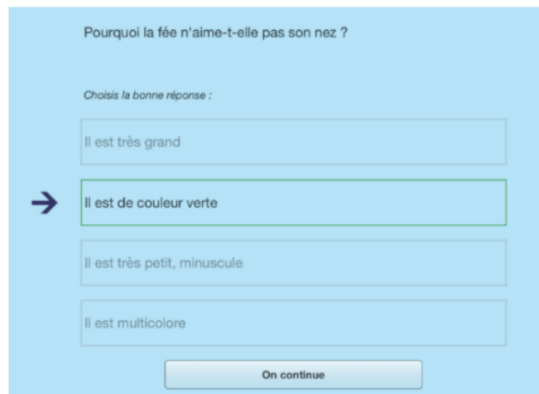


Figure 7. Tablette avec application ludique pour la lecture de textes et tests de compréhension, réalisée par S. Dufau (Javourey-Drevet (2021)).

L'ensemble du corpus ALECTOR est consultable sur un site hébergé par HumaNum¹⁴, 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 :

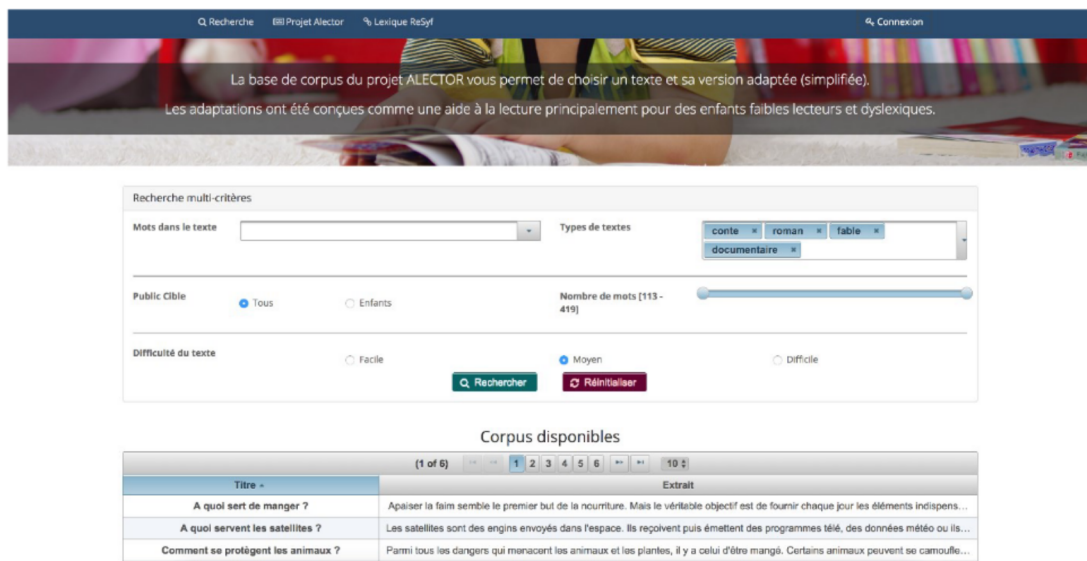


Figure 8. Corpus ALECTOR en ligne.

¹⁴ <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). *Accessing 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.



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*.

Siddhantan, 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.