

Effets des variables linguistiques sur les gains de fluidité et de compréhension en lecture grâce à la simplification de textes écrits

Ludivine Javourey – Drevet₁, Thomas François₂, Núria Gala₃

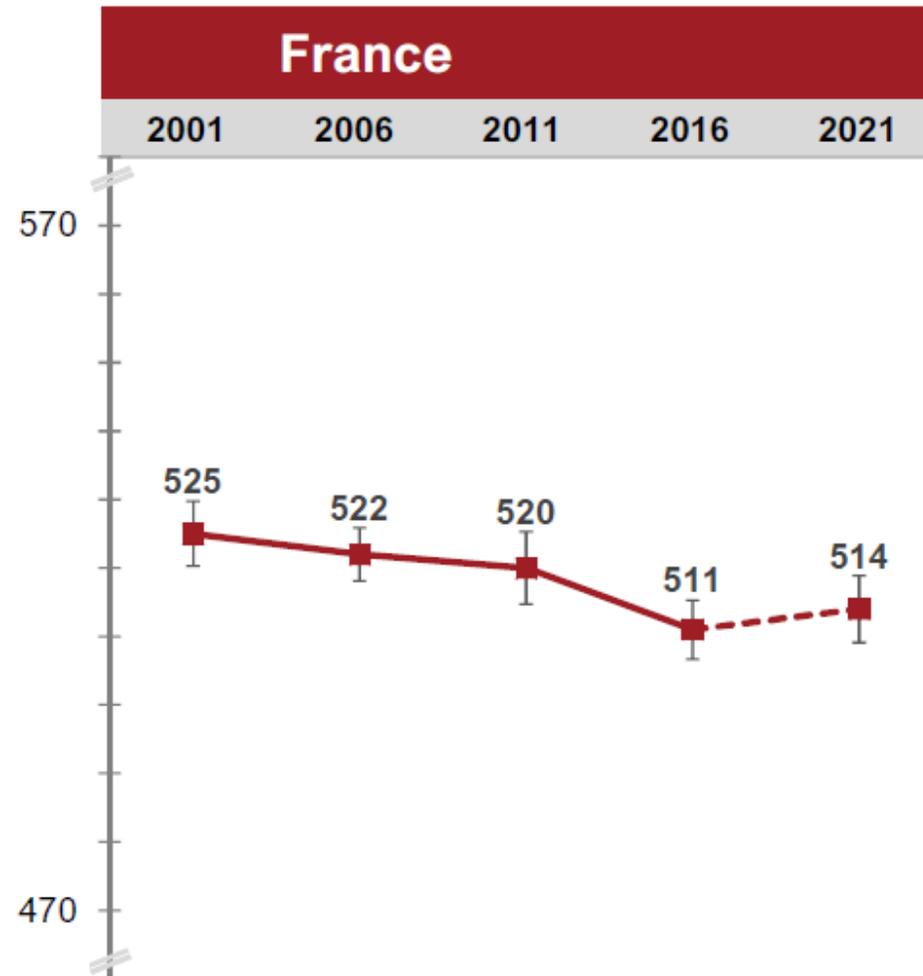
1 Université de Lille, Laboratoire des Sciences Cognitives et Affectives, SCALab UMR 9193

2 Université Catholique de Louvain

3 Aix Marseille Université, Laboratoire Parole et Langage, LPL UMR 7309

Colloque International AFLS 2023 Le français et ses frontières

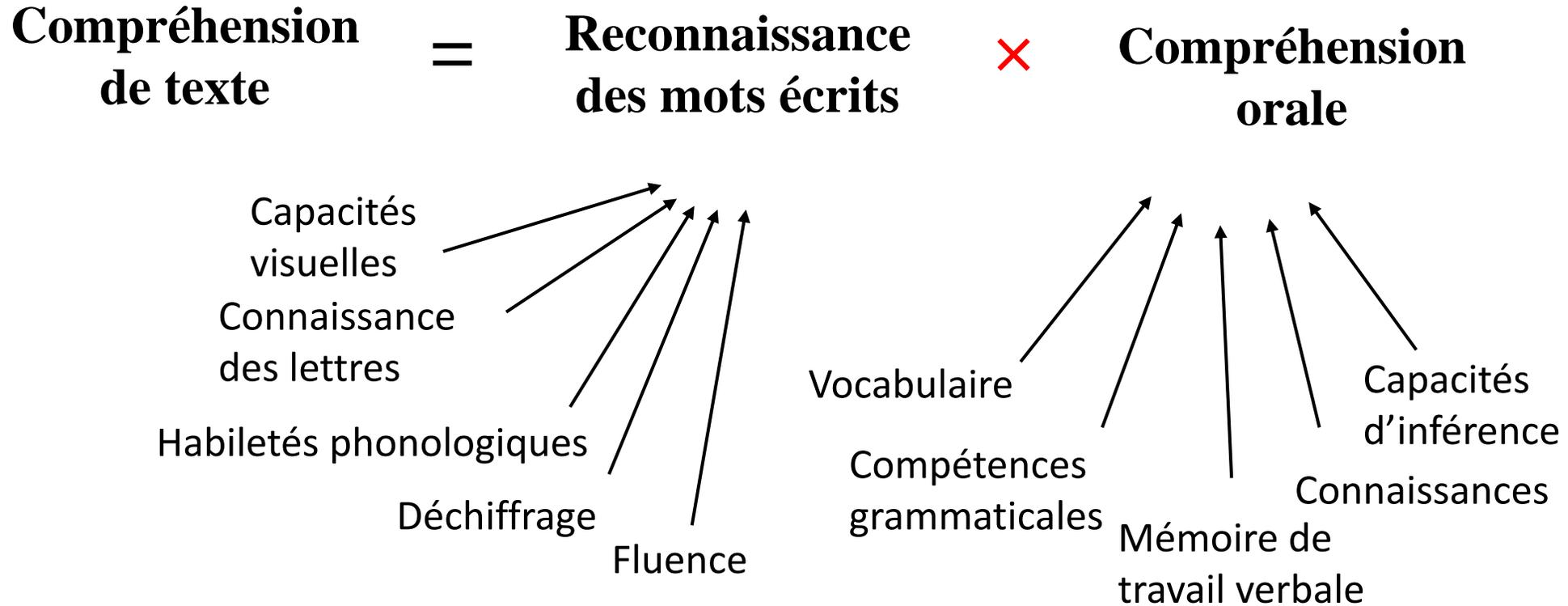
Evaluation PIRLS en France



Mullis, I. V., & Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Modèle simple de la lecture

Castles, A., Rastle, K., & Nation, K. (2018); Hoover & Gough, 1986)



La simplification de textes

- « La simplification de texte consiste à réduire la complexité d'un texte, tout en conservant sa signification d'origine.
- L'objectif de la simplification est de rendre accessible à un public cible le contenu du texte et les informations qui en découlent.
- Transformer un texte en un autre texte qui, véhiculant idéalement le même message, sera plus facile à lire et à comprendre par un public plus large. »

Siddhartan (2014)

Saggion (2017)

Exemples de simplifications des textes

ORIGINAL

- Ce qui les **chagrina** encore, c'est que le plus jeune était **fort** délicat et ne **disait mot** : **prenant** pour bêtise ce qui était une marque de la bonté de son esprit.
- **Bien que** les moulins à vent **aient été remplacés** par les moulins **industriels**, **on peut encore trouver** le moulin à café et le moulin à poivre dans **nos** cuisines.

SIMPLIFIÉ

- Ce qui les **attristait** encore, c'est que le plus jeune était délicat et ne **parlait pas**. **Ils prenaient** pour bêtise ce qui était une marque de la bonté de son esprit.
- **Aujourd'hui**, des moulins **modernes** **ont remplacé** les moulins à vent. **On trouve encore** le moulin à café et le moulin à poivre dans **les** cuisines.

changements lexicaux
changements morpho-syntaxiques
changements discursifs

Protocole expérimental

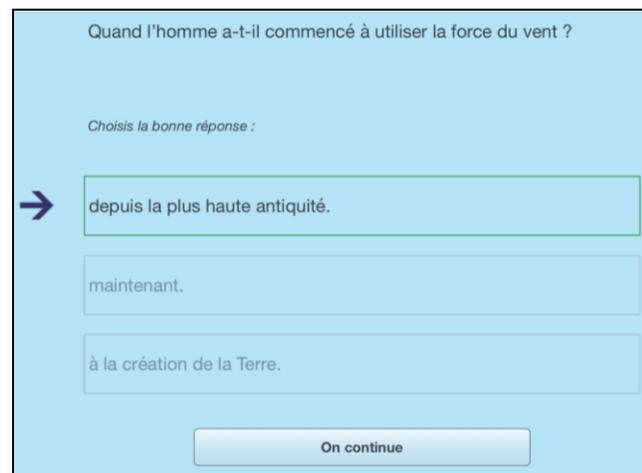
Lecture sur tablette (en autonomie)

Types de texte

- Documentaires scientifiques
- Textes littéraires

Effets de la simplification

- Textes originaux
- Textes simplifiés



Protocole expérimental

Tests cognitifs et langagiers (en individuel)

- - VITESSE DE LECTURE
- - VOCABULAIRE (lexique)
- - COMPREHENSION ORALE
- - PHONOLOGIE
- - MEMOIRE DE TRAVAIL
- - CONSCIENCE MORPHOLOGIQUE
- - COMPETENCES ORTHOGRAPHIQUES
- - DECISION LEXICALE AUDITIVE
- - ...

Timé - 3 

Test d'identification de mots écrits
Feuille de passation

Nom :
Prénom :
Né(e) le :
Date de passation :
Âge (mois) :
Classe :
Enseignant :
Ecole :

	chapeou chapo chapeau chameau cpaheua	bateau	mre mer merle mère men
	moufle monfle moufle moule moufe		thermonètre terminaison termomètre themomètre thermomètre
	disque dique dispue diske discours		tompette trompète tromper trampette trompette
	lire litre lite littre lifre		catédrale cathédrale caténaire cathébrale cathédale
	veste vete veiste veste vessie		cascade casquette casquette caquette casquelte
	gace gloce glacier glasse glace		croissant croïçant craissant croiser coissant
	coche cloche clauche clache clocher		service seruiette serviette serviette serviète
	glode glotte globe glaube gobe		hélicopter hélicotère hélicoïdal hélicoqtère hélicoptère
	brosse brasse broce bosse brode		corbeil corbeille cobeille corbeau cordeille
	zèbe zéro zèdre zèbre zaibre		margarine maguerite marguerite marguerite marguerite
	corde caurde cordée corbe code		grenouilte grenouille grenouil genouille grenadine

membre pate pâtre patte potte pâte	premier secand second cegont secon secoue	main pointe pouin point poing paint	conours examen examen exomen exagérer egsamin
pelle saut sean séant saux seau	cirque clounne clawn clown cloue clowne	doigt pousse pouce pousse poule pauce	livre album atbum alebun albinos albome
buste cor come corps cart cort	chauffage poète poêle poual poël pôle	plante serne sert sert serbe serre	poisson acoiriome aquarium aquorium aquarium aqueduc
objectif bête bul but butoir butte	légume oiguon oignon augnont oinion ogre	oiseau cigne ciguë cygne sigue signe	bûche faou faon fand faond fond
marcher canot câne cane canne cone	alcool rhum ruhm rhum rhume romme	bateau encre aure anche hencre ancre	chanter choralle chonale chorale corral corail

Questions de recherche

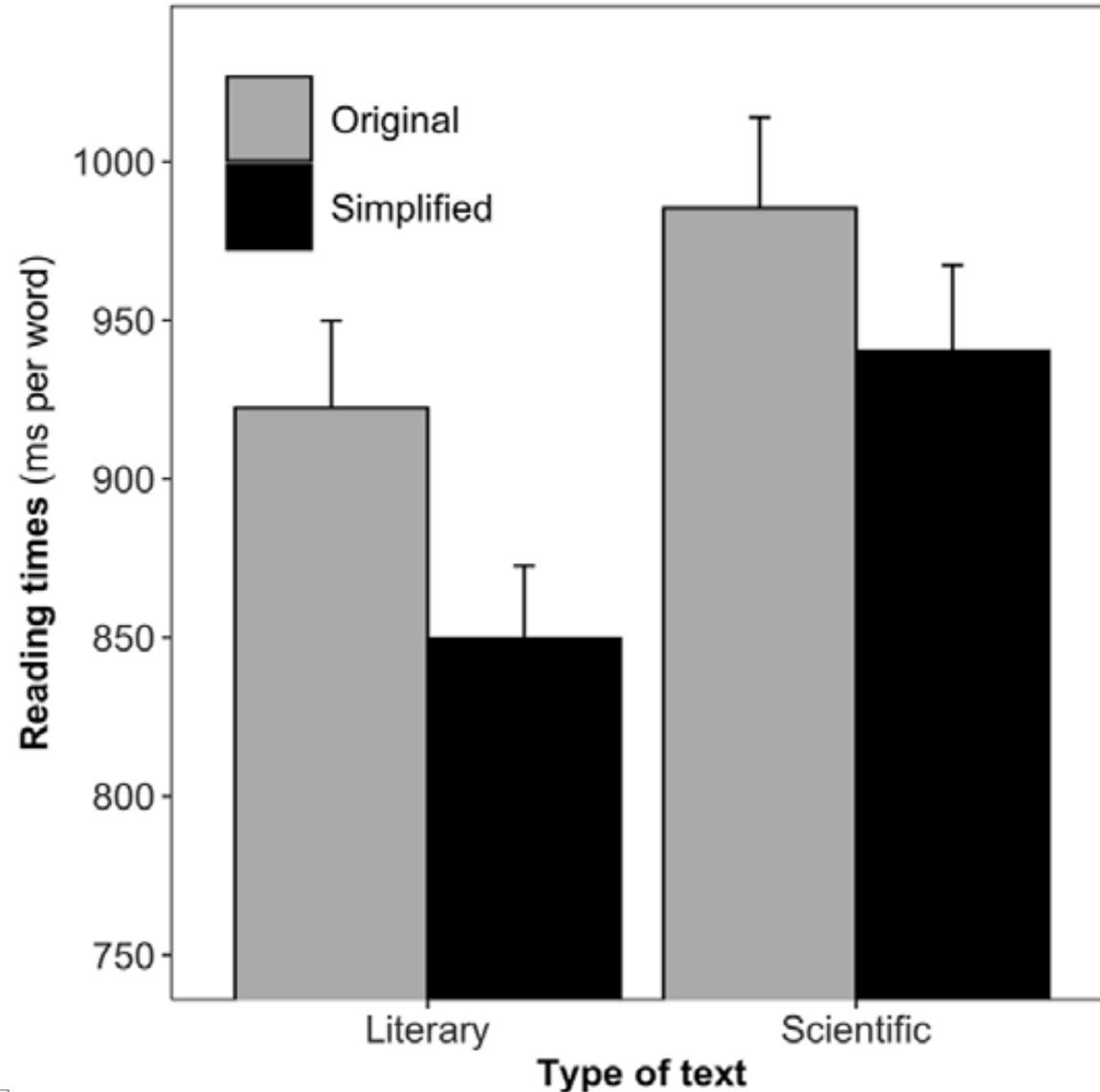
- La simplification des textes améliore-t-elle la fluidité et la compréhension ?
- Quels sont les **facteurs linguistiques** qui expliquent les gains ?
- Textes simplifiés aux niveaux lexical, morpho-syntaxique et discursif
- 10 textes littéraires et 10 documentaires informatifs
- Suivi de la cohorte de la deuxième à la quatrième année (138 élèves)

La simplification des textes améliore-t-elle la fluidité?

Temps de lecture (en millisecondes, ms) par mot pour les versions originales et simplifiées de textes littéraires et scientifiques.

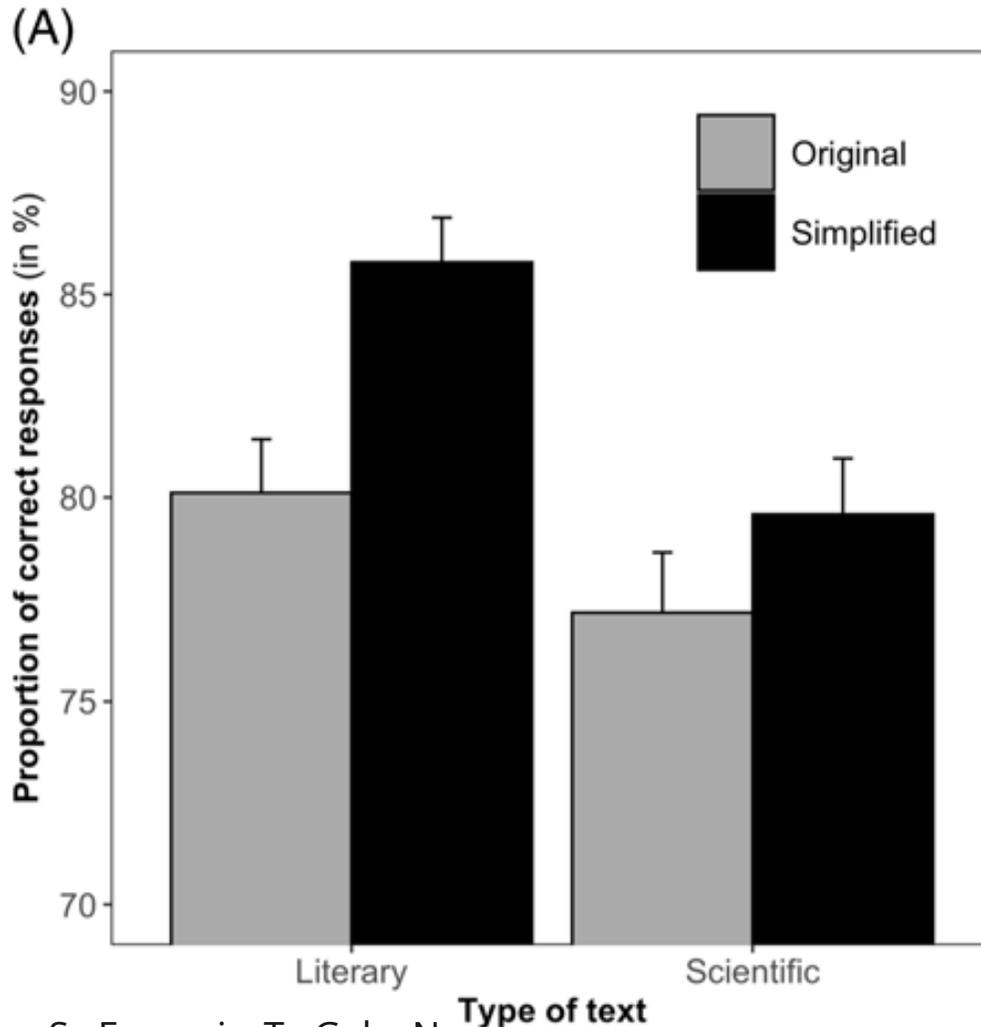
Les barres d'erreur représentent les erreurs standard

Javourey-Drevet, L., Dufau, S., François, T., Gala, N., Ginestié, J., & Ziegler, J. (2022).

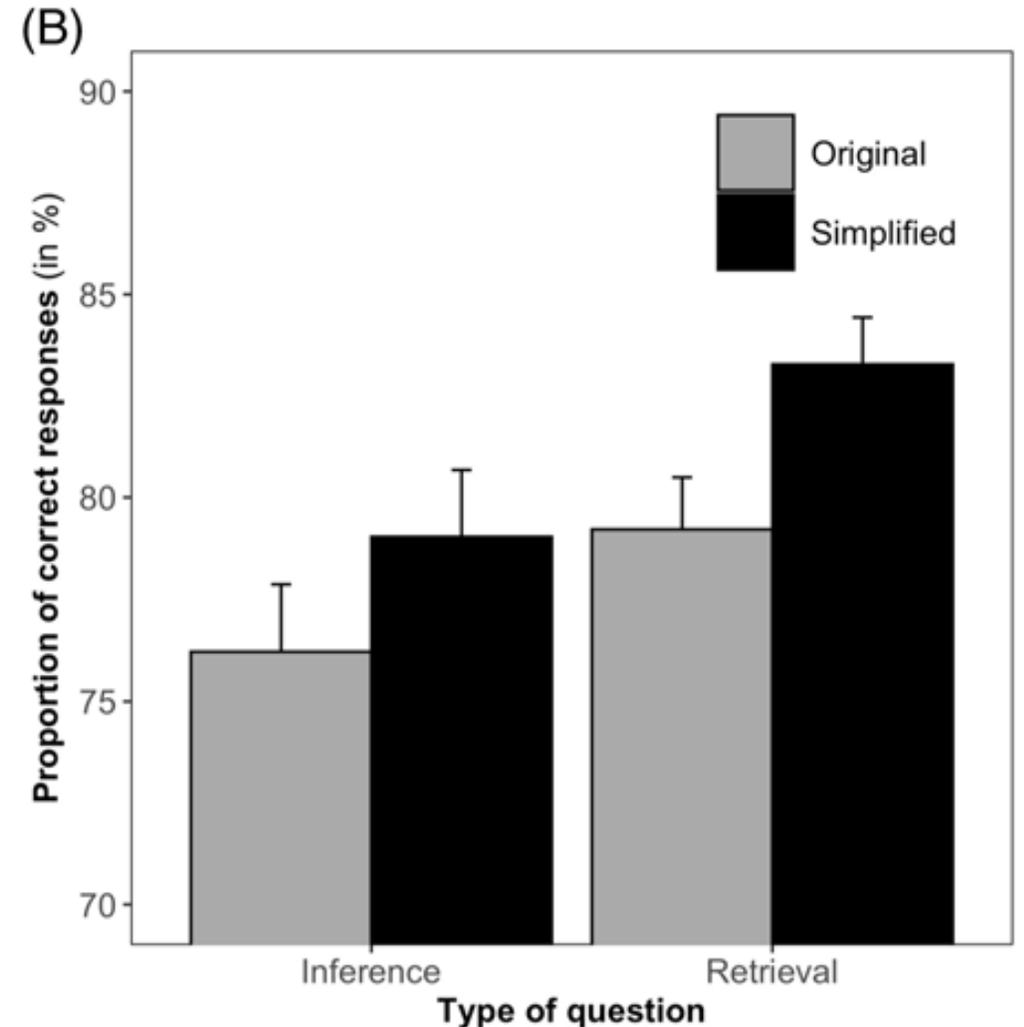


La simplification des textes améliore-t-elle la compréhension ?

A. Proportion de réponses correctes (en pourcentage) pour les textes simplifiés, originaux en fonction du type de texte (littéraire ou scientifique).



B. Proportion de réponses correctes pour les textes simplifiés et originaux en fonction du type de question (recherche ou déduction).



Quels sont les facteurs textuels qui expliquent les gains en fluidité et en compréhension ?

Variables lexicales	RT simple	RT interact	Comprehension simple	Comprehension interact
75 ^e percentile de la distribution de fréquence des lemmes	-21.27	-0.01	0.43	1.38
90 ^e percentile de la distribution de fréquence des lemmes	6.90	-9.47	0.86	0.85
Fréquence moyenne des adjectifs (lemmes)	44.57	-55.69	0.18	1.77 *
Fréquence moyenne des noms (lemmes)	0.42	3.77	-0.12	0.69
Médiane de la distribution de fréquence des adjectifs (lemmes)	-11.30	-24.61 **	-0.17	1.60 *
Modèle d'unigramme basé sur les formes fléchies désambiguïsées par les parties du discours	-23.33 *	-29.87 ***	0.45	1.80 *
Proportion de mots de plus de 8 lettres	11.66	-4.39	-0.45	-1.18
Proportion de mots absents de la liste de 1 063 mots de Gougenheim	38.91 **	18.08 *	-0.62	-1.27
Proportion de mots absents de la liste de 2 000 mots de Gougenheim	9.77	16.90	-0.84	-0.87
Nombre moyen de voisins orthographiques à haute fréquence	-12.80 *	16.59	-0.03	1.48

Coefficients de régression non standardisés entre les variables textuelles et les effets de la simplification en termes de temps de lecture (TR, millisecondes par mot) et de scores de compréhension (pourcentage de mots corrects). Les coefficients correspondent à l'interaction entre les effets de la simplification et une variable textuelle.

Modèle d'unigramme basé sur les formes fléchies désambiguïsées par les parties du discours gains en fluidité et en compréhension

R10.L4. Unités lexicales plus courtes et plus fréquentes

On privilégiera les unités lexicales courtes. La longueur étant corrélée à la fréquence, très souvent les synonymes seront plus courts et plus fréquents (Gala et Ziegler, 2016).

Cela impliquera également de privilégier les graphèmes simples et d'éviter les unités lexicales avec des consonnes doubles (consonnes étymologiques : ph, ff, ll, pp).

- 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.

Proportion de mots absents de la liste de 1 063 mots de Gougenheim

Gains en fluidité

Dans certains cas, les 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.

- [ReSyf est une ressource lexicale en français avec des synonymes gradués en fonction de leur niveau de difficulté.](https://cental.uclouvain.be/resyf/)
- <https://cental.uclouvain.be/resyf/>
- <http://www.lexique.org>
- <http://www.manulex.org>

Parfois, 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.

Médiane de la distribution de fréquence des adjectifs (lemmes) gains en fluidité

Fréquence moyenne des adjectifs (lemme) gains en compréhension

Unités lexicales fréquentes et de moins de sept caractères à privilégier, au détriment d'unités lexicales longues et moins fréquentes

- 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.

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).

- 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

Quels sont les facteurs textuels qui expliquent les Gains en fluidité et en compréhension ?

Variables morphologiques	RT simple	RT interact	Comprehension simple	Comprehension interact
Proportion de verbes au futur	-21.25 *	-5.08	0.72	1.30
Proportion de verbes au participe passé	11.29	2.46	1.47	-1.01
Proportion de verbes au présent simple	22.34	-12.16	0.75	-0.23
Proportion de verbes au participe présent	-0.80	0.86	-0.53	-0.92

Coefficients de régression non standardisés entre les variables textuelles et les effets de la simplification en termes de temps de lecture (TR, millisecondes par mot) et de scores de compréhension (pourcentage de mots corrects). Les coefficients correspondent à l'interaction entre les effets de la simplification et une variable textuelle.

Quels sont les facteurs textuels qui expliquent les gains en fluidité et en compréhension ?

Variables syntaxiques	RT simple	RT interact	Comprehension simple	Comprehension interact
75ème percentile de la distribution du nombre de mots par phrase	-8.89	-25.80 ***	-1.03	-0.04
90ème percentile de la distribution du nombre de mots par phrase	-11.28 *	-40.06 ***	-0.52	-0.19
Nombre médian de mots par phrase	-3.64	-21.80	-0.80	0.12
Nombre moyen de mots par phrase	-12.36 *	-38.05 ***	-0.51	-0.21
Proportion de prépositions par rapport à l'ensemble des articles	-18.42 *	-2.97	-1.23	-0.74
Proportion de pronoms	16.68	-2.82	-3.88 *	1.76 *
Proportion d'articles par rapport à l'ensemble des mots grammaticaux	20.44	-1.64	3.41	-1.02
Ratio des noms propres par rapport à l'ensemble des mots du contenu	11.70	-10.26	-2.04	1.08
Rapport entre les noms propres et les adjectifs	18.53	-3.66	-2.24	0.86
Proportion de noms propres	15.36	-5.35	-1.76	1.10
Rapport entre les noms propres et les pronoms	-4.33	-9.35	0.15	0.34
Rapport entre les pronoms et les mots du contenu	5.48	-2.92	-3.78 *	1.56 *
Ratio pronoms/noms communs	6.15	-2.26	-3.50 *	1.87 **

75ème percentile de la distribution du nombre de mots par phrase

90ème percentile de la distribution du nombre de mots par phrase

Nombre moyen de mots par phrase - Gains en fluidité

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)

R25. S9. Incises et parenthèses

Suppression de structures qui apportent des informations non essentielles à la compréhension

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.

Proportion de pronoms- Rapport entre les pronoms et les mots du contenu - Ratio pronoms/noms communs

Gains en compréhension

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. 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.

Valorisation

Outre les résultats encourageants concernant l'amélioration de la lecture pour les différentes cohortes testées, ce travail a permis :

- La création d'un corpus parallèle (plus de 100 textes originaux et simplifiés), accessible via deux applications en libre accès :
 - Corpus ALECTOR : <http://corpusalector.huma-num.fr/>
 - Application HIBOU : <https://lib.isiaccess.com/HIBOU/connection>
- La définition d'un ensemble de règles de simplification de textes et des recommandations pour la création des questions ([document accessible en ligne – ALECTOR](#))

Conclusion

- La simplification a été étudiée comme levier pour observer des améliorations dans la fluidité et la compréhension en lecture.
- Des nuances sémantiques peuvent être perdues lors de la simplification, sans pour autant entraîner une perte de signification du contenu total.
- Read at the right level – proposer des textes adaptés aux différents niveaux des apprenants!

Références bibliographiques

- Andreu, S., Cioldi, I., Conceicao, P., Etève, Y., Fabre, M., Le Breton, S., Persem, E., Portelli, T., Rocher, T., Rue, G., Vourc'h, R. & Wuillamier, P. (janvier 2021) Évaluations repères 2020 de début de CP et de CE1 Note d'Information n° 21.02, MENJ Depp
- Andreu, S., Cioldi, I., Conceicao, P., Eteve, Y., Fabre, M., Laskowski, C., Le Breton, S., Persem, E., Portelli, T., Rocher, T., Rue, G. & Vourc'h, R. (2021). Progression des performances des élèves de CP à mi-parcours entre 2020 et 2021, Note d'Information, n° 21.29, MENJ Depp
- Bentolila, A. (2011). Le vocabulaire: pour dire et lire. *Ressources pour l'école primaire*.
- Bianco, M. (2015). Du langage oral à la compréhension de l'écrit. PUG.
- Bianco, M., & Lima, L. (2017). Comment enseigner la compréhension en lecture ? Hatier.
- Colé, P. (2011). éduSCOL. Le vocabulaire et son enseignement. http://imagesetlangages.fr/pe2/docs_application/Developpementvoc_ecprimaire.pdf
- Colé, P., & Sprenger-Charolles, L. (2021). *La dyslexie: de l'enfant à l'adulte*. Dunod.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5-51.
- Gala, N., Todirascu, A., Javourey-Drevet, L., Bernhard, D., Wilkens, R. & Meyer, J.-P. (2020) Recommandations pour des transformations de textes en français afin d'améliorer leur lisibilité et leur compréhension. Rapport du projet ANR ALECTOR (ANR-16-CE28-0005).
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, 7(1), 6-10.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, 2(2), 127-160.
- Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., ... & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, 23(6).
- Saggion, H. (2017). *Automatic Text Simplification: Synthesis Lectures on Human Language Technologies*, 1 ed. Morgan & Claypool Publishers; 2017.
- Seidenberg, M. S., & McClelland, J. L. (1989). Un modèle distribué et développemental de reconnaissance et de dénomination des mots. *Revue psychologique*, 96(4), 523.
- Siddharthan, A. (2014). A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2), 259-298
- Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120397

Merci de votre attention

ludivine.javourey@univ-lille.fr

Ce travail a été financé par deux projets ANR :

- Morphème (ANR-15-FRAL-0003)
- Alector (ANR-16-CE28-0005) <https://alectorsite.wordpress.com>

Ce travail se poursuit avec :

- 100% PlusSimple – PIA 3 100%IDT

