Impact of Text Simplification on Eye-Tracking-Based Reading Profiles Across Domains

Oksana Ivchenko Natalia Grabar
CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France
oksana.ivchenko.etu@univ-lille.fr, natalia.grabar@univ-lille.fr

Abstract

Understanding how text readability affects reading behaviour is crucial for improving accessibility and health communication. We analyse sentence-level eye-tracking data from the French Eye-TrAcking (FETA) corpus, which includes original and manually simplified texts from three domains: general, medical, and clinical. Using clustering of fixation-based features, we identify recurrent processing patterns and examine how these patterns change under text simplification. Cluster quality is evaluated using silhouette scores and participant-level bootstrap stability. Simplification does not uniformly reduce reading effort but reorganises processing in domain-dependent ways. Medical texts show strong diversification, general texts moderate diversification, and clinical texts show a reduction in the number of distinct reading profiles. Hence, rather than uniformly facilitating reading, simplification redistributes effort across sentences, underscoring the need for domain-sensitive readability approaches.

Keywords: Readability, Eye-tracking, Text Simplification, Clustering, Reading profiles

1. Introduction

Understanding how text complexity affects reading behaviour is essential for improving accessibility and health communication. This issue is particularly important in specialised domains such as medicine, where texts are often dense and conceptually demanding (Eklics et al., 2024; Brown, 2008). Text simplification is commonly used to improve readability (François and Lefer, 2022), yet its effects on reading dynamics are not always uniform and may vary across domains. Identifying which textual elements drive reading difficulty is crucial for developing more effective simplifications (Cheng Sheang et al., 2022).

Eye-tracking provides an objective way to study reading processes by capturing fixation durations, regressions, and other indicators of cognitive effort (Salvucci and Goldberg, 2000; Duchowski, 2007; Radach and Kennedy, 2013). These measures reveal how readers process complex sentences and where difficulties arise. Reading effort is not homogeneous across sentences, and simplification may reorganise rather than uniformly reduce processing demands. Unsupervised methods, including clustering approaches, have been applied to eye-tracking data to identify patterns of reading behaviour and cognitive processing (Kucharský et al., 2020; Göbel and Martin, 2018; Guan et al., 2025), enabling the discovery of latent processing profiles without predefined categories.

We analyse sentence-level eye-tracking data from original and manually simplified texts in three domains–general, medical, and clinical–to identify recurrent processing profiles. Using clustering of fixation-based features, we examine how these profiles vary across domains and how they change under simplification. This approach allows us to characterise different configurations of reading effort and to assess whether simplification leads to more homogeneous or more diverse processing patterns.

This study addresses the following questions: (i) Can sentence-level eye-tracking features identify distinct processing profiles? (ii) Do these profiles vary across text domains? (iii) How does simplification reshape these profiles?

2. Method

Dataset. Experiments are conducted on the FETA (French Eye-TrAcking) corpus (Ivchenko and Grabar, 2025), a French eye-tracking dataset combining *original* and manually *simplified* texts from three domains: *general* Wikipedia articles, *medical* Wikipedia articles, and *clinical* case reports (toxicology and gastroenterology), thus enabling an analysis of reading behaviour across increasing conceptual complexity. The corpus contains gaze recordings from 46 native French participants (32 women, 14 men; age 18–43 years, $M = 23.3$, $SD = 6.7$) collected with a Tobii Pro Spectrum eye tracker. All participants had normal or corrected-to-normal vision, and no medical training. Texts are distributed across two counterbalanced sets, such that each participant reads only one version (original or simplified) of each text.

Simplification modifies text structure by increasing the number of sentences through syntactic segmentation. By domain, sentence counts increased from 32 to 42 in clinical texts (+31.3%), from 73 to 107 in general texts (+46.6%), and from 144 to 179 in medical texts (+24.3%). As a result, simplified texts contain more, shorter sentences than their original counterparts.

Proceedings of The Second International Workshop onEye-Tracking Resources and Evaluation for Human-Aligned NLP, pages 24–29
May 12, 2026.

Eye-tracking data representation. We perform clustering on sentence-level observations derived from eye-tracking recordings, where word-level measures are aggregated into sentence representations by averaging across words within each sentence. Each observation corresponds to a *participant–sentence* pair, where sentence boundaries follow the experimental segmentation. For each observation, we use a vector of standardized eye-tracking features (z-scores) capturing early and late reading processes (Cook and Wei, 2019; Vasishth et al., 2013). The following set of five non-redundant features was selected for clustering:

- **First-pass first fixation duration** (ms): reflects early lexical processing effort upon first encountering a word;
- **Average duration of fixations** (ms): reflects overall processing effort per word;
- **First-pass regression** (proportion): indicates whether the reader made a backward eye movement from a word during its first encounter, reflecting early integration difficulty;
- **Number of fixations**: represents how many times the reader's gaze landed on the word or sentence (AOI);
- **Re-reading duration** (ms): measures the time spent revisiting a text region after the initial reading pass, calculated as the difference between regression-path duration and first-pass duration.

Observations are analysed separately for each condition defined by `text type` ∈ {clinical, medical, general} and `version` ∈ {original, simplified}.

Outlier removal. Eye-tracking measures are known to be heavy-tailed and sensitive to occasional tracking artifacts. To reduce the impact of extreme values on clustering solutions, we remove outlier observations within each condition. Within each `text type` × `version` condition, features are z-standardised and sentence-level outliers (|z| > 3 on any selected feature) are removed.

Clustering. We cluster observations with k-means (Lloyd's algorithm) using Euclidean distance in the standardized feature space. For each condition, we fit k-means with `n_init`=20 random initializations and select the solution with minimal within-cluster sum of squares. We report the silhouette score (Rousseeuw, 1987) as an internal validation measure. This metric compares how close each observation is to points within its assigned cluster relative to points in the nearest alternative cluster. Higher values indicate better separation and cohesion. In our work, k is fixed per condition based on stability and interpretability considerations.

Stability under participant bootstrap. To ensure that cluster structure is not driven by a small number of participants, we evaluate robustness using participant-level bootstrap resampling. For each condition and chosen k, we first fit k-means on the full dataset and obtain cluster centroids. We then perform $B = 30$ bootstrap iterations, sampling participants with replacement and refitting k-means on the aggregated sentence-level observations. Because cluster labels are arbitrary across runs, bootstrap centroids are aligned to the full-data solution via minimum-cost matching (Hungarian algorithm; Kuhn, 1955). Stability is defined as the mean Euclidean distance between matched centroids, averaged across bootstrap samples. Lower values indicate more stable clustering. Results are reported in table 1.

Text	Version	k	Silhouette	Stability	n
Clinical	Orig.	3	0.215	0.538	638
	Simpl.	2	0.272	0.168	1023
Medical	Orig.	2	0.247	0.194	3372
	Simpl.	5	0.251	0.203	4864
General	Orig.	2	0.256	0.214	1672
	Simpl.	3	0.219	0.202	2242

Table 1: Clustering diagnostics after outlier removal (k: number of clusters; higher silhouette scores and lower centroid distances indicate better clustering quality).

Hierarchical clustering validation. To assess the robustness of the K-means clustering solutions, we performed a cross-validation using agglomerative hierarchical clustering (Ward, 1963) with Ward's linkage and Euclidean distance. Unlike K-means, hierarchical clustering does not rely on random initialization and therefore provides an independent view of the underlying structure.

For each condition (text type × version), we fit hierarchical clustering with the same number of clusters k selected for the K-means analysis. Agreement between the two algorithms was quantified using the Adjusted Rand Index (ARI), which measures the similarity between two partitions while correcting for chance. The index ranges from -1 to 1, where 1 indicates identical partitions and 0 indicates agreement expected by chance.

The comparison revealed moderate-to-high agreement between K-means and hierarchical clustering across most conditions (Figure 1). For the medical and general texts, ARI values ranged from 0.45 to 0.58, indicating that the identified reading profiles are largely algorithm-independent. Lower agreement was observed for the clinical simplified condition (ARI ≈ 0.25), suggesting that while distinct reading strategies exist, cluster boundaries in this condition are less sharply defined.

Overall, the consistency between two independent clustering approaches supports the structural validity of the reported reading profiles and indicates that the main low-effort vs. high-effort distinction is not specific to a single algorithm.

However, it should be noted that simplification in the corpus often involves sentence splitting or merging, which prevents a strict one-to-one alignment between original and simplified sentences. Consequently, rather than performing direct pairwise comparisons, the present approach adopts a condition-level perspective and analyses distributions of sentence-level processing patterns within each condition. As such, the results reflect global differences in reading behaviour across conditions rather than direct sentence-level effects of simplification.

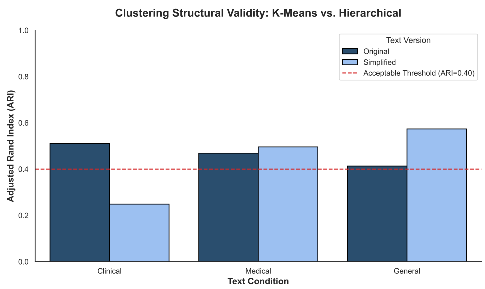

Figure 1: Agreement between k-means and hierarchical clustering across conditions, showing moderate-to-high consistency between solutions.

3. Interpretation of cluster profiles

General Texts. In **original** general texts (Figure 2), sentences cluster into two processing profiles, mainly distinguished by early lexical processing:

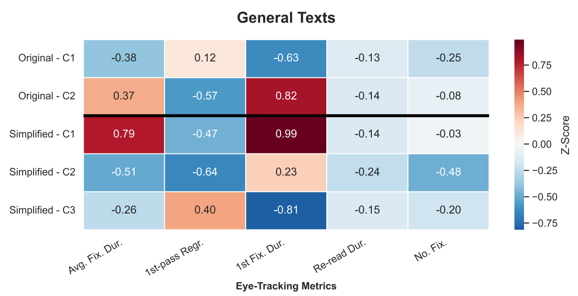

Figure 2: Cluster centroids for general texts (z-scored eye-tracking features).

- **C1.** Shorter fixation durations, reduced first-fixation duration, minimal re-reading, and fewer fixations overall, with slightly elevated early regressions. This profile reflects relatively fast processing with occasional brief corrective regressions.

- **C2.** Longer early fixation durations and higher average fixation durations, combined with few regressions and limited re-reading. This pattern indicates deeper first-pass lexical processing followed by stable linear reading.

In the **simplified** general texts, three distinct profiles emerge, indicating diversification rather than homogenisation of sentence-level processing:

- **C1.** Longer fixation durations and elevated first-fixation durations, with low regression and re-reading rates. This profile reflects deeper but stable first-pass processing.

- **C2.** Shorter fixation durations, reduced regressions, minimal re-reading, and fewer fixations overall, corresponding to fluent processing.

- **C3.** Relatively short fixation durations with increased first-pass regressions and limited re-reading, suggesting fast processing with occasional corrective regressions.

Overall, simplification does not uniformly reduce processing effort but increases the diversity of sentence-level processing patterns, enabling fluent processing for some sentences while others remain more effortful.

Medical Texts. In the **original** medical texts (Figure 3), sentences cluster into two primary processing profiles:

- **C1.** Elevated average and first-fixation durations combined with reduced regression rates. This pattern reflects substantial first-pass lexical integration with stable linear processing.

- **C2.** Shorter fixation durations and lower first-fixation durations, with slightly increased first-pass regressions. This profile corresponds to relatively fluent processing with occasional corrective eye movements.

In the **simplified** medical texts, five distinct sentence-level profiles emerge, indicating strong diversification of processing dynamics:

- **C1.** Elevated fixation durations and high first-fixation durations with low regression and re-reading rates, reflecting intensive but stable first-pass processing.

- **C2.** Very low first-fixation duration and reduced fixation counts, indicating highly fluent early lexical access.

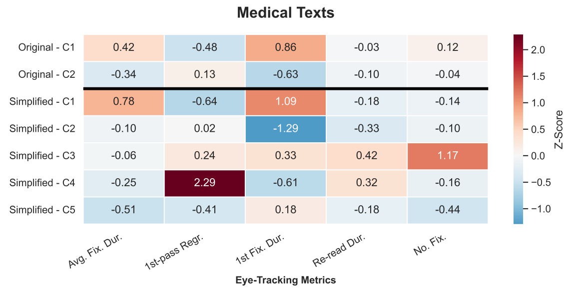

Figure 3: Cluster centroids for medical texts (z-scored eye-tracking features).

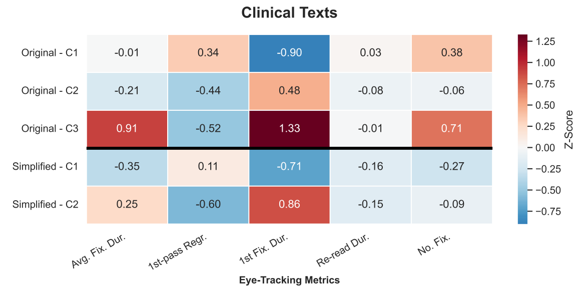

Figure 4: Cluster centroids for clinical texts (z-scored eye-tracking features).

- **C3.** Increased regressions, elevated re-reading duration, and high fixation counts, suggesting reanalysis and repeated inspection.
- **C4.** Extremely high early regressions with short fixation durations, indicating rapid initial processing followed by substantial corrective eye movements.
- **C5.** Globally reduced fixation durations and low regression rates, corresponding to fluent and stable processing.

Taken together, whereas original medical texts exhibit two main processing modes, simplification leads to marked diversification of sentence-level patterns. Rather than uniformly reducing effort, it redistributes processing across distinct profiles.

Clinical Texts. In **original** clinical texts (Figure 4), we observe three sentence-level reading profiles:

- **C1.** Sentences associated with short first-fixation durations and elevated regression rates, combined with increased fixation counts. This pattern suggests rapid initial processing followed by corrective eye movements and additional inspection.
- **C2.** Sentences characterised by longer first-fixation durations but reduced regression activity, indicating careful early lexical integration followed by stable linear reading.
- **C3.** Sentences showing elevated average fixation duration, strongly increased first-fixation duration, and higher fixation counts, reflecting substantial lexical integration effort during the first pass without extensive reanalysis.

In the **simplified** clinical texts, two processing profiles remain:

- **C1.** Sentences associated with reduced fixation durations, lower fixation counts, and minimal re-reading, indicating globally fluent processing.
- **C2.** Sentences characterised by elevated first-fixation durations and slightly increased average fixation duration but reduced regression rates, suggesting deeper initial lexical processing that remains stable and linear.

Overall, while original clinical texts exhibit multiple distinct processing regimes, simplification appears to reduce heterogeneity and promote more stable processing dynamics across sentences.

3.1. Cross-domain comparison of processing profiles

Sentence-level clustering reveals that the impact of simplification on reading dynamics varies across text domains.

In the **general texts**, the number of clusters increases from two in the original condition to three in the simplified condition, while within-cluster variance decreases (2.08 to 1.55). This indicates moderate diversification of processing profiles accompanied by greater internal consistency.

In the **medical texts**, the effect is considerably stronger: clusters increase from two to five, and within-cluster variance drops markedly (2.23 to 1.55). This suggests that simplification does not uniformly reduce processing effort but redistributes it across multiple internally coherent regimes.

In contrast, the **clinical texts** exhibit the opposite pattern. The original texts yield three clusters, whereas the simplified texts yield two, with a slight increase in within-cluster variance (1.83 to 2.02). This indicates a reduction in distinct processing configurations and more homogeneous reading dynamics across sentences.

This pattern suggests that simplification reorganises sentence-level processing in a domain-dependent manner: medical texts show strong diversification, general texts moderate diversification, and clinical texts relative homogenisation, reflecting how readers balance early decoding and later integration processes across domains. Detailed cluster proportions and variance values are reported in Table 2.

Text	Version	k	Avg. Var.	Cluster Sizes (%)	Total Inertia
Clinical	Original	3	1.83	35.3 / 41.8 / 22.9	1166.91
	Simplified	2	2.02	52.0 / 48.0	2066.74
Medical	Original	2	2.23	46.9 / 53.1	7508.05
	Simplified	5	1.55	22.3 / 24.8 / 12.1 / 10.1 / 30.8	7524.70
General	Original	2	2.08	54.1 / 45.9	3475.72
	Simplified	3	1.55	29.4 / 36.3 / 34.3	3481.84

Table 2: Detailed clustering diagnostics per condition.

4. Conclusion

We investigated whether sentence-level eye-tracking behaviour can be grouped into interpretable processing profiles across text domains and levels of simplification. Clustering analyses revealed stable and meaningful patterns distinguished by fixation duration, regressions, and overall processing effort.

Simplification did not produce a single homogeneous processing pattern. Instead, it reorganised reading dynamics in domain-dependent ways. Medical texts showed strong diversification of processing profiles, general texts showed a moderate diversification, and clinical texts showed a relative homogenisation. These findings suggest that simplification redistributes processing demands across sentences rather than uniformly reducing effort.

This work provides a basis for future research on predicting reading effort from textual features and on text simplification. Thus, identifying sentence-level processing profiles can help detect passages that remain effortful even after simplification and guide targeted revisions. Such profiles could also inform predictive models that estimate reading effort directly from text, enabling adaptive simplification and accessibility-oriented writing tools.

Acknowledgement

This work was partially funded by the French National Research Agency (ANR) through the CLEAR project (Communication, Literacy, Education, Accessibility, Readability), ANR-17-CE19-0016-01, and by the French State under the France-2030 programme and the Initiative of Excellence of the University of Lille, which are acknowledged for the funding and support granted to the R-CDP-25-002-PRIME-NEXT-GEN.

We thank the reviewers for their helpful comments and questions, which improved the overall quality of the paper. We also thank Jamal Abdul Nasir (University of Galway) for valuable assistance with the methodological analysis.

5. Bibliographical References

Jo Brown. 2008. How clinical communication has become a core part of medical education in the UK. *Medical Education*, 42(3):271–278.

Kim Cheng Sheang, Anaïs Koptient, Natalia Grabar, and Horacio Saggion. 2022. Identification of complex words and passages in medical documents in French. In *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, pages 116–125, Avignon, France. ATALA.

Anne E. Cook and Wei Wei. 2019. What can eye movements tell us about higher level comprehension? *Vision*, 3(3).

Andrew Duchowski. 2007. *Eye Tracking Methodology: Theory and Practice*. Computer Science. Springer London.

K. Eklics, A. Csongor, A. Hambuch, and J.-D. Fekete. 2024. Diverse integration of simulated patients in medical education for communication, language, and clinical skills in Hungary. *Advances in Medical Education and Practice*, 15:301–312.

Thomas François and Marie-Aude Lefer. 2022. Revisiting simplification in corpus-based translation studies: Insights from readability research. *Meta*, 67(1):50–70.

Zheng-Hong Guan, Sunny S. J. Lin, and Liwei Hsu. 2025. Profiling readers in multiple-text reading: Affective engagement, metacognition, and mediation effects. *2025 IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 338–342.

Fabian Göbel and Henry Martin. 2018. Unsupervised clustering of eye tracking data. In *Spatial Big Data and Machine Learning in GIScience*, pages 25–28. Spatial Big Data. Workshop at the 10th International Conference on Geographic Information Science (GIScience 2018).

Šimon Kucharský, Ingmar Visser, George Ovidiu Truţescu, Patrick G Laurence, Maria Zaharieva, and Maartje EJ Raijmakers. 2020. Cognitive strategies revealed by clustering eye movement transitions. *Journal of Eye Movement Research*, 13(1).

H. W. Kuhn. 1955. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Ralph Radach and Alan Kennedy. 2013. Eye movements in reading: Some theoretical context. *Quarterly journal of Experimental Psychology (2006)*, 66.

Peter J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, page 71–78, New York, NY, USA. Association for Computing Machinery.

Shravan Vasishth, Titus von der Malsburg, and Felix Engelmann. 2013. What eye movements can tell us about sentence comprehension. *WIREs Cognitive Science*, 4(2):125–134.

Joe H. Ward. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

6. Language Resource References

Oksana Ivchenko and Natalia Grabar. 2025. A French eye-tracking corpus of original and simplified medical, clinical, and general texts - FETA. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 37–43, Varna, Bulgaria. INCOMA Ltd., Shoumen, BULGARIA.