Study of Medical Text Reading and Comprehension Through Eye-Tracking Fixations

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

Abstract

Reading plays a crucial role in cognitive processes, acting as the primary way in which people access and assimilate information. However, the ability to effectively comprehend and understand text is significantly influenced by various factors related to people and text types. We propose to study the reading easiness and comprehension of texts through the eye-tracking technology, which tracks gaze and records eye movement during reading. We concentrate on the study of eye-tracking measures related to fixations (average duration of fixations and number of fixations). The experiments are performed on several types of texts (clinical cases, encyclopedia articles related to the medical area, general-language texts, and simplified clinical cases). Eye-tracking measures are analysed quantitatively and qualitatively to draw the reading patterns and analyse how the reading differs across the text types.

Keywords: Medical Texts, Types of Texts, Simplification, Reading, Eye-Tracking, Fixations

1. Introduction

Reading plays a crucial role in cognitive processes, acting as the primary way in which people access and assimilate information (Wylie et al., 2018; K. and Ismail, 2011). However, the ability to effectively comprehend and understand text is significantly influenced by various factors, including a person's level of education, proficiency in the language of the text, and general health (Aarsland et al., 2021; Javourey Drevet et al., 2022; Gala et al., 2018; Pandey et al., 2021). These elements can both facilitate and impede the reading process by impacting how information is processed and comprehended. Therefore, understanding the relationship between these factors and reading comprehension is essential for developing strategies to make information more comprehensible and accessible to a wider audiences.

In addition, text comprehension also varies considerably depending on the text type, and is influenced by key factors that interact with both reader capabilities and text properties (Pugh et al., 2023; Fulcher, 1997). Hence, in the medical area, several types of texts can be distinguished, such as scientific literature, drug inserts, patient leaflets, clinical documents, clinical trial protocols, and encyclopedia articles. In each case, the texts have specific structure, expected recipients, and content (Zweigenbaum et al., 2001; Biber and Finegan, 1994). The reading and comprehension of the different text types may vary due to these different reasons.

Eye-tracking technology provides possibility to study the reading easiness through specific measures. Researchers employ eye-tracking, a technique that tracks gaze patterns and records eye movements while reading, in different contexts. Eye tracking provides objective measures of reading behaviour such as saccade duration, fixation size, pupil dilation and regression frequency (Rayner et al., 1989; Ekstrand et al., 2021; Clifton et al., 2007; Singh et al., 2016).

Longer fixation durations are often indicative of increased cognitive effort, as individuals spending more time at a specific point likely face challenges in interpreting the information presented (Just and Carpenter, 1980; Holmqvist et al., 2011; Ozeri-Rotstain et al., 2020). Thanks to this technique, the past decade has witnessed considerable advancements at the intersection of linguistics, cognitive science, and computer science, enhancing our understanding of cognitive processes, including text readability and comprehension. One of the known limitations of eye-tracking is that only short spans of texts can be processed (Duchowski, 2007). Hence, eye-tracking permits to collect precise reading measures but requires to split the texts into segments that fit the screen.

In our study, we aim to investigate the ease of reading and comprehension of different types of texts (medical and general-language). We employ eye-tracking methodologies, which enable the observation of gaze behaviour and the recording of eye movement data during the reading.

We hypothesise that eye-tracking measures can identify the complexity of texts, highlighting technical passages and terms, that are difficult to read and may require simplification. To investigate this hypothesis, we conducted an experiment where participants were exposed to four types of texts: original clinical texts (clinical cases), medical texts from Wikipedia, general-language texts, and simplified clinical cases. Our objective is to compare the average fixation duration and the number of fixations across these text types to statistically discern their differences. As already indicated, this experiment provides objective measures about the complexity of texts and points out the content which should be simplified.

In what follows, we first introduce the eye-tracking technology and metrics in more detail (Section 2). We then describe the data used in the experiment (Section 3). Sections 4 and 5 are dedicated to the presentation of the principles for statistical analysis of eye-tracking measures related to fixation measures (the average fixation duration and the number of fixations), and the obtained results. Section 4 is focused on the analysis of each type of texts individually, while Section 5 provides an analysis across the types of texts making their comparison. A qualitative analysis of words that require most readers' attention is presented in Section 6. Finally, Section 7 is dedicated to the general conclusion and outlines some directions for future work.

2. Fixations in Eye-Tracking

Eye-tracking provides several objective measures collected during the reading process. Among them, the measures related to fixations occupy an important place.

Fixations, characterised by brief pauses during reading, are pivotal for information processing, serving as indicators of cognitive engagement and interaction with the text. Longer fixations often signal processing difficulty or heightened interest, while more frequent fixations may indicate the text's challenging nature or its ability to engage the reader. The average duration of an eye fixation on a word during reading varies depending on several factors, including the complexity of the text, the reader's familiarity with the content, and the purpose of reading (Hyönä and Kaakinen, 2019).

However, in general, research in the field of eye movement and reading suggests that the average eye fixation duration on a word is approximately 200-250 milliseconds (ms) for adults reading texts in their native language under normal conditions (Rayner and Reingold, 2015; Rayner et al., 2006). Hence, fixations tend to be longer for less common or more complex words, as the reader may require additional time to process the meaning of such words. Conversely, familiar or highly predictable words may receive shorter fixations, or even be skipped entirely, as the reader's brain can efficiently predict their meaning based on context.

Figure 1 illustrates a visual representation of eye-tracking data superimposed on a passage of a medical case text. Each circle represents a fixation point, where the number within the circle indicates the sequential order of fixations, and the size of the

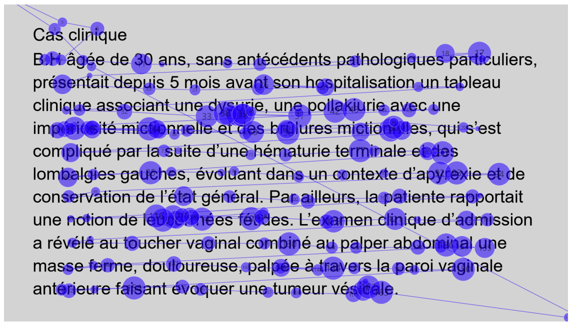

Figure 1: Example of fixations in a medical text. Translation of the excerpt from the clinical case: *Clinical case. B.H, a 30-year-old female with no particular pathological history, presented with symptoms for 5 months before her hospitalization, which included dysuria, frequent urination with urinary urgency, and burning sensations during urination. These symptoms were later complicated by terminal hematuria and left-sided lower back pain, progressing in a context of no fever and maintained general health. Additionally, the patient reported having foul-smelling leukorrhea. The admission clinical examination revealed through a combined vaginal touch and abdominal palpation a firm, painful mass palpated through the anterior vaginal wall, suggesting a bladder tumor.*

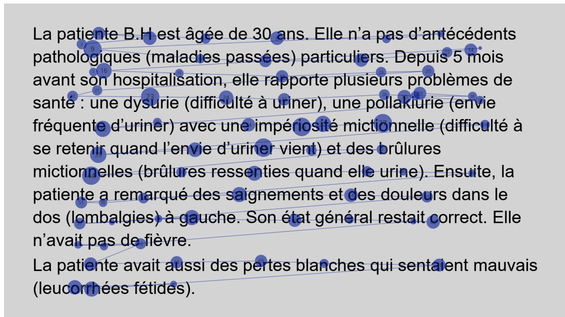

Figure 2: Example of fixations in a simplified clinical case. Translation of the excerpt from the clinical case: *Clinical Case. Patient B.H. is 30 years old. She has no particular past medical history. For five months before her hospitalization, she reported several health problems: dysuria (difficulty urinating), pollakiuria (frequent urge to urinate), urinary urgency (difficulty holding urine when the urge comes), and micturition burns (burning sensations when urinating). Subsequently, the patient noticed bleeding and pain in the left side of her back (left lumbar pain). Her overall health remained stable. She did not have a fever. The patient also experienced foul-smelling white discharge (fetid leukorrhea).*

circle corresponds to the duration of the fixation: the larger the circle, the longer the reader looked at that point in the text. The lines connecting the circles show saccadic movements between fixations, demonstrating the trajectory of the reader's gaze as the reader progresses through the text. In contrast, Figure 2 displays the reading of a simplified version of the clinical case, showing a more uniform spread of colours with fewer fixations. This map of visual attention allows us to determine which parts of the text are subjected to deeper cognitive processing, as indicated by the number and size of fixations.

3. Data and Experiment

For our experiment, we utilise a medical corpus in French that includes the CLEAR corpus (Grabar and Cardon, 2018) and a corpus with Clinical Cases (Grabar et al., 2020). These corpora encompass a variety of materials, such as articles from Wikipedia, reviews, leaflets, and medical cases.

Clinical cases are detailed accounts of the symptoms, diagnosis, treatment and follow-up of an individual patient or a small group of patients. Their content is close to clinical documents, such as discharge summaries (Grabar et al., 2020). Hence, clinical cases are rich in medical terminology. Additionally, they may include a discussion of the rationale for treatment choices, making them complex and rich in specialised information. Such texts are invaluable for medical education and practice, providing insight into the practical application of theoretical knowledge in real-life scenarios. In the context of our study, clinical case texts serve as a key component of the corpus, providing a deep immersion into medical scenarios that require significant cognitive effort of patients and their families to process and comprehend medical information relevant to their health and care.

Overall, for the eye-tracking experiments, we compiled a corpus of 16 texts. The texts are in French. This collection was thoughtfully divided into two distinct sets to ensure a balanced representation of text types in each.

Specifically, Set 1 comprised one assortment of texts, while Set 2 featured a different assortment, with each set containing an equal mix of medical texts, clinical cases, and ordinary texts to maintain a uniform distribution of text complexity and subject matter across both sets. We manually simplified these texts, resulting in two variations for each text within the sets: Version A and Version B. In Version A, some texts were presented in their original form, while others were simplified. Version B reversed this configuration, providing a mirrored counterpart to Version A in terms of which texts were simplified. Participants in the study were assigned to read texts from either Set 1 or Set 2, but not both, to ensure focused exposure to a specific subset of texts.

For the purpose of the work presented here, we chose 8 texts with the aim to cover the variety of the available medical text types (Table 1 indicates the size of these texts):

- two clinical case texts, chosen for their complexity and their specificity to the clinical context. Indeed, the patients and their families often have to face such documents during their healthcare process;

- two Wikipedia articles related to medical topics: autopsy and erythema. These texts illustrate medical information freely accessible to the general public. As observed by researchers and associations, general population look for medical and health information online increasingly frequently (Fox, 2014);

- two general-language texts about popcorn and quince. They are provided from Wikipedia as well. They are selected to illustrate general-language topics and provide some contrast in terms of content and lexical density by comparison with medical texts;

- the simplified versions of clinical cases. Since the content of clinical cases is too technical for common people, we manually simplified these clinical cases to make their reading more friendly. The simplification was done at lexical (lexical substitutions with synonyms, hyperonyms, hyponyms, definitions...), syntactic (sentence structure modification) and semantic (addition of contextual and semantic information) levels.

Table 1: Text size

Text Category	**Nb words**
Clinical Cases	534
Medical Encyclopedia Texts	1,594
General Encyclopedia Texts	1,545
Simplified Clinical Cases	630

As indicated, we created two sets of texts and, in each set, there is a version A and a version B :

- version *A* contains text 1 in its original form, text 2 in its simplified form, text 3 in its original form, text 4 in its simplified form,

- version *B* contains text 1 in its simplified form, text 2 in its original form, text 3 in its simplified form, text 4 in its original form.

Hence, each person reads all the texts from a given set in either original or simplified forms. Yet, in this study, we analyze only the simplified versions of clinical cases. As indicated above, due to the screen size limitations (Duchowski, 2007), the texts are divided into smaller segments.

The main purpose of this experiment is to collect eye-tracking indications on text reading. In addition, we also collect information on text understanding. For this, comprehension questions are asked after the reading of a given segment. The questions are related to the segment the participants just read. The possible answers to these questions are: *True*, *False*, or *I don't know*. In order to make the reading as natural as possible, the questions are asked only on some segments of text.

Participants read the texts using a Tobii Pro Spectrum eye-tracker, operating at 600 Hz.

For the experiment presented in this paper, we analyse the results from two groups of 5 participants each, totaling 10 participants. These individuals are French-speaking with French as first language, aged between 19 and 33 years, with no medical education, and coming from various social backgrounds, including students, PhD students, and full-time employees. Each group is tasked with reading texts of different types (clinical case, medical text, general text, and a simplified medical case), and sometimes answering the understanding questions.

In this preliminary analysis, we selectively focused on eye movement data from ten participants and specific text types to investigate characteristic patterns of eye movements during reading. The responses to comprehension questions from this subset of participants were not considered in the current analysis. This decision was made because the primary goal of this phase was to examine eye movement performance, and the limited sample size precludes a comprehensive analysis of text comprehension across the entire participant group based on their responses to questions. At this stage, we considered the comprehension questions mainly as active engagement with the text, thus ensuring that the recorded eye movements accurately reflect actual reading behaviour. It is important to note that this approach does not negate the value of comprehension data. As we move beyond this preliminary phase, we intend to conduct a more extensive analysis that includes eye movement characteristics in conjunction with comprehension responses for all texts and participants. This future analysis aims to offer deeper insights into how text comprehension correlates with specific eye movement patterns.

In the two following sections, we present the analysis principles and the results first for each type of texts individually (Section 4) and then across the types of texts making their comparison (Section 5). In addition, Section 6 concentrates on a qualitative analysis of words that require most readers' attention.

4. Statistical Analysis of Fixation Measures

To describe the fixation measures in different text types, and to infer the cognitive effort required to process the text content, we analyse the fixation metrics for four types of texts (clinical case, medical text, general text, and simplified clinical case). For each type of texts, the results are presented and discussed across three lines: general statistical analysis of fixation measures (Section 4); normality test (Section 4.2) to assess the normality of the data distribution and to define which further statistical tests can be applied; correlation of fixation measures (Section 4.3).

4.1. Collected Values of Fixation Measures

The collected average measures of the fixations are summarised in two tables: Table 2 details the average total duration of fixations, while Table 3 presents the average number of fixations for each text type. We indicate information on Mean values, the Standard deviation and Median values.

Table 2: Average Total Duration of Fixations Across Text Types in ms

Text Category	Mean	SD	Median
Clinical Case	395.89	328.25	307.8
Medical Text	359.29	239.37	321.8
General Text	323.78	239.95	271.9
S. Clinical Case	255.75	197.50	223.25

Table 3: Average Number of Fixations per Word

Text Category	Mean	SD	Median
Clinical Case	1.83	1.33	1.5
Medical Text	1.62	0.97	1.4
General Text	1.54	0.99	1.4
S. Clinical Case	1.27	0.87	1.2

The average total duration of fixations (Table 2) was the highest for clinical case texts, indicating these require more time to process, likely due to their complexity. In contrast, simplified clinical case texts showed the shortest average duration, suggesting that simplification effectively reduces cognitive load. Similarly, the average number of fixations (Table 3) followed a comparable pattern (clinical cases show the highest number of fixations indicating that this type of texts require more attention), further supporting the notion that text complexity influences reading behaviour.

The standard deviation for both fixation duration and number was notably higher in clinical case texts, underscoring a variability in complexity within this category. Indeed, clinical cases typically contain technical terms (*dysuria, hematuria, leukorrhea...*) aside more common words with medical meaning (*pathological, symptoms, hospitalization, urgency, burning...*) or not (*female, sensation, frequent, combined...*). This variability was less pronounced in the simplified texts, indicating a more consistent level of difficulty.

The two other types of text (medical and general language encyclopedia articles) show intermediate values. Yet, it is worth to notice that general language texts show high SD values.

4.2. Normality Test

We apply the normality Shapiro-Wilk test (Shapiro and Wilk, 1965) to assess the normality of the data distribution. Understanding whether our data conform to a normal distribution is necessary to determine the most appropriate further statistical tests.

After conducting the Shapiro-Wilk test, the data for both number and duration of fixations across all text categories yielded p-values very close to 0, decisively indicating their non-normal distributions. This result is consistent across the diverse text types examined (clinical case, encyclopedia medical text, general text, and simplified clinical case) highlighting significant variability in fixation metrics that could be attributed to differences in syntactic and lexical density. The anticipation of such variability, given the distinct characteristics of each text type, underscores the complexity of the cognitive processes involved in text reading and comprehension. We included the mean measure in our analysis to provide a central tendency of eye-tracking measures, revealing how text complexity affects reader engagement. Despite the non-normal distribution of the data, the mean values illustrate the general reading behaviour across different text types (Tables 2 and 3), highlighting longer and more frequent fixations on complex texts, such as clinical cases. The median offers a more accurate reflection of central tendency than the mean, as it is less influenced by extreme values. Therefore, focusing on the median provides a clearer understanding of the typical reader engagement and comprehension levels across different text types.

4.3. Correlation of Fixation Measures

After establishing that the fixation data, both duration and number, do not follow a normal distribution across the various text types, we next explore the relationship between these two metrics. Understanding the correlation between average total duration of fixations and average number of fixations can offer deeper insights into how text complexity influences reading behaviour. Given the non-normal distribution of our data, we employ Spearman's rank correlation coefficient (Spearman, 1904; Zar, 2005), a non-parametric measure that assesses the strength and direction of association between two ranked variables. This test is particularly suitable for our dataset given its observed non-normality.

The purpose is to verify whether a higher number of fixations correlates with longer total durations, suggesting more cognitive effort or processing time, in different types of texts. Such an analysis is crucial for understanding the nuances of reading patterns and how textual characteristics impact reader engagement and comprehension. Hence, employing this test, we analyse the correlation between average total duration of fixations and average number of fixations within each of the four text types.

Table 4: Spearman's Correlation Coefficients and P-values for Different Text Types

Text Type	Correl. Coeff.	P-value
Clinical Case	0.971	< 0.001
Medical Text	0.903	< 0.001
General Text	0.955	< 0.001
S. Clinical Case	0.968	< 0.001

Table 4 presents the Spearman's correlation coefficients and p-values for different text types, indicating a consistently strong positive correlation between the average total duration of fixations and the average number of fixations. These results suggest a robust relationship across all text types: increased fixation duration is associated with a higher number of fixations, reflecting varying levels of text complexity and cognitive engagement of readers.

5. Statistical Comparison between the Types of Texts

To determine whether significant differences exist across the text types in terms of fixation duration and frequency, we apply the Kruskal-Wallis test (Kruskal and Wallis, 1952). The analysis for average total duration of fixations yielded a test statistic of 85.137, with a corresponding p-value near 0.001. This indicates significant differences between the text types. Similarly, for average number of fixations, the test statistic was 55.191, with a p-value near 0.001, further confirming significant disparities between the text types.

Hence, we further apply the Dunn's post-hoc test (Dunn, 1961) to specify which text types differ significantly in fixation duration and frequency. This analysis facilitates pairwise comparisons between text types, elucidating the specific nature of the differences indicated by the Kruskal-Wallis test. We first present and discuss the findings for the average duration of fixations (Section 5.1) and then for the average number of fixations (Section 5.2).

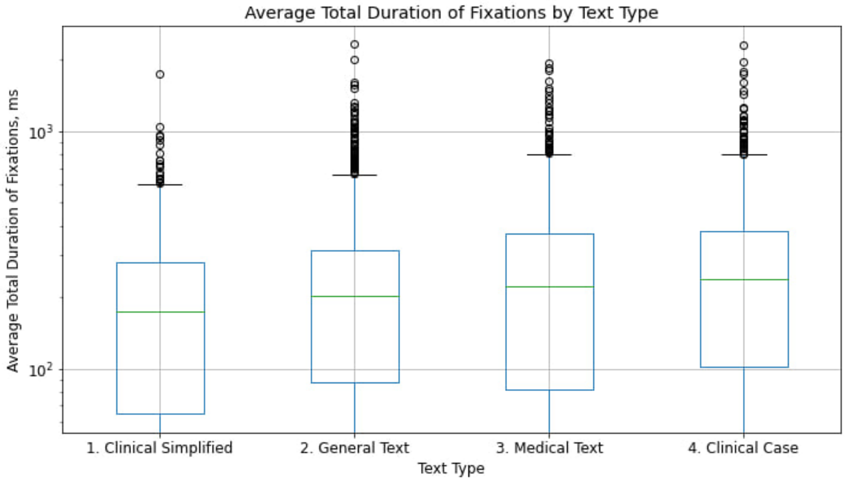

Figure 3: Boxplot of Average Total Duration of Fixations Across Text Types.

5.1. Average Total Duration of Fixations

Table 5: Post-hoc Dunn's Test Results for Average Total Duration of Fixations. The following text types are indicated: CC (clinical case), SCC (simplified clinical case), MT (medical text), and GT (general text).

Text Pairs for Comparison	**P-value**
CC vs. SCC	2.61×10^{-11}
CC vs. GT	0.009975
CC vs. MT	1.000
SCC vs. GT	1.84×10^{-6}
SCC vs. MT	2.49×10^{-17}
GT vs. MT	0.000035

As demonstrated in Table 5, the Post-hoc Dunn's test reveals a statistically significant variation in the average total duration of fixations across different types of texts. Specifically, a Post-hoc Dunn's test highlights a notable difference between clinical case texts and their simplified versions ($p = 2.61 \times 10^{-11}$), indicating that text simplification leads to a measurable reduction in cognitive load. This is further supported by comparisons between simplified clinical case texts against general ($p = 1.84 \times 10^{-6}$) and medical texts ($p = 2.49 \times 10^{-17}$), suggesting that simplified texts are read and comprehended more efficiently by readers.

Interestingly, no significant difference is found between clinical case and medical texts ($p = 1.0$), which suggests a comparable level of complexity from a cognitive load perspective. This observation is crucial for understanding the nuances of text engagement and comprehension, reinforcing the importance of text types for cognitive processing.

In Figure 3 the plot illustrates the distribution of average fixation durations across four categories of texts. The central line in each box represents the median duration, while the top and bottom edges of the box delineate the third and first quartiles, respectively. We can observe that the texts present an increasing difficulty starting with simplified clinical case, going through general-language text, then up to medical text and clinical case. From this boxplot, it is evident that the simplified clinical text has the lowest range of fixation durations as measured by eye-tracking characteristics, followed by general text. Medical text exhibits higher fixation durations, and clinical case texts have the highest, indicating a trend of increasing fixation duration across the complexity of the text types.

5.2. Average Number of Fixations

Table 6: Post-hoc Dunn's Test Results for Average Number of Fixations. The following text types are indicated: CC (clinical case), SCC (simplified clinical case), MT (medical text), and GT (general text).

Text Pairs for Comparison	**P-value**
CC vs. SCC	3.99×10^{-10}
CC vs. GT	0.01323
CC vs. MT	1.000
SCC vs. GT	1.38×10^{-5}
SCC vs. MT	2.85×10^{-10}
GT vs.MT	0.112714

Similarly, the average number of fixations across text types indicates significant disparities, reinforcing the insights gained from the analysis of fixation durations. The Post-hoc Dunn's test results (Table 6) demonstrate a significant reduction in fixations when comparing clinical case texts with their simplified counterparts ($p = 3.99 \times 10^{-10}$), underlining the effectiveness of simplification in enhancing text accessibility.

Furthermore, while general and medical texts exhibit no significant difference in the number of fixations ($p = 0.112714$), the distinct contrast with simplified clinical case texts emphasises the impact of simplification on reader engagement and cognitive effort. Taken together, these results highlight the potential of targeted text simplification to improve their comprehension and accessibility, especially for non-expert readers who are exposed to complex medical content.

6. Analysis of Words that Require Most Readers' Attention

We propose an analysis of words with the longest fixation durations and the highest number of fixations across the four types of texts (Tables 7 to 10). For 8 to 10 top-words, we indicate the average duration of fixations and the average number of fixations.

Table 7: Top Words by Average Total Duration (ms) and Number of Fixations in Clinical Case Texts

Words	Avg. Dur.	Avg. Nb
urétéro-hydronéphrose	2322.4	9.6
immunohistochimique	1954.2	9.2
impériosité	1780.8	6.0
vésicoacétabulaire	1748.2	7.2
pollakiurie	1617.6	5.6
leucorrhées	1486.0	5.4
latéro-trégonale	1428.0	6.8
cystoscopie	1267.2	5.4
extrapéritonisation	1242.0	5.0
47/48/52 mm	1164.4	5.0

Table 8: Top Words by Average Total Duration (ms) and Number of Fixations in Medical Texts

Words	Avg. Dur.	Avg. Nb
télangiectasie	1936.8	7.6
n°2011-525	1852.4	6.8
d'anatomo-pathologie	1820.0	7.4
«médico-hospitalière»	1637.8	6.4
anatomo-pathologiques	1512.8	6.6
1335-11	1487.0	4.6
ataxie-télangiectasie,	1423.8	6.4
spécialistes	1376.4	6.0
polypathologies	1316.4	4.6
scarlatiniformes	1256.0	5.2

In Tables 7 and 8, we can observe that words with the longest average total duration of fixations in medical texts are predominantly medical terms (*immunohistochimique, vésicoacétabulaire, ataxie-télangiectasie...*). Moreover, there is a noticeable overlap between words with the highest number of fixations and those with the longest fixation durations, indicating again a correlation between the complexity of medical terms and the cognitive effort required for their processing.

Table 9: Top Words by Average Total Duration in ms and Number of Fixations in General Texts

Words	Avg. Dur.	Avg. Nb
2-méthyle-2-butenoate	2337.6	9.8
proanthocyanidols	2017.8	8.6
anthocyanidols	1616.4	6.2
(sclérenchymateuses)	1576.0	5.8
Aromatnaya	1515.6	6.2
Balkans	1321.2	5.6
fabrique	1258.4	3.6
cuisson	1212.8	4.6

Table 9 demonstrates similar information for general-language text. Here, complex chemical names, such as *2-méthyle-2-butenoate* and *proanthocyanidols* (chemical substances contained in quince) attract the gaze of readers and show the highest fixation durations and frequencies when reading the general text. This suggests that these terms pose a cognitive challenge to readers, despite being part of general texts. They may also be the cause of the high standard deviation in this type of texts, where they neighbour common and frequent words (Section 4.1). Besides, the fixation data reveal that even non-medical, general texts may contain words that necessitate significant cognitive effort to be processed.

Table 10: Top Words by Average Total Duration (ms) and Number of Fixations in Simplified Clinical Case Texts

Words	Avg. Dur.	Avg. Nb
vésicoacétabulaire	1745.00	6.20
Staphylococcus	1052.40	4.80
(cystoscopie)	963.40	4.60
pollakiurie	956.25	3.5
(tomodensitométrie)	916.50	4.50
cystoscopie)	883.00	4.75
(intraveineuse)	810.00	3.60
d'orthopédie	759.40	3.80
(aminoside)	758.80	2.8
dysurie	727.75	3.50

Finally, Table 10 shows words from the simplified clinical cases, highlighting that medical terms, often enclosed in brackets and explained with simpler language, significantly reduce both the duration of fixations and the number of fixations per word when compared to the fixation values observed on original clinical case texts. This reduction indicates that simplification efforts effectively lower the cognitive load required to comprehend these medical terms and that the comprehension is helped by the context.

These findings collectively underscore the influence of text types on reader engagement and cognitive processing. Indeed, the proposed qualitative and quantitative analysis of eye-tracking measures indicates the complexity elements and zones within the analysed texts.

7. Conclusion and Future Work

In this work, we propose an experiment using eye-tracking technology with the purpose of revealing text complexity and the inherent cognitive load it presents to readers. Four text types are considered: clinical cases, encyclopedia medical texts, encyclopedia general texts, and manually simplified clinical cases. These texts are read by 10 people. The collected eye-tracking measures related to fixations are analysed. We perform quantitative and qualitative analysis. The normality statistical test shows that the fixation values do not have a normal distribution, which means that the reading difficulty is uneven. This also suggests the complexity of cognitive processed involved in text reading. Further quantitative analysis with different statistical tests indicates that the text types significantly impact the reading easiness, as testified by the fixation measures (duration of fixations and their number): in our experiment, (1) clinical cases are the most difficult to read, (2) they are followed by medical and general language encyclopedia articles, (3) while the simplified version of clinical cases eases a lot the reading process. The statistical analysis also indicates that duration of fixations and their number are correlated: complex words usually require longer fixations and their number is higher. Such words need a stringer cognitive effort. Interestingly, this quantitative analysis indicates that clinical cases present the highest difficulty, yet the simplification of clinical cases makes these texts much easier to read and comprehend. As for the qualitative analysis, we presented top-words which require the most attention from readers in each type of texts. These words usually correspond to technical medical terms.

These measures from eye-tracking records across various text types can be utilised to detect the complexity zones within these texts. Besides, such an eye-tracking annotation of texts can be used to train a language model, thus enabling the automatic prediction of reading patterns for texts of different types. This is one of our objectives for future. Other objectives are related to the collection of eye-tracking measures from more people and on more texts.

8. Ethical Considerations and Limitations

Participation in this study is voluntary, with informed consent obtained from all participants, ensuring compliance with the European General Data Protection Regulation (EU) 2016/679 and the modified French Data Protection Act of January 6, 1978. All personal data collected in the course of this research are anonymized to protect participant privacy and are accessible only by the designated project manager. This study has been registered in the University of Lille's registry under reference 2022-075, affirming our commitment to upholding the highest standards of data protection and participant rights.

9. Acknowledgements

This work was partly funded by the French National Agency for Research (ANR) as part of the *CLEAR* project (*Communication*, *Literacy*, *Education*, *Accessibility*, *Readability*), ANR-17-CE19-0016-01.

We thank the FR "Sciences et Cultures du visuel" for the facility to perform the eye-tracking experiments.

We also would like to thank the anonymous reviewers for their helpful comments.

10. Bibliographical References

References

Dag Aarsland, Lucia Batzu, Glenda Halliday, Gert Geurtsen, Clive Ballard, Kallol Ray Chaudhuri, and Daniel Weintraub. 2021. Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers*, 7:47.

Douglas Biber and E. Finegan. 1994. Intra-textual variation within medical research articles. *Corpus-based research into language*, 12:201–222.

Charles Clifton, Adrian Staub, and Keith Rayner. 2007. Eye movements in reading words and sentences. *Eye movements: A window on mind and brain*.

Andrew Duchowski. 2007. *Eye Tracking Methodology. Theory and practice*. Springer, London, UK.

Anna Ekstrand, Mattias Nilsson, and Gustaf Öqvist Seimyr. 2021. Screening for reading difficulties: Comparing eye tracking outcomes to neuropsychological assessments. *Frontiers in Education*, 6.

Susannah Fox. 2014. The social life of health information. Technical report, Pew Internet & American Life Project, Washington DC.

Glenn Fulcher. 1997. Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25:497–513.

Núria Gala, Thomas François, Ludivine Javourey-Drevet, and Johannes C Ziegler. 2018. La simplification de textes, une aide à l'apprentissage de la lecture. *Langue française*, 199(3):123–131.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. Eye tracking: A comprehensive guide to methods and measures.

Jukka Hyönä and Johanna K. Kaakinen. 2019. *Eye Movements During Reading*, pages 239–274. Springer International Publishing, Cham.

Ludivine Javourey Drevet, Stephane Dufau, Thomas François, Núria Gala, Jacques Ginestié, and Johannes Ziegler. 2022. Simplification of literary and scientific texts to improve reading fluency and comprehension in beginning readers of french. *Applied Psycholinguistics*, 43:1–28.

Marcel Adam Just and Patricia A. Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87 4:329–54.

Ooi K. and Khaidzir Ismail. 2011. The relationship between cognitive processing and reading. *Asian Social Science*, 7.

William H. Kruskal and W. Allen Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.

Aya Ozeri-Rotstain, Ifaat Shachaf, Rola Farah, and Tzipi Horowitz-Kraus. 2020. Relationship between eye-movement patterns, cognitive load, and reading ability in children with reading difficulties. *Journal of Psycholinguistic Research*, 49.

Mamata Pandey, Geoffrey Maina, Jonathan Amoyaw, Yiyan Li, Rejina Kamrul, C. Michaels, and Razawa Maroof. 2021. Impacts of english language proficiency on healthcare access, use, and outcomes among immigrants: a qualitative study. *BMC Health Services Research*, 21.

Alia Pugh, Devin Kearns, and Elfrieda Hiebert. 2023. Text types and their relation to efficacy in beginning reading interventions. *Reading Research Quarterly*, 58:710–732.

Keith Rayner, Kathryn Chace, Timothy Slattery, and Jane Ashby. 2006. Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading - SCI STUD READ*, 10.

Keith Rayner and Eyal Reingold. 2015. Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, 1:107–112.

Keith Rayner, Sara Sereno, Robin Morris, Anna Schmauder, and Charles Clifton. 1989. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes - LANG COGNITIVE PROCESS*, 4.

Abhinav Singh, Poojan Mehta, Samar Husain, and Rajkumar Rajakrishinan. 2016. Quantifying sentence complexity based on eye-tracking measures.

Judith Wylie, Jennifer Thomson, Paavo Leppanen, Rakefet Ackerman, Laura Kanniainen, and Tanja Prieler. 2018. *Chapter 3. Cognitive processes and digital reading*, pages 57–90.

Pierre Zweigenbaum, Pierre Jacquemart, Natalia Grabar, and Benoit Habert. 2001. Building a text corpus for representing the variety of medical language. In *MEDINFO*, pages 290–294.

11. Language Resource References

Olive Jean Dunn. 1961. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64.

N. Grabar, Clément Dalloux, and V. Claveau. 2020. Cas: corpus of clinical cases in french. *Journal of Biomedical Semantics*, 11.

Natalia Grabar and Rémi Cardon. 2018. CLEAR – simple corpus for medical French. In *Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA)*, pages 3–9, Tilburg, the Netherlands. Association for Computational Linguistics.

S. S. Shapiro and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.

C Spearman. 1904. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Jerrold H Zar. 2005. Spearman rank correlation. *Encyclopedia of Biostatistics*, 7.