Study of readability of health documents with eye-tracking approaches

Natalia Grabar
CNRS, UMR 8163, F-59000 Lille, France;
Univ. Lille, UMR 8163 - STL,
F-59000 Lille, France
natalia.grabar@univ-lille.fr

Emmanuel Farce, Laurent Sparrow
Univ. Lille, CNRS,
UMR 9193 - SCALab,
F-59000 Lille, France
emmanuel.farce@univ-lille.fr
laurent.sparrow@univ-lille.fr

Abstract

Medical area is an integral part of our lives due to health concerns, but the availability of medical information does not guarantee its correct understanding by patients. Several studies addressed this issue and pointed out real difficulties in the understanding of health contents by patients. We propose to use eye-tracking methods for studying this issue. For this, original technical and simplified versions of a deidentified clinical document are exploited. Eye-tracking methods permit to follow and to record the gaze of participants and to detect reading indicators such as duration of fixations, regressions and saccades. These indicators are correlated with answers to questionnaires submitted to participants after the reading. Our results indicate that there is statistically significant difference in reading and understanding of original and simplified versions of health documents. These results, in combination with another experiment, permit to propose a typology of medical words which need to be explained or simplified to non-expert readers.

1 Introduction

Medical area is an integral part of our lives due to health concerns and to presence of health information in media and literature. With the evolution of Internet, medical and health information is becoming widely available and accessible online. It has been noticed that, across the world, Internet is positioned at the second place where patients are searching for health information, while the first source of information is still occupied by consultations with medical doctors (Pletneva et al., 2011; Fox, 2011). According to these surveys, up to 24% of the population uses Internet at least once a day to find information on health issues and up to 80% of population is looking for health information on Internet in general. Yet, the availability of medical information does not guarantee its correct understanding by patients. Medical area conveys indeed very specific terminology, like *abdominoplasty, hepatic* or *metatarsophalangeal*. This fact has been stressed by several studies dedicated to poor understanding of health information (McCray, 2005; Patel et al., 2002; Williams et al., 1995; Berland et al., 2001) and to complicated communication between patients and medical staff (Jucks and Bromme, 2007; Tran et al., 2009).

Text complexity is studied in several disciplines, such as linguistics (Iacobini, 2003; Lüdeling et al., 2002), psychology (Bertram et al., 2011; Lüttmann et al., 2011; Bozic et al., 2007; Dohmes et al., 2004; Cain et al., 2009), and NLP (Natural Language Processing) with traditional (Flesch, 1948; Dale and Chall, 1948) and computational (Zeng et al., 2005; Chmielik and Grabar, 2011) approaches. The purpose of our work is to study further the understanding of health documents by non-expert people. We work with data in French. More particularly, we propose to address the reading and understanding of health information through methods and tools provided by eye-tracking. Indeed, study of eye movements during the reading is indicative about the cognitive processes involved. More particularly, text difficulty and readability can be measured with several indicators (Duchowski, 2007; Rayner, 1998; Sparrow et al., 2003; Miellet et al., 2008). Among the eye-tracking indicators, we can mention the following, which are the most exploited in the existing work:

- *Saccades* are rapid movements of eyes during the reading to go from one spot in the

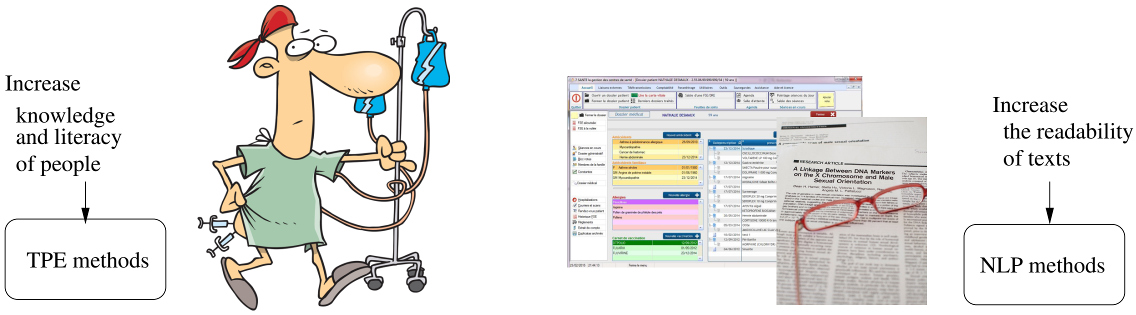

Figure 1: Two aspects related to the understanding of health documents: health literacy of people and readability of documents.

text to another. When the text is easy to read and understand, saccades are longer, and they become shorter when the text is complicated because readers need more time for reading;

- *Fixations* are periods during which the eyes are stable. Fixations correspond to moments when visual information is analyzed. Duration of fixations is increased when the texts are difficult because such texts require longer time for the assimilation of information. Correspondingly, the duration of fixations decreases when the text is easy to read and understand;

- *Regressions* occur when the reader goes back to the text spans already read. When the text is difficult it usually requires more regressions.

Hence, the comparison of eye-tracking parameters (duration of fixations, length of saccades, regressions, etc.) recorded during the reading of texts permits to evaluate with more precision difficulties and blocking points of readers.

According to our hypotheses, reading of complex texts and unknown terms condition our attention and the movements of our eyes present then typical and observable patterns. Such indicators can be directly correlated with difficulties occurring during the text reading and understanding: for instance, when a text contains technical terms, the reading speed and fixations are longer, and require more time for the assimilation. We propose to exploit the findings from the eye-tracking experiments for the detection of blocking reading points in medical texts and for providing a typology of medical words for which reading and understanding may present some difficulties. This typology is also supported by another set of experiments performed with medical terms in French. One issue is that these medical words and terms should be simplified or explained to laypeople and patients for making better their understanding of the medical contents.

Eye-tracking and its indicators are exploited in several contexts for the detection of text spans that attract or block the eye movements and the reading, such as: relation between speech and eye movements, when participants are looking at picture segments which correspond to the sentences they are hearing (Cooper, 1974; Tanenhaus et al., 1995; Wendt et al., 2014); reading of texts in first and second languages (Altarriba et al., 1996; Bisson et al., 2014); reading of texts by dyslexic people (Rubino and Minden, 1973; Elterman et al., 1980; Nilsson Benfatto et al., 2016) and autists (Yaneva et al., 2015); processing of syntactic structures (Frenck-Mestre and Pynte, 1997; Clifton and Staub, 2011; Trueswell et al., 1994; Singh et al., 2016); detection and processing of errors (Keating, 2009); evaluation of text complexity during the translation (Sharmin et al., 2008) and language acquisition (Balakrishna, 2015).

Usually, in relation with understanding of texts, two closely related aspects are distinguished (Figure 1):

1. On one side, patients have a given level of literacy and of health literacy, when situated in the health area. This aspect is researched by Therapeutic Patient Education (TPE), which purpose is to diagnose and to improve the health literacy of patients (Golay et al., 2007; Pélicand et al., 2009; Glasgow et al., 2012). Such actions are usually done by experts in patient education (specialized medical doctors, speech therapists, sociologists, psychologists, nurses...);

EXAM: SONOGRAPHY OF HANDS AND FEET
REASON: Arthralgia
Hands: Tenosynovitis and arthrosynovitis cannot be observed.
Forefoot: Interesting reorganization of the first metatarsophalangeal can be seen, in relation with the history of surgery of hallux valgus.
Absence of arthrosynovitis at the level of metatarsophalangeal articulations.

EXAM: SONOGRAPHY OF HANDS AND FEET
REASON: Pain in articulations
Hands: Inflammation of tendons or of articulation membrane cannot be observed.
Forefoot: Interesting reorganization of the first foot articulations can be seen, in relation with the history of surgery of foot deformation.
Absence of inflammation of the membrane at the level of foot articulations.

Figure 2: Translated examples with original (upper) and simplified (lower) texts.

2. On the other side, health documents show a given readability level and can be more or less difficult to read and to understand. Here, the purpose is to diagnose the difficulty of information and to make this information more easily accessible for laypeople. Typically, this process is addressed by researchers in NLP for the readability diagnosis and for the text simplification (Biran et al., 2011; Brouwers et al., 2012; Glavas and Stajner, 2015).

Our work is related to the second aspect: diagnosis of text readability.

In what follows, we first present the material used (Section 2) and the protocol of the approach (Section 3) to reach the objectives. Section 4 is dedicated to the description and discussion of the results obtained, and Section 5 draws the conclusion and proposes some issues for the future work. All experiments are performed with the French-language data.

2 Material

Two short excerpts of deidentified clinical documents are used: summary discharge in cardiology and radiology report of feet and hands. These texts are used in two versions: original (technical) and manually simplified (see Figure 2). Due to the experimental setting of eye-tracking experiments, the texts used are short: 48 words in $text_1$ and 112 words $text_2$. For the simplification, we use automatically built resources (Grabar and Hamon, 2014; Antoine and Grabar, 2016), which provide pairs of equivalent terms such as {*myocard*; *heart muscle*}, {*desmorrhexy*; *rupture of ligaments*}, and pairs of hyperonyms such as {*metatarsophalangeal→foot*}. Synonyms and paraphrases are used in priority, and hyperonyms are used when synonyms and paraphrases are not available. The simplification is typically done for words and terms which have been judged as non-understandable in previous research, for which almost 30,000 medical words from the UMLS (Lindberg et al., 1993) and Snomed International (Côté, 1996) terms have been manually categorized as understandable or non-understandable (Grabar and Hamon, 2016). Overall, the $text_1$ has undergone seven modifications, and the $text_2$ ten modifications. After the simplification, $text_1$ contains 65 words and $text_2$ 82 words. As a matter of fact, $text_1$ has become longer because its original version contains several compoundings which simplification requires paraphrasing with several words.

These texts are used to build two testsets, in which the order of technical and simplified texts varies:

- $testset_1$: original $text_1$ and simplified $text_2$,
- $testset_2$: simplified $text_1$ and original $text_2$.

Figure 2 presents the English translation of the $text_1$ in the original and simplified versions.

3 Approach

We first describe the inclusion criteria of this study, and then the protocol of the eye-tracking experiments, and the analysis of the obtained data (Sections 3.1 to 3.3).

3.1 Inclusion Criteria

50 participants are recruited and each testset is read by 25 of them, so that statistical significance

between original and simplified versions can be computed. Can be included in the study:

1. undergraduate students from different disciplines (psychology, linguistics, history, communication studies...). Medical and paramedical students are excluded. Usually, 5 levels of literacy are distinguished (Bernèche and Perron, 2006):
   - levels 1 and 2 correspond to persons who have serious difficulties in reading, understanding and assimilation of information;
   - level 3 gathers people who usually have standard readability and literacy level. They are fluent in reading and can understand general language purposes;
   - levels 4 and 5 correspond to persons who show the capacity to read, understand and make complex deductions, which is often specific to persons with high school education.

   Undergraduate students are usually associated with the third level of literacy, and are representative of the average citizens (Bernèche and Perron, 2006);
2. people without chronic disorders because in that case they may be familiar with medical terminology;
3. people without dyslexia because they have difficulties with reading, which are not specifically due to the reading of specialized texts, such as aimed in our study;
4. people with French as first language, which provide the common basis for all participants and guarantees that difficulties in reading and understanding are not due to other causes than specificity of the medical field.

3.2 Eye-tracking Protocol

The proposed approach is based on exploitation of eye-tracking, which purpose is to measure the fluidity of reading with objective measures like the number of saccades, the duration of saccades, the number of fixations, the duration of the first fixation, or the regressions (Sparrow et al., 2003; Miellet et al., 2008). These indicators typically permit to detect text zones which obstruct the reading and the understanding, as the two of them are related. The texts are presented on a display, and specific camera (EyeLink 1000) permits to capture eye movements and to relate them with the text.

> *The heart is supplied in blood by coronary arteries which are fed by another artery: the aorta. When the diameter of coronary arteries is reduced because of progressive formation of fat patches, cardiac muscle is no more supplied in oxygen and nutrients: it is suffering. If the artery is blocked completely, infarctus may be close... Bypass and stent have the same purpose: restore normal blood flow.*

Figure 3: Translation of the control text (step 4).

After the presentation of the objectives of the study, each participant goes though:

1. parameterizing of the eye-tracking camera,
2. reading of a general text for training,
3. reading of the testset$_1$ or of the testset$_2$, with medical texts in original and simplified versions,
4. reading of the control text with lay medical contents (Figure 3).
5. After the reading of each text, the participant has to answer multiple choice questionnaires (two questions per text) to control the understanding of these texts. On the text$_1$, these two questions are asked:
   - *The sonography is done for: (1) shoulder, (2) hands and feet, (3) I do not know*
   - *Which inflammations are looked for: (1) articulations only, (2) articulations and tendons, (3) I do not know*

   On the text$_2$, these two questions are asked:
   - *The patient has problems: (1) cardiac, (2) cerebral, (3) I do not know*
   - *The patient is treated with: (1) surgery, (2) genetically, (3) I do not know*

   On the control text, these two questions are asked:
   - *The arteries can be damaged with: (1) fat patches, (2) calcium patches, (3) I do not know*

	Text$_1$						*Text*$_2$					
	O	*S*	*std.*	*p*	*dof*	*t-test*	*O*	*S*	*std.*	*p*	*dof*	*t-test*
TRN	60,55	63,63	-3,08	0,23	45,00	1,22	62,73	59,67	3,06	0,22	45,00	1,24
CRL	58,88	62,06	-3,19	0,22	45,00	1,25	61,04	57,84	3,20	0,21	45,00	1,29
DFF	227,41	215,75	11,66	0,11	45,00	1,65	214,73	214,69	0,04	0,50	45,00	0,68
TNF	587,61	370,48	217,14	**0,00**	45,00	7,38	395,71	372,22	23,49	0,16	45,00	1,43
AMP	3,50	3,80	-0,30	**0,02**	45,00	2,44	3,33	3,82	-0,49	**0,00**	45,00	5,38
REG	27,26	21,21	6,06	**0,05**	45,00	2,05	21,47	19,30	2,18	0,24	45,00	1,18
MCQ	1304,35	869,57	434,78	**0,02**	21,00	2,08	602,77	538,95	63,82	**0,00**	21,00	2,08

Table 1: Results for the two versions of the texts (original O and simplified S) and their statistical analysis. The indicators are the following: training text *TRN* and control text *CRL*; duration of the first fixation *DFF*, total number of fixations *TNF*, amplitude of saccades *AMP*, number of regressions *REG*; answers to questions *MCQ*. Statistically significant p is marked with bold characters.

- *When the artery is blocked, there is risk of: (1) headaches, (2) infarctus, (3) I do not know*

6. At the end, if desired, the results recorded further to his reading are presented and explained to the participant.

Overall the experiment lasts for 15 to 20 minutes.

3.3 Analysis of Eye-tracking Data

The data collected during the eye-tracking experiments are analyzed with several statistical measures, such as *t test*, statistical significance, degree of freedom (Walker, 1940) and standard deviation. The objective is to assess the difference of indicators when reading original and simplified versions of texts. We expect that the simplified version of texts is read more easily, *e.g.* with short fixations, long saccades and no regressions.

4 Results and Discussion

We first present the results obtained from the presented eye-tracking experience (Section 4.1) and discuss them, we then indicate some advantages of using the eye-tracking methodology (Section 4.2) and some known limitations (Section 4.3), and propose a typology of words that may present reading difficulties in medical texts (Section 4.4).

4.1 Results from the Eye-tracking Protocol

In Figure 4, we present an example of reading of the text$_1$: the original (upper graphics) and simplified (lower graphics) versions. In Figure 5, we present another example obtained while reading the text$_2$ in original and simplified versions. On these Figures, we can easily observe differences in reading of original and simplified versions.

In Table 1, we present the average reading indicators for the two tested versions (original O and simplified S) and their statistical analysis: the reading time for the original O and the simplified S versions, the standard deviation $sdt.$, the p-value p, the degree of freedom dof, and the $t-test$ value. These results are provided for the text$_1$ and the text$_2$, and for each indicator studied. The results indicate that:

- *Reading time of the training (TRN) and control (CRL) texts.* No statistical difference is observed with the reading time of the training and control texts. This indicates that the participants have the same reading capacity and that their reading results on medical texts are comparable. This is a good observation which points out that further results are comparable;

- *Duration of the first fixation (DFF).* No difference is observed for the duration of the first fixation. This indicates that reading of the two versions of texts starts in a similar way, that the participants do not anticipate on the nature of the texts (original or simplified), and again that further results are comparable;

- *Total number of fixations (TNF).* Statistically significant difference is observed for the total duration of fixations for the text$_1$: on the original version the fixations are more frequent. This can be observed on Figure 4: on the original (upper) text, the blue dots are more frequent than on the simplified (lower)

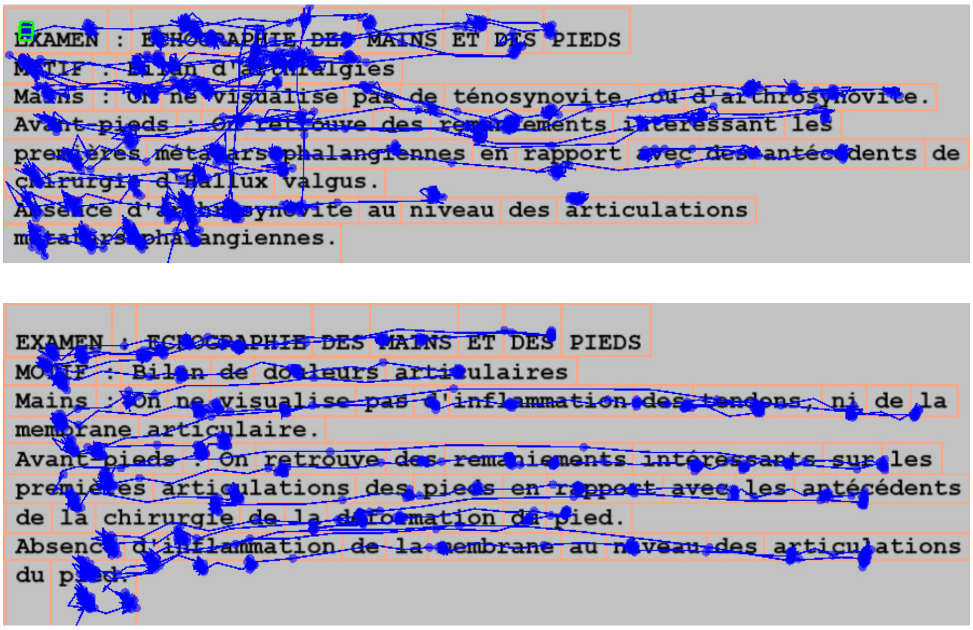

Figure 4: Examples of reading of original (upper) and simplified (lower) versions of $texts_1$.

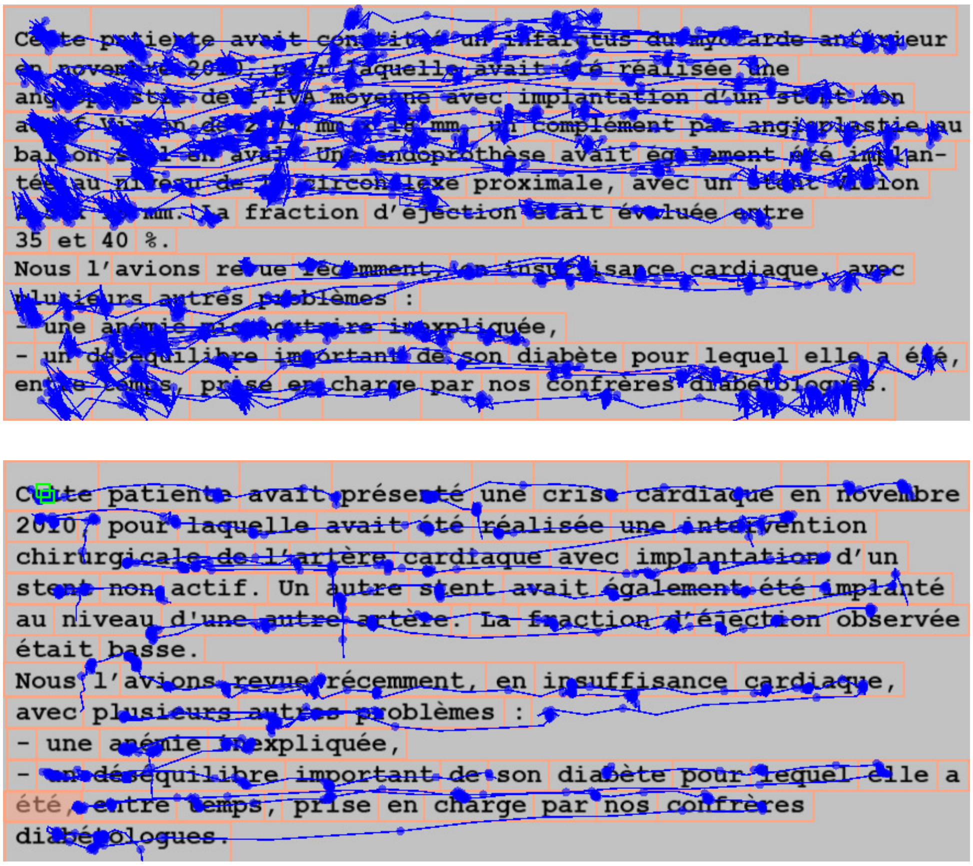

Figure 5: Examples of reading of original (upper) and simplified (lower) versions of $texts_2$.

text. For instance, in the original text, compound words like *arthralgia*, *arthrosynovitis* or *metatarsophalangeal* can require several fixations, which may correspond to the syllables of these words. Besides, this kind of terms also show longer fixations by the participants (the dots are larger). The technical version of the $text_2$ does not require that many fixations, may be due to the fact that it does not contain compounds;

- *Amplitude of saccades (AMP).* The simplification of the texts causes the increasing of the amplitude of saccades. As indicated in Table 1, this indicator is statistically significant for the $text_1$ and the $text_2$. This means that simplification decreases the reading difficulty. Hence, on Figures 4 and 5, the horizontal blue lines are shorter on the original texts than on the simplified texts;

- *Regressions (REG).* The number of regressions is statistically important for the $text_1$ but not for the $text_2$. This suggests that the reading of the simplified version of the $text_1$ is more fluent. This can also be observed on Figure 4: on the original text, we can see vertical blue lines;

- *Answers to the questionnaires (MCQ).* The analysis of the answers to questions indicates that the understanding of the simplified version is always better for the simplified versions than for the technical versions. The difference is statistically significant for the two texts. Hence, we obtain 54% of correct answers for the original versions and up to 81% of correct answers for the simplified versions.

On the whole, we can observe that the simplification of text improves all the reading indicators: (1) the total duration of fixations is lesser, (2) the amplitude of saccades is bigger, and (3) the regressions are less frequent. These results provide coherent and stable reading indicators and reading patterns specific to the technical original and the simplified versions of the medical texts. These results also indicate that the simplification of health documents is an efficient way to improve their reading and understanding by non-expert readers. As indicated all indicators show statistically significant differences on the $text_1$ and some of them are also statistically significant on the $text_2$. On both texts, correct answers to questions are correlated with the text difficulty.

4.2 Advantages of Eye-tracking

The eye-tracking technology offers several advantages which can be exploited in different tasks, such as those presented in Section 1. We present here some of these advantages, which have been very useful in our experiments:

- Several indicators on the reading process can be computed and exploited. Typically, these indicators are: the number and duration of fixations, the amplitude of saccades, the number of regressions;

- The eye-tracking indicators are objective: they are non-conscious and non-controlled by people. They are directly impacted by the individual reading habits acquired during the scholar and family learning;

- The eye-tracking indicators can be correlated with other types of information such as understanding of texts, social and professional status of participants, etc.;

- During the reading, the words and terms are considered within their contexts and the global perception of texts is usually expected from participants.

4.3 Limitations of Eye-tracking

Yet, the eye-tracking technology presents some known limitations, which are usually taken into account in experiments (Duchowski, 2007):

- Eye-tracking camera permits to detect and to record the gaze of the participants. The common hypothesis is that the gaze is correlated with the attention of participants, while in reality attention can also be oriented on objects which are located on peripheral areas of the gaze. Human vision system is indeed very sophisticated and currently it is not fully decoded yet. This is one of the known limitations of the eye-tracking methodology and it requires that the two possibilities are accepted: the gaze matches with the explicit attention of participants or does not match. In our case, with the reading of short medical texts, the requirement to answer questions after the reading, and the absence of distractors (the tests have been performed in experimental lab conditions), we assume that the attention of readers matches with their gaze;

- With some participants, due to their physiological specificities (long eyelashes, make-up, heavy eyelids...), it can be complicated to parameter the eye-tracking camera, and to track and record the eye movements. This can lead to loss of data or to wrong superposition of gaze recording on the texts. Nevertheless, when the data are exploitable, there is no impact on the reading indicators;

- Similarly, eyeglasses and contact lenses can be problematic for the tracking of the pupil and of its movements;

- For a given text or picture, the attention and the gaze of participants vary according to the task and the questions they are being asked. In our experiment, all the participants had to do the same task which consisted in text reading and answering to questions. The instructions have been presented clearly at the beginning of the test and before each reading;

- Eye-tracking cameras also have some limitations: (1) they work with a given frequency (60 Hz) and some eye movements can be missed and not recorded; (2) the recorded signal is cleaned up, such as with blinking or some peripheral eye movements, which can also remove some important eye movements;

- Due to the test requirements, the tests can be performed only with short texts which can be easily displayed and read by all participants from a computer screen. This means that several tests and experiments are necessary to cover more texts and to increase their diversity.

4.4 Typology of difficulties

The results obtained from the presented experiments permit to propose a typology of some medical words and terms that may present reading and understanding difficulties to laypeople. Notice that this typology is confirmed and completed by larger experiments done with medical terminologies in French: almost 30,000 medical words from the UMLS (Lindberg et al., 1993) and Snomed International (Côté, 1996) in French. These terms have been manually categorized as understandable or non-understandable (Grabar and Hamon, 2016).

The complete proposed typology contains the following types of linguistic units:

- abbreviations (*IVA, NIHSS, OAP, NaCl, VNI, OG, VG, PAPS, j, bat, cp*);

- borrowings from Latin or English (*stent, Hallux valgus*);

- proper names (*Gougerot, Sjögren, Bentall, Glasgow, Babinski, Barthel, Cockcroft*);

- drug names (*CALCIPARINE*);

- neoclassical compounds meaning disorders, procedures or treatments (*endoprothesis, pseudohémophilie, sclérodermie, hydrolase, tympanectomie, arthrodèse, synesthésie*);

- human anatomy (*metatarsophalangeal, microcytic, cloacal, pubovaginal, nasopharyngé, mitral, antre, inguinal, strontium, érythème, maxillo-facial, mésentère*);

- lab test results with numeral values and their interpretation.

Such units are very frequent in different types of medical texts and potentially present an important understanding difficulty. We assume that such words and terms must be explained or simplified to laypeople to guarantee a more correct understanding of medical texts by them. This task can be typically addressed during the automatic text simplification or adaptation.

Due to the experimental set-up, only two short excerpts from medical texts have been used (160 words in technical versions and 147 words in simplified versions, in total). Currently, it is difficult to link the typology classes to the eye-tracking indicators. Nevertheless, we can present here some first observations:

- *Abbreviations*. The $text_2$ contains one abbreviation (*IVA*), which required longer fixations of the participants;

- *Borrowings*. The $text_1$ contains one borrowing from Latin (*Hallux valgus*), which was read normally by participants. One possible explanation collected from participants is that uppercased *H* in *Hallux valgus* associated this term with a proper name;

- *Proper names*. No real proper names occur in the two texts;
- *Drug names*. No drug names occur in the two texts;
- *Neoclassical compounds*. The tested texts, and especially the text_1, contain several compounds. As already indicated above, compounds require several fixations and these fixations are longer. Compounds may also require regressions;
- *Human anatomy*. Several terms related to human anatomy occur in the two texts. Excepting very frequent terms (like *foot* or *hand*), human anatomy terms usually require several fixations and these fixations are longer;
- *Numeral values*. The text_2 contains several numerical values. These values require longer fixations and also regressions.

These are just first raw observations obtained from two small medical texts in French.

5 Conclusion and Future Work

We proposed an experiment on studying the effect of simplification of medical texts addressed through the use of eye-tracking methods. In this way, we can obtain several objective reading indicators, such as duration of fixations, amplitude of saccades and regressions. The collected indicators are then compared between the original and simplified versions of a given text with statistical measures to analyze if there is statistically significant differences when reading technical and simplified medical contents. Then, two understanding questions (multiple choice questionnaires) are asked to the participants after the reading of each text.

The results obtained indicate that reading of the two versions of the texts, original and simplified, provide coherent and stable reading patterns. For instance, when reading the simplified version, the fixations are shorter, the saccades are longer and the regressions absent or infrequent. Additionally, the analysis of the answers to questions indicates that the understanding of the simplified version is better: the number of correct answers varies between 54% for the original text and up to 81% for the simplified text. This also indicates that medical texts can be efficiently simplified in order to obtain their better understanding by non-expert persons.

These tests, together with data obtained from previous experiments, also permitted to propose a typology of medical words and terms that may present blocking points and difficulties with understanding. This typology include abbreviations, borrowed words, proper names, drug names, compounds, terms related to human anatomy and numbers. We assume that these kinds of terms should be simplified for a better understanding of medical texts by patients. Our first results permitted to associate some of these classes with eye-tracking indicators. For instance, compounds require more fixations and these fixations are longer. They may also require regressions.

We have several directions for future research. For instance, it would be interesting to study the relation between the text length and its readability and understanding. The hypothesis is that longer texts, even if they are simpler, may yet present reading and understanding difficulties. The impact of other factors (such as definitions, favorable contexts, pictures) can also be studied. Due to the experimental constraints, only short excepts of texts are used. For this reasons, it may be interesting to perform additional tests with a greater variety of text types and of simplification versions. Terms related to the proposed typology will be addressed in other works in order to perform their automatic explanation or simplification. In order to address different levels of literacy, different principles may be used when performing the simplification. These principes and the corresponding rules will be defined and tested in future work. Besides, like with manual simplification, the efficiency of automatic simplification methods can also be tested and evaluated using eye-tracking protocols.

6 Acknowledgements

This work was 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. The excerpts of the texts used have been obtained during the RAVEL project funded by the French National Agency for Research (ANR) and the DGA, under grant Tecsan ANR-11-TECS-012.

We would like to thank again the participants in the eye-tracking study. Also, we would like to thank the anonymous reviewers for their helpful comments.

References

J Altarriba, J Kroll, A Sholl, and K Rayner. 1996. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition*, 24:477–92.

E Antoine and N Grabar. 2016. Exploitation de reformulations pour l'acquisition d'un vocabulaire expert/non expert. In *Traitement Automatique des Langues Naturelles (TALN)*.

SV Balakrishna. 2015. *Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications*. Thèse de doctorat, Eberhard Karls Universität Tübingen, Tübingen, Germany.

GK Berland, MN Elliott, LS Morales, JI Algazy, RL Kravitz, MS Broder, DE Kanouse, JA Munoz, JA Puyol, and M Lara et al. 2001. Health information on the Internet. Accessibility, quality, and readability in english ans spanish. *JAMA*, 285(20):2612–2621.

Francine Bernèche and Bertrand Perron. 2006. Développer nos compétences en littératie: un défi porteur d'avenir. Enquête internationale sur l'alphabétisation et les compétences des adultes. Technical report, Institut de la statistique du Québec, Canada.

Raymond Bertram, Victor Kuperman, Harald R Baayen, and Jukka Hyönä. 2011. The hyphen as a segmentation cue in triconstituent compound processing: Its getting better all the time. *Scandinavian Journal of Psychology*, 52(6):530–544.

Or Biran, Samuel Brody, and Noémie Elhadad. 2011. Putting it simply: a context-aware approach to lexical simplification. In *Annual Meeting of the Association for Computational Linguistics*.

MJ Bisson, W Van Heuven, K Conklin, and R Tunney. 2014. Processing of native and foreign language subtitles in films: An eye-tracking study. *Applied Psycholinguistics*, 35:399–418.

Mirjana Bozic, William D. Marslen-Wilson, Emmanuel A. Stamatakis, Matthew H. Davis, and Lorraine K. Tyler. 2007. Differentiating morphology, form, and meaning: Neural correlates of morphological complexity. *Journal of Cognitive Neuroscience*, 19(9):1464–1475.

Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat, and Thomas François. 2012. Simplification syntaxique de phrases pour le français. In *Traitement Automatique des Langues Naturelles (TALN)*, pages 211–224.

Kate Cain, Andrea S. Towse, and Rachael S. Knight. 2009. The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102(3):280–298.

J Chmielik and N Grabar. 2011. Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques. *TAL*, 51(2):151–179.

C Clifton and A Staub. 2011. Syntactic influences on eye movements in reading. In S Liversedge, I Gilchrist, and S Everling, editors, *The Oxford handbook of eye movements*, pages 895–909. Oxford University Press.

RM Cooper. 1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychol*, 6:84–107.

RA Côté. 1996. *Répertoire d'anatomopathologie de la SNOMED internationale, v3.4*. Université de Sherbrooke, Sherbrooke, Québec.

E Dale and JS Chall. 1948. A formula for predicting readability. *Educational research bulletin*, 27:11–20.

Petra Dohmes, Pienie Zwitserlood, and Jens Bölte. 2004. The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language*, 90(1-3):203–212.

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

RD Elterman, LA Abel, RB Daroff, LF DellOsso, and JL Bornstein. 1980. Eye movement patterns in dyslexic children. *J Learn Disabil*, 13:16–21.

R Flesch. 1948. A new readability yardstick. *Journ Appl Psychol*, 23:221–233.

Susannah Fox. 2011. Health topics. 80% of Internet users look for health information online. Technical report, Pew Internet & American Life Project, Washington DC.

C Frenck-Mestre and J Pynte. 1997. Syntactic ambiguity resolution while reading in a second and native languages. *The Quarterly Journal of Experimental Psychology*, 50(1):119–48.

Russell E. Glasgow, Deanna Kurz, Diane King, Jennifer M. Dickman, Andrew J. Faber, Eve Halterman, Tim Woolley, Deborah J. Toobert, and Lisa A. Strycker et al. 2012. Twelve-month outcomes of an Internet-based diabetes self-management support program. *Patient Education and Communication*, 87(1):81–92.

Goran Glavas and Sanja Stajner. 2015. Simplifying lexical simplification: Do we need simplified corpora? In *ACL-COLING*, pages 63–68.

A. Golay, G. Lagger, and A. Giordan. 2007. Motivating patient with chronic diseases. *Journ of Med and the Person*, 5(2):57–63.

Natalia Grabar and Thierry Hamon. 2014. Automatic extraction of layman names for technical medical terms. In *ICHI 2014*, Pavia, Italy.

Natalia Grabar and Thierry Hamon. 2016. A large rated lexicon with French medical words. In *LREC (Language Resources and Evaluation Conference)*, pages 1–12.

C Iacobini. 2003. Composizione con elementi neoclassici. In Maria Grossmann and Franz Rainer, editors, *La formazione delle parole in italiano*, pages 69–96. Walter de Gruyter.

R Jucks and R Bromme. 2007. Choice of words in doctor-patient communication: an analysis of health-related internet sites. *Health Commun*, 21(3):267–77.

G Keating. 2009. Sensitivity to violations of gender agreement in native and non-native Spanish: An eye-movement investigation. *Language Learning*, 59:503–35.

DA Lindberg, BL Humphreys, and AT McCray. 1993. The Unified Medical Language System. *Methods Inf Med*, 32(4):281–291.

A Lüdeling, T Schmidt, and S Kiokpasoglou. 2002. Neoclassical word formation in German. *Yearbook of Morphology*, pages 253–283.

Heidi Lüttmann, Pienie Zwitserlood, and Jens Bölte. 2011. Sharing morphemes without sharing meaning: Production and comprehension of German verbs in the context of morphological relatives. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie exprimentale*, 65(3):173–191.

A McCray. 2005. Promoting health literacy. *J of Am Med Infor Ass*, 12:152–163.

S Miellet, L Sparrow, and S Sereno. 2008. The effects of frequency and predictability in French: An evaluation of the E-Z Reader model. *Psychonomic Bulletin & Review*, 14:762–769.

M Nilsson Benfatto, G qvist Seimyr, J Ygge, T Pansell, A Rydberg, and C Jacobson. 2016. Screening for dyslexia using eye tracking during reading. *PLoS ONE*, 11(12):e0165508.

V Patel, T Branch, and J Arocha. 2002. Errors in interpreting quantities as procedures : The case of pharmaceutical labels. *Int Journ Med Inform*, 65(3):193–211.

J Pélicand, C Fournier, and I Aujoulat. 2009. Observance, auto-soin(s), empowerment, autonomie: quatre termes pour questionner les enjeux de l'éducation du patient dans la relation de soins. *ADSP*, 66:21–23.

N Pletneva, A Vargas, and C Boyer. 2011. How do general public search online health information? Technical report, Health On the Net Foundation.

K Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372–373.

CA Rubino and HA Minden. 1973. Analysis of eye-movements in children with reading disability. *Cortex*, 9:217–220.

S Sharmin, O Spakov, KJ Rih, and AL Jakobsen. 2008. Effects of time pressure and text complexity on translators fixations. In *ETNA*, pages 123–126.

AD Singh, P Mehta, S Husain, and R Rajkumar. 2016. Quantifying sentence complexity based on eye-tracking measures. In *Workshop on Computational Linguistics for Linguistic Complexity*, pages 202–212.

Laurent Sparrow, S Miellet, and Yann Coello. 2003. The effects of frequency and predictability on eye fixations in reading: An evaluation of the E-Z reader model. *Behavioral and Brain Sciences*, 26:503–505.

MK Tanenhaus, MJ Spivey-Knowiton, KM Eberharda, and JC Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1634.

TM Tran, H Chekroud, P Thiery, and A Julienne. 2009. Internet et soins : un tiers invisible dans la relation médecine/patient ? *Ethica Clinica*, 53:34–43.

J Trueswell, M Tanenhaus, and S Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33:285–318.

HM Walker. 1940. Degrees of freedom. *Journal of Educational Psychology*, 31(4):253–269.

D Wendt, T Brand, and B Kollmeier. 2014. An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities. *PLoS ONE*, 9(6):e100186.

MV Williams, RM Parker, DW Baker, NS Parikh, K Pitkin, WC Coates, and JR Nurss. 1995. Inadequate functional health literacy among patients at two public hospitals. *JAMA*, 274(21):1677–1682.

V Yaneva, I Temnikova, and R Mitkov. 2015. Accessible texts for autism: An eye-tracking study. In *Int ACM SIGACCESS Conference on Computers & Accessibility*, pages 49–57.

Qing T Zeng, Tony Tse, Jon Crowell, Guy Divita, Laura Roth, and Allen C Browne. 2005. Identifying consumer-friendly display (CFD) names for health concepts. In *Ann Symp Am Med Inform Assoc (AMIA)*, pages 859–63.