

Fine-grained text simplification in French: steps towards a better grammaticality

Anaïs Koptient¹, Natalia Grabar¹

¹CNRS, UMR 8163, Univ. Lille, UMR 8163 - STL – Savoirs Textes Langage, F-59000 Lille, France; anaïs.koptient.etu@univ-lille.fr, natalia.grabar@univ-lille.fr

The question of easy access to medical and health information by patients has attracted attention of the society and researchers. It has indeed been observed that poor understanding of medical and health information by patients may be harmful for their healthcare process. We assume that simplification and adaptation of technical documents may provide a solution to this situation. While the dedicated guidelines to the simplification summarize different kinds of criteria to consider, actually, it is still difficult to respect all these criteria. Usually, automatic systems for text simplification only address some lexical or syntactic transformations. Besides, little work is done on simplification and adaptation of documents from specialized areas, such as medical and health texts. We propose to combine lexical and syntactic simplification within a rule-based system, and to make the simplification process more fine-grained through additional processing. More particularly, we consider transformation of passive sentences into active sentences, and we control the grammatical concordance within sentences. We work with technical medical documents in French. The results are mainly evaluated according to the three measures specifically dedicated to the simplification: semantics, simplicity and grammaticality.

Keywords

automatic text simplification, evaluation, grammaticality, medical texts, natural language processing

1. Introduction

As in any specialized area, the medical area is using specific and technical notions and terms. They are usually not understandable by patients, who have not specific education in medicine. Yet, during the healthcare process, patients are confronted with technical medical documents and information. Usually, patients have some difficulties in understanding such documents. Consider for instance these few examples: information on drug intake [1,2], such as understanding of steps necessary to drug preparation and intake; clinical documents [3] which contain important information for patients and their healthcare process; clinical brochures and informed consents [4] created specifically for patients; more globally, information which can be found on health websites also created for patients [5-10]. For all these reasons, poor and difficult understanding between patients and medical doctors may be complicated and harmful for the healthcare process [11,12]. Hence, the issue related to the easy access to information by patients has gained an important place and attracted attention of researchers and society. Recently, simplification guidelines have been proposed [13-15] thanks to the activity of different associations and initiatives. Such guidelines allow us to understand what should be modified when simplifying a text in order to make it more understandable from the point of view of common readers. Hence, simplification can be done at several levels: lexical, semantic, syntactic, pragmatic, and stylistic. As for the automatic text simplification, three types of methods can be exploited: *methods based on distributional probabilities* with, for instance, word embeddings [16,17], *methods based on automatic translation* [18-21], and *methods based on rules* [22-24]. Most of those works have been done on English texts. Besides, very few works addressed the simplification of medical texts in French [25].

According to the existing guidelines for the creation of simple or simplified texts [13-15], several linguistic phenomena should be taken into account. Among the desiderata, we can find for instance: using short words that are frequent and non-ambiguous, avoiding abbreviations, limiting the variability of the vocabulary used, making sentences that are syntactically simple, avoiding sentences that are in the passive or negative voice, using personal style, explaining difficult concepts, using pictures. Yet nowadays, automatic text simplification is more about lexical phenomena, is especially addressing the

use of short and frequent words when replacing difficult words [16,24,26]. Syntactic phenomena linked to sentence simplicity [27,28] and the use of the active voice [22,27] are also taken into account, but less frequently.

The purpose of our work is to provide automatic simplification of medical texts in French in order to help human users to understand better such texts, and thus to help them to also understand better their pathology or treatments. We use an existing corpus built with documents from different genres (drug leaflets found on the government website, abstracts from systematic reviews and online encyclopedia articles) for fitting our system and testing it. We use a rule-based method. These methods have shown good precision but low recall [24]. We are especially interested by the capacity of this kind of methods to target and to treat simplification phenomena more precisely. Concerning the recall, we think that it might be improved if the lexical resources used have a better coverage and quality.

We first present the method proposed (Section 2). It is rule-based and addresses lexical and syntactic simplification. We then present and discuss the results obtained (Section 3). Finally, we conclude and indicate some issues for future works (Section 4).

2. Methods

The texts used are first syntactically tagged and parsed by Cordial [29], because the system uses the syntactic information of the text in order to do the simplification. Then, the simplification is done at lexical and syntactic levels.

2.1 Lexical simplification

The detection of difficult words and lexical simplification are done with a dedicated lexicon issued from previous work. Currently, this lexicon contains over 6,000 entries. Each entry is within this format: *technical word => simpler equivalent* (like for instance, *abcès (abscess) => accumulation de pus (collection of pus)*, *tachycardie (tachycardia) => accélération du coeur (heart speeding)*, *erythème (erythema) => rougeur (redness)*, *ACO (OAC) => anticoagulant oral (oral anticoagulant)*, *acétonémie (acetoneamia) => taux d'acétone dans le sang (level of acetone in blood)*, *hémarthrose (hematrosis) => saignements musculo-articulaires (musculo-articular bleedings)*). The entries are the technical terms, which are related to non-specialised and easier to understand equivalents, definitions and explanations, or, in case of abbreviations, to their developed forms. This lexicon is also syntactically tagged by Cordial, which provides the morpho-syntactic characteristics for the words of each entry: *abcès/NCMIN (abscess/masculine common noun with the same form in singular and plural)*, *concentration localisée de pus/NCFS VPARPFS PREP NCMIN (localized aggregation of pus/singular feminine common noun, past participle singular feminine, preposition, common masculine noun singular or plural)* is a noun phrase which syntactic head is a singular feminine noun (*concentration*). A given entry can have several possible simplifications. For instance, the word *hypoxie (hypoxia)* has several equivalents: *faible oxygénation (low oxygenation)*, *faible teneur en oxygène (low oxygen level)*, *manque d'oxygène dans le sang (deficit in blood oxygen)*, *manque d'oxygène dans les organes (deficit of oxygen in organs)*, *taux d'oxygène bas dans le sang (low oxygen level in blood)*.

For each syntactic phrase delimited by Cordial, we search if it exists in the lexical resource and, if this is the case, the phrase is considered as difficult to understand and is treated during the simplification. If the term is an abbreviation, then its developed form is added between brackets:

- simplified sentence: *L'OMS (Organisation mondiale de la santé) recommande un calendrier de vaccination antitétanique durant l'enfance de 5 doses. (WHO (World Health Organization) advises a five-injection calendar for the tetanus vaccine during childhood).*

In other cases, the simple equivalent is substituted:

- source sentence: *Elle peut conduire très rarement à une syncope (It may rarely cause a syncope).*
- simplified sentence: *Elle peut conduire très rarement à un évanouissement ou sensation d'évanouissement (It may rarely cause a fainting or a faint feeling).*

When several equivalents are available, the shorter one (which is also considered as the simpler) is chosen, for now. We can also add that in sentences that already contain explanation for a given technical term, this term is therefore not replaced. This is a case of non-simplification context, like in this example, which already explains the meaning of *CARE*:

- source sentence: *L'étude Cholesterol and Recurrent Events (CARE), est un essai randomisé, en double aveugle, contrôlé contre placebo (The study Cholesterol and Recurrent Events (CARE), is a randomized trial, double blinded, controlled against the placebo).*

Lexical substitutions do not always provide satisfying results, especially concerning the sentence grammaticality. Indeed, if a technical term and its simpler equivalent do not share the same part-of-speech category or have different gender (*acétonémie (acetonemia)* which is feminine => *taux d'acétone dans le sang (level of blood acetone)* which is masculine) or number (*hémathrose (hematrosis)* which is singular => *saignements musculo-articulaires (musculo-articular bleedings)* which is plural), then it is necessary to make other modifications within the sentence in order to respect its grammaticality. Thus, when the gender or the number differ within source and simplified phrases, we must concord adjectives, past participles and determiners accordingly:

- source sentence: *les essais de génotoxicité se sont révélés négatifs (genotoxicity trials(plural masculine) appear to be negative(plural masculine))*
- simplified sentence: *les études de génotoxicité se sont révélées négatives (the studies(plural feminine) on genotoxicity appear to be negative(plural feminine)).*

When the number is different, we also change the conjugation of verbs:

- source sentence: *le traitement doit être arrêté progressivement (the treatment must(third person of the singular) be stopped progressively).*
- simplified sentence: *vous devez arrêter le traitement progressivement (you must(second person of the plural) stop the treatment progressively).*

When necessary, the style of the sentence becomes more personal, as in the example above. Besides, we can also take into account both regular and irregular forms of adjectives and verbs. Here is one example with irregular adjective form:

- source sentence: *Les effets hypotenseurs sont additifs (hypotensive effects(plural masculine) are additive).*
- simplified sentence: *Les diminutions de la pression artérielle sont additives (the decrease of arterial pressure(plural feminine) is additive).*

2.2 Syntactic simplification

When the sentences are syntactically complex and contain more than one clause, it is possible, on the one hand, to divide them in order to create simpler sentences, and, on the second hand, to delete information which can be considered as secondary. Here, we are interested in sentence segmentation. Thus, when a sentence includes two clauses, it becomes a candidate for sentence division. Currently, we are using two types of markers to do the division. These markers are related to types of clauses and to discursive markers:

- We consider two types of clauses, detected further to the syntactic analysis of Cordial: subordinate clauses and coordinate clauses. For instance, in the following example, the sentence is segmented on the coordination clause and provides two sentences:
 - source sentence: *L'administration concomitante du chlorhydrate de tamsulosine avec la paroxétine a entraîné une augmentation de la Cmax et de l'ASC du chlorhydrate de tamsulosine d'unfacteur 1,3 et 1,6 respectivement, mais ces augmentations ne sont pas considérées comme étant cliniquement significatives (Concomitant administration of tamsulosin hydrochloride with paroxetine leads to the increase of Cmax and tamsulosin hydrochloride ASC of respectively 1.3 and 1.6, but this increase is not considered to be clinically significative).*
 - simplified sentence: *L'administration concomitante du chlorhydrate de tamsulosine avec la paroxétine a entraîné une augmentation de la Cmax et de l'ASC du chlorhydrate de tamsulosine d'unfacteur 1,3 et 1,6 respectivement (Concomitant administration of tamsulosin hydrochloride with paroxetine leads to the increase of Cmax and tamsulosin hydrochloride ASC of respectively 1.3 and 1.6.). Ces augmentations ne sont pas considérées comme étant cliniquement significatives (This increase is not considered to be clinically significative).*
- On the basis of relative markers (*qui (who)*, *ce qui (which)*, *celui (the one)*, *celle-là (that one)*), we also take into account relative clauses. Indeed, we observed that Cordial cannot detect relative clauses exhaustively. Here is an example of segmentation of sentences with relative clause into two sentences with the recovery of the antecedent of *qui (which)*:

- source sentence: *Le tramadol peut provoquer chez les nouveau-nés des modifications de la fréquence respiratoire, qui sont généralement sans conséquences cliniques préjudiciables (Tramadol can cause in newborn babies modifications of breathing frequency, which do not have harmful clinical consequences)*
- simplified sentence: *Le tramadol peut provoquer chez les nouveau-nés des modifications de la fréquence respiratoire (Tramadol can cause in newborn babies modifications of breathing frequency). Elles sont généralement sans conséquences cliniques préjudiciables (These have no harmful clinical consequences).*
- On the basis of discursive markers (*cependant (however), aussi (also)...*), we take into account other types of syntactically complex sentences, such as in:
 - source sentence: *La ranitidine est éliminée par voie rénale, aussi les taux plasmatiques du médicament augmentent chez les patients présentant une insuffisance rénale (Ranitidine is eliminated by the kidney, also the plasma levels of the medication increase in patients suffering from renal insufficiency).*
 - simplified sentence: *La ranitidine est éliminée par voie rénale (Ranitidine is eliminated by kidney). De cette manière, les taux plasmatiques du médicament augmentent chez les patients présentant une insuffisance rénale (In this way, the plasma levels of the medication increase in patients suffering from renal insufficiency).*

Another kind of syntactic modifications is related to the transformation of sentences with passive voice into sentences with active voice. The transformation rules passive to active are also based on syntactic analysis by Cordial. Hence, if the sentence is in passive voice, our system extracts: (1) the verbal phrase in passive voice, and (2) its subject and object. The verb is then transformed into the active verb form, while the syntactic positions of the object and subject are inverted. In some cases, the verb must also be modified in order to become concorded with the new subject. The following example illustrates this situation:

- source sentence: *Une prudence particulière devra être observée par les conducteurs d'automobiles et les utilisateurs de machines (A particular attention should be paid by car drivers and vehicle users)*
- simplified sentence: *Les conducteurs d'automobiles et les utilisateurs de machines devront observer une prudence particulière (Car drivers and vehicle users should be particularly attentive).*

2.3 Evaluation

In order to evaluate the simplification results, we consider three metrics [19,30] specifically dedicated to this task. They are usually evaluated by human users:

- *Simplicity*. The simplicity judgment has to define whether the simplification makes indeed the sentences easier to understand. In order to measure it, we provide both the technical and the simplified sentences to human evaluators. In each pair of sentences, the sentence order is random. The evaluators are then instructed to indicate which sentence is the simpler one. If the simpler sentence is detected correctly, then we consider that the simplification is real and provides the expected result (simpler sentences);
- *Fluency*. The fluency judgment has to decide whether the simplified sentence is correct grammatically. In other words: is the simplified sentence fluent and understandable? Indeed, in medical texts, the lexical (word substitution) and syntactic (sentence segmentation, transformation from passive voice to active voice) simplification can disrupt the sentence grammaticality. In order to measure the grammaticality of simplification, the evaluators are instructed to indicate the fluency and understandability of the simplified sentence with a Likert scale going from 1 (not grammatical sentence) to 5 (perfectly grammatical sentence);
- *Adequacy*. The adequacy judgment has to define whether the simplified sentence has the same meaning as the technical sentence. In other words: is the simplified sentence semantically close to the technical source sentence? The evaluators are then instructed to indicate the semantic similarity between the two sentences with a Likert scale going from 1 (sentences with independent semantics) to 5 (semantically identical sentences).

The evaluation has been done on ten documents from the CLEAR corpus [31]: seven drug leaflets, two systematic review abstracts and one encyclopedia article. These documents have been randomly selected among all the documents that contain the fine-grained simplification cases treated by our method. The ten evaluated documents contain 3,674 sentences and 53,247 word occurrences. These

documents have not been used for the development and fitting of the system. One hundred sentences, all containing at least one type of transformation aimed, have been manually evaluated by two evaluators. These evaluators had to provide their judgment on these three measures. The evaluators have no medical training and represent common knowledge of medical notions.

As for the evaluation of the intermediate steps of our method (lexical substitution, development of abbreviations, gender, number, verbal and determiner agreement, passive to active transformation), we also compute the precision of the results. Besides, we compute the Spearman correlation [32] for fluency and adequacy.

3. Results of the Simplification

In Table 1, we indicate the total number of transformations that have been done within the ten test documents (column *Total number of transformations*). In the next column, we indicate the number of transformations contained in the hundred sentences that have been manually evaluated (column *Number of evaluated transformations*). The column *Number of correct transformations* indicates the number of transformations that have been considered as correct by the evaluators. Finally, the last column *Precision* indicates the precision obtained at these steps.

Table 1 Transformations done during the simplification and their evaluation.

| Transformation types | Total number of transformations | Number of evaluated transformations | Number of correct transformations | Precision |
|---|---------------------------------|-------------------------------------|-----------------------------------|-----------|
| <i>Concordance (gender, number)</i> | 14 | 14 | 6 | 0.43 |
| <i>Concordance (verbs)</i> | 138 | 49 | 32 | 0.65 |
| <i>Concordance (determiners)</i> | 575 | 28 | 22 | 0.79 |
| <i>Passive to active transformation</i> | 103 | 31 | 23 | 0.74 |
| <i>Lexical substitutions</i> | 1,153 | 1,153 | - | - |
| <i>Abbreviation developments</i> | 159 | 159 | - | - |

The adjustment of grammatical concordance shows 0.62 of the overall precision. We can distinguish three types of concords within sentences, which we present, explain and exemplify in what follows:

- Cases, in which it is necessary to concord words on gender and number, have low frequency (14 occurrences only) and low precision (0.43) as well. In the following example, we can see that the noun phrase, upon which the verb must be concorded, has a nested structure, which may cause the failure for the automatic system:
 - simplified sentence: *Cependant une surveillance devra être effectuée chez ces patients car certains symptômes_(plural masculine) de la baisse du taux de sucre dans le sang peuvent être masquées_(plural feminine). (Yet, the surveillance must be done in such patients because some symptoms due to the decrease of sugar level in blood can be hidden.)*

In this example, the system fails to compute the correct noun with which the concordance must be done: it decides on putting the past participle *masqué* (*hidden*) in plural feminine, while the correct form would be plural masculine related to *certaines symptômes* (*some symptoms*).

- We count 138 cases in which it is necessary to concord the verbal forms with the new subjects. Among the 49 transformations that have been evaluated, 32 are correct, which gives 0.65 precision. In the following example, the negation occurring in the sentence introduces specific syntactic structure and causes the failure of the correct transformation:
 - simplified sentence: *La sertraline n'a provoquer_(infinitive verb) l'apparition ni la stimulation ni l'anxiété associées à la d-amphétamine, ni la sédation et l'altération psychomotrice associées à l'alprazolam. (Sertraline did not cause occurrence, or stimulation, or anxiety associated to*

d-amphetamine neither did it cause sedation or psychomotor alteration associated with alprazolam.)

Here, the main verb is in the infinitive form *provoquer* (cause) instead of the past participle form *provoqué* (caused).

- We count 575 cases in which it is necessary to concord the determiner. Among the 28 transformations that have been evaluated, 22 are correct, which gives 0.79 precision. In the following example, it is necessary to contract the preposition with the determiner that follows, which the system fails to do:
 - simplified sentence: *Dans les cas extrêmes de résistance de les_(preposition+determiner) battements cardiaques très lents au traitement, un stimulateur cardiaque pourra être mis en place.* (In cases of extreme resistance to treatment of slow heartbeats, cardiac pacemaker may be inserted.)

In this example, the *de les* sequence must be contracted in *des*.

Concerning the two other syntactic transformations:

- The ten documents analyzed contain 103 sentences in the passive voice which have been transformed into sentences with active voice. The 31 sentences that have been evaluated show 0.74 precision. Globally, this type of transformation is performed correctly. Among the errors observed, we can show this example:
 - source sentence: *Les bêta-bloquants doivent être administrés avec prudence* (The beta blockers must be administrated with caution.)
 - simplified sentence: *Les bêta-bloquants vous devez administrer avec prudence* (The beta blockers you should administrate with caution).

The error is due to the syntactic structure chosen to make the sentence style personal when addressing the patient. Hence, the style of the sentence could not be respected.

- The evaluation of the 11 segmented sentences shows 0.64 precision. These segmentations are triggered in the following cases: four relative propositions and three coordinated propositions that have been correctly divided, while two discursive markers (*aussi* [also]), and two subordinate propositions both introduced by *comme* (as) have been segmented incorrectly. Due to its inherent semantic complexity, we assume that this last case (subordination introduced by *comme* (as)) should be excluded from the sentence segmentation triggers.

As for the lexical substitution, we count 1,153 cases of substitutions and 159 cases of abbreviation developments. First of all, the quality of such transformations depends on the quality of the exploited resource. Their quality is also evaluated according to the three metrics specific to simplification:

- *Simplicity*: among the 100 evaluated sentences, 30 have been judged as badly simplified, which corresponds to 0.70 precision. The errors in the detection of simplified sentences may be due to the fact that simplified sentences become longer (contain more words): lexical substitutions of technical terms by their paraphrases and development of abbreviation make the sentences longer. Hence, in the following example, the source sentence contains 19 words while its simplified version contains 34 words:
 - source sentence: *Cependant une surveillance devra être effectuée chez ces patients car certains symptômes de l'hypoglycémie peuvent être masqués (tachycardie, palpitations).* (Yet, the surveillance must be done on such patients because some hypoglycemia symptoms may be hidden (tachycardia, palpitations).)
 - simplified sentence: *Cependant une surveillance devra être effectuée chez ces patients car certains symptômes de la baisse du taux de sucre dans le sang peuvent être masquées (accélération des battements du coeur, battements du coeur gênants).* (Yet, the surveillance must be done on such patients because some symptoms related to the decrease of blood sugar level may be hidden (increase of the heart rate, t palpitations).)

Since several terms (*hypoglycemia, tachycardia, palpitation*) are replaced by the equivalent paraphrases, the sentence becomes indeed much longer and, for this reason, it may also become more complicated to parse and to understand.

- *Fluency*: the average index of fluency is between 4.52 and 4.71, according to the evaluators. It is a relatively high value, which shows that globally the fluency is kept. We think that the fine-grained treatment of sentences (grammatical concordance and anaphora resolution) permits to preserve the fluency during the simplification. The Spearman correlation index is 0,2419.
- *Adequacy*: the average index of adequacy is between 4.52 and 4.83, according to the evaluators. This is also a high value, which shows that the evaluators consider that semantics is preserved

during the simplification. We think that this index is mainly dependent on the quality of the resource used for lexical substitutions. Besides, the fact that the evaluators do not have medical training may bias this evaluation measure, as the evaluators present the potential to incorrectly judge the semantics of the evaluated sentences. The Spearman correlation index is 0,4409.

These three evaluation measures are not used systematically in the existing works. We can only compare our results with those obtained in one previous work [19] done with texts in English. Globally, our system shows better scores for adequacy and fluency, which are between 3 and 4 in the cited work. Our simplicity scores are comparable with those from this previous work. As the correlation index between the annotators is quite low, we plan to involve more annotators in the process.

If our method attempts to satisfy different *desiderata* from the existing simplification guidelines, it also shows that some aspects have to be improved yet. The evaluation proposed mainly indicates that the issues related to the syntactic transformation (sentence division, concordance, transformation from passive voice to active voice) are performed quite well but need still some additional improvements. Another perspective is related to the improvement of adequacy. For instance, abbreviations can be ambiguous and have several possible developments, like *AFS*, which means *Agence Française du Sang* (*French Blood Agency*) but also *American Fertility Society*. In our resource, 171 abbreviations are affected by ambiguity. We are currently working on context-based disambiguation of abbreviations with supervised methods.

4. Conclusions

In this paper, we described and evaluated a rule-based method dedicated to the simplification of technical documents from medical domain in French. In its functionalities, the method is inspired by previous research works in simplification, but also by the *desiderata* expressed in the existing simplification guidelines. Thus, we take into account different lexical and syntactic criteria. This leads to the simplification process performed at two levels and addressing several fine-grained issues (grammatical concordance, transformation passive→active, sentence segmentation, non-simplification contexts). An evaluation done on 100 sentences with specific simplification metrics shows that fluency and adequacy are well respected while the simplicity must be improved. This last issue is mainly related to the cases when lexical substitutions lead to the increase of the sentence size, due to the fact that simpler lexical equivalents are usually longer lexical units (paraphrases, definitions, abbreviation developments). We have several perspectives, such as taking into account other linguistics criteria and aspects and the improvement of simplicity.

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 authors would like to thank the reviewers for their questions and comments that helped in improving the article.

References

- [1] Vander Stichele R. Promises for a measurement breakthrough. In : Sons JW, editor. Drug regimen compliance. Issues in clinical trials and patient management. JM Metry and UA Meyer ; 1999. p. 71–83.
- [2] Patel V, Branch T, Arocha J. Errors in interpreting quantities as procedures: *The case of pharmaceutical labels*. Int Journ Med Inform. 2002 ;65(3) :193–211.
- [3] Zeng-Treiler Q, Kim H, Goryachev S, Keselman A, Slaughter L, Smith C. Text characteristics of clinical reports and their implications for the readability of personal health records. In : MEDINFO. Brisbane, Australia ; 2007. p. 1117–1121.
- [4] Williams M, Parker R, Baker D, Parikh N, Pitkin K, Coates W, et al. *Inadequate functional health literacy among patients at two public hospitals*. JAMA. 1995 ;274(21) :1677–1682.
- [5] Rudd R, Moeykens B, Colton T. In : Annual Review of Adult Learning and Literacy. NCSALL ; 1999. p. ch 5.

- [6] Berland G, Elliott M, Morales L, Algazy J, Kravitz R, Broder M, et al. *Health information on the Internet. Accessibility, quality, and readability in English and Spanish*. JAMA. 2001 ;285(20) :2612–2621.
- [7] Mcgray A. *Promoting Health Literacy*. J of Am Med Infor Ass. 2005 ;12 :152–163.
- [8] Oregon Practice Center. *Barriers and Drivers of Health Information Technology Use for the Elderly, Chronically Ill, and Underserved*. Agency for healthcare research and quality. Oregon Evidence-based Practice Center; 2008.
- [9] D'Alessandro D, Kingsley P, Johnson-West J. *The readability of pediatric patient education materials on the World Wide Web*. Arch Pediatr Adolesc Med. 2001 ;155(7) :807–12.
- [10] Brigo F, Otte M, Igwe S, Tezzon F, Nardone R. *Clearly written, easily comprehended ? The readability of websites providing information on epilepsy*. Epilepsy & Behavior. 2015 ;44 :35–39.
- [11] Jucks R, Bromme R. *Choice of words in doctor-patient communication: an analysis of health-related internet sites*. Health Commun. 2007 ;21(3) :267–77.
- [12] Tran T, Chekroud H, Thiery P, Julienne A. *Internet et soins : un tiers invisible dans la relation médecin/patient ?* Ethica Clinica. 2009 ;53 :34–43.
- [13] Ruel J, Kassi B, Moreau A, Mbida-Mballa S. *Guide de rédaction pour une information accessible*. Gatineau : Pavillon du Parc ; 2011. Available from : <https://www.autisme.qc.ca/assets/files/07-boite-outils/Communication/Guide-pavillon-parc.pdf>.
- [14] OCDE. *Guide de style de l'OCDE Troisième édition : Troisième édition*. OECD Publishing ; 2015. Available from : <https://books.google.fr/books?id=3EK8CgAAQBAJ>.
- [15] UNAPEI. *L'information pour tous*. UNAPEI ; 2019. Available from : https://easy-to-read.eu/wp-content/uploads/2014/12/FR_Information_for_all.pdf.
- [16] Glavas G, Stajner S. *Simplifying Lexical Simplification : Do We Need Simplified Corpora ?* In : ACL-COLING ; 2015. p. 63–68.
- [17] Kim YS, Hullman J, Burgess M, Adar E. *SimpleScience: Lexical Simplification of Scientific Terminology*. In: EMNLP ; 2016. p. 1–6.
- [18] Zhao S, Wang H, Liu T. *Leveraging Multiple MT Engines for Paraphrase Generation*. In: COLING; 2010. p. 1326–1334.
- [19] Wubben S, van den Bosch A, Krahmer E. *Sentence Simplification by Monolingual Machine Translation*. In: Annual Meeting of the Association for Computational Linguistics ; 2012. p. 1015–1024.
- [20] Sennrich R, Haddow B, Birch A. *Improving Neural Machine Translation Models with Monolingual Data*. In : Proc of the Ann Meeting of the Assoc for Comp Linguistics. Berlin, Germany ; 2016. p. 86–96.
- [21] Nisioi S, Stajner S, Ponzetto SP, Dinu LP. *Exploring Neural Text Simplification Models*. In: Ann Meeting of the Assoc for Comp Linguistics; 2017. p. 85–91.
- [22] Carroll J, Minnen G, Pearce D, Canning Y, Devlin S, Tait J. *Simplifying Text for Language- Impaired Readers*. In: EACL; 1999. p. 269–270.
- [23] Bautista S, Gervás P, Madrid RI. *Feasibility analysis for semi-automatic conversion of text to improve readability*. In: Int Conf on Inform and Comm Technology and Accessibility (ICTA) ; 2009. p. 33–40.
- [24] De Belder J, Deschacht K, Moens MF. *Lexical Simplification*. In: ITEC; 2010. 1-4.
- [25] Cardon R. *Approche lexicale de la simplification automatique de textes médicaux*. In Proc RECITAL 2018. p 1-15
- [26] Specia L, Jauhar S, Mihalcea R. *SemEval-2012 Task 1: English Lexical Simplification*. In: *SEM 2012 ; 2012. p. 347–355.
- [27] Chandrasekar R, Doran C, Srinivas B. *Motivations and methods for text simplification*. In: COLING; 1996. p. 1041–1044.
- [28] Brouwers L, Bernhard D, Ligozat AL, François T. *Simplification syntaxique de phrases pour le français*. In: Traitement Automatique des Langues Naturelles (TALN) ; 2012. p. 211–224.
- [29] Laurent D, Nègre S, Séguéla P. *L'analyseur syntaxique Cordial dans Passage*. In: Traitement Automatique des Langues Naturelles (TALN); 2009. .
- [30] Biran O, Brody S, Elhadad N. *Putting it Simply: a Context-Aware Approach to Lexical Simplification*. In: Annual Meeting of the Association for Computational Linguistics ; 2011.
- [31] Grabar N, Cardon R. *CLEAR – Simple Corpus for Medical French*. In: Workshop on Automatic Text Adaption (ATA) ; 2018. p. 1–11.
- [32] Spearman C. (1904). *“The proof and measurement of association between two things”*. American Journal of Psychology. **15** (1): 72–101