Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles
Avignon, France, 27 juin au 1er juillet 2022

Evaluation of Automatic Text Simplification: Where are we now, where should we go from here

Natalia Grabar[1] Horacio Saggion[2]
(1) CNRS, Univ Lille, UMR 8163 – STL, F-5900 Lille, France
(2) LaSTUS Lab/TALN Group, Universitat Pompeu Fabra, Spain
natalia.grabar@univ-lille.fr, horacio.saggion@upf.edu

RÉSUMÉ

The purpose of automatic text simplification is to adapt the content of documents in order to make them easier to understand by a given population or to improve the performance of NLP tasks such as summarization or information extraction. The main steps for the automatic text simplification systems are quite well defined and researched in existing work but the evaluation of the simplification output remains understudied. Indeed, contrary to other NLP tasks, like information retrieval and extraction, terminology structuring, or question-answering, which expect factual and consensual outputs of the systems, it is difficult to define a standard output of simplification. There is considerable subjectivity in the simplification process, and it is not consensual because it is heavily based on own knowledge of people. Hence, several factors are involved in the simplification process and its assessment. In this paper, we present and discuss some of these factors : the role of end users, the reference data, the domain of source documents, and the evaluation measures.

ABSTRACT

Évaluation de la simplification automatique de textes : où nous en sommes et vers où devons-nous aller.

L'objectif de la simplification automatique de textes consiste à adapter le contenu de documents afin de les rendre plus faciles à comprendre par une population donnée ou bien pour améliorer les performances d'autres tâches TAL, comme le résumé automatique ou extraction d'information. Les étapes principales de la simplification automatique de textes sont plutôt bien définies et étudiées dans les travaux existants, alors que l'évaluation de la simplification reste sous-étudiée. En effet, contrairement à d'autres tâches de TAL, comme la recherche et extraction d'information, la structuration de terminologie ou les questions-réponses, qui s'attendent à avoir des résultats factuels et consensuels, il est difficile de définir un résultat standard de la simplification. Le processus de simplification est très subjectif et souvent non consensuel parce qu'il est lourdement basé sur les connaissances propres des personnes. Ainsi, plusieurs facteurs sont impliqués dans le processus de simplification et son évaluation. Dans ce papier, nous présentons et discutons quelques uns de ces facteurs : le rôle de l'utilisateur final, les données de référence, le domaine des documents source et les mesures d'évaluation.

MOTS-CLÉS : Simplification, evaluation of simplification, end user, reference data, measures.

KEYWORDS: Simplification, évaluation de la simplification, utilisateur final, données de référence, mesures.

Volume 1 : conférence principale, pages 453–463.

1 Introduction

The purpose of automatic text simplification is to adapt the content of documents in order to make them easier to understand (Saggion, 2017) by a given population including children (Son *et al.*, 2008; De Belder & Moens, 2010; Vu *et al.*, 2014), foreigners or people with low literacy (Paetzold & Specia, 2016), people with neurodegenerative disorders (Chen *et al.*, 2016), or people without specialized knowledge (Arya *et al.*, 2011; Leroy *et al.*, 2013; Cardon & Grabar, 2020). Text simplification can also help other NLP applications and has been used for example to prepare documents further for syntactic analysis (Chandrasekar & Srinivas, 1997; Jonnalagadda *et al.*, 2009), semantic annotation (Vickrey & Koller, 2008), summarization (Blake *et al.*, 2007), automatic translation (Stymne *et al.*, 2013; Štajner & Popović, 2016), indexing (Wei *et al.*, 2014), or information retrieval and extraction (Beigman Klebanov *et al.*, 2004).

Simplification can be performed at different linguistic levels : lexical, syntactic, semantic, or pragmatic. The main steps and requirements for the automatic text simplification systems are now quite well defined and researched. We can mention for instance text difficulty assessment, complex words identification, disambiguation, word difficulty ranking, building of models and resources, or sentence alignment. Yet, the evaluation of simplification remains understudied. Indeed, contrary to other NLP tasks, like information retrieval and extraction, terminology structuring, or question-answering, which expect factual and consensual outputs of the systems, it is difficult to define standard output of simplification : (1) it is not factual because it relies on series of transformations more or less managed by human simplificators and automatic systems, and (2) it is not consensual because it is heavily based on own knowledge of people and because everyone has an opinion on the simplification output. As it has been noticed, *native simplified-language speaker does not exist* (Siddharthan, 2014). Hence, several factors are involved in the simplification process and its assessment. In this paper, we propose to discuss the role of four important factors : the role of end users (Section 2), the reference data (Section 3), the domain of source documents (Section 4), and the evaluation measures (Section 5).

2 End user

Different users may need different simplification strategies. For instance, children do not have the same needs as people with neurodegenerative disorders or foreigners : types of documents and their content differ, as well as required adaptation of documents. Besides, the level of literacy of people varies a lot, including within the a given population (children, foreigners...). Six standard literacy levels are usually distinguished (Bernèche & Perron, 2006; OECD, 2019), and rely on the following skills :

0. read brief texts on familiar topics, locate a single piece of information, know basic vocabulary,
1. read short texts, locate synonymous information, recognize basic vocabulary, determine the meaning of sentences,
2. match between the text and information, paraphrase, make low-level inferences,
3. read and navigate in dense, lengthy or complex texts,
4. integrate, interpret information from complex texts, identify and understand non-central ideas, interpret or evaluate subtle evidence-claim or persuasive discourse relationships,

5. search for, and integrate, information across multiple texts, construct syntheses of similar and contrasting ideas, evaluate evidence based arguments, understand subtle cues, make high-level inferences, use specialized background knowledge.

The levels 2 or 3 correspond to the high school level : persons can read and understand quite complex non-specialized texts, make some inferences and paraphrase content. To ensure that a given person belongs to a given literacy level, specific tests are used. Such tests are specific to a given population, needs, and language, like those dedicated to health literacy in different languages (Lee *et al.*, 2010; Mancuso, 2009; Rouquette *et al.*, 2018).

Usually, the real user is not involved in the evaluation leaving unexplored the question of how useful the simplification is. To perform a correct evaluation of the simplification, the evaluation must be user-dependent and consider the needs of final users. This accounts for instance for the type of target population, its literacy level, the type of documents, the readability level of these documents, the expected simplification rules (lexical, syntactic...), and the expected simplification level.

3 Reference data

The reference data play a very important role in the development and evaluation of simplification systems. Several approaches exist for the creation of such data, which we present and discuss :

— *Expert judgment* : experts use their theoretical knowledge about needs of the target population (Clercq *et al.*, 2014). The main limitation is that experts do not always know the real needs of the population and that the expert judgements are difficult to generalize over the target population ;

— *Textbooks* exploited as the reference data. They are indeed created for a given population and respect a given readability level, such as school books (François & Fairon, 2013; Gala *et al.*, 2013). The main limitation is that such textbooks are also created following a particular theoretical framework and thus are not always generalizable ;

— *Crowdsourcing* (Clercq *et al.*, 2014; Xu *et al.*, 2016; Alva-Manchego *et al.*, 2020b). The advantage is that crowdsourcing involves large population : the output can be representative of a large set of people. However, inclusion/exclusion criteria are often difficult to implement, and the population involved is uncontrolled ;

— *Eye-tracking* records eye movements during reading : it indicates precisely which words or segments require more attention (fixations are longer and saccades are shorter). The advantage of this approach is that it provides fine-grained and objective analysis of reading difficulties (Yaneva *et al.*, 2015; Grabar *et al.*, 2018). The main limitation is that usually short text spans of texts are used and, for this reason, generalization to longer segments might be impossible.

— *Manual annotation or simplification* (Audiau, 2009; Liffran, 2015; Grabar & Hamon, 2016; Gala *et al.*, 2020). The main limitations of this approach : (1) large variability across the annotators with small chance to reach concensus because each person has his own knowledge and understanding feeling, (2) variability within the same annotator because, when a given complex word is read several times, this word may become more familiar (Grabar & Hamon, 2017), (3) inconsistency in annotation due for instance to fatigue or to limited knowledge of simplification rules and principles, (4) face-saving strategies, which may prevent from annotating exhaustively the documents. (5) Besides, this is a very long and tedious process which may limit the size of the data annotated.

In all these approaches, annotators or experts represent a given population through their own expe-

rience or theoretical knowledge. The annotations are expected to be extrapolated to similar target population and task. As we discussed, each approach has its own advantages and limitations, which should be taken into account during the evaluation process. In any case, we can argue that final users must be involved in the creation of the reference data in order to better represent their own needs.

Hence, the reference data must fit a given literacy level for a given population. This is an important yet complicated issue because there is no clear indications on the simplification requirements for each literacy level. Existing simplification guidelines (Ruel *et al.*, 2011; OCDE, 2015; UNAPEI, 2019) are very vague and indicate only general principles without specification of the literacy levels : use short words, use frequent and non-ambiguous words, avoid abbreviations, limit the variability of the vocabulary, make syntactically simple sentences, avoid sentences in passive or negative voice, use personal style, explain difficult concepts, use pictures, etc. This means that each work on simplification should reimplement such principles in order to create clear and exploitable simplification rules.

The reference data created for simplification are available for several languages : Basque (Gonzalez-Dios *et al.*, 2018), Danish (Klerke & Søgaard, 2012), English (Daelemans *et al.*, 2004; Petersen & Ostendorf, 2007; Specia *et al.*, 2012; Zhang & Lapata, 2017a), French (Grabar & Cardon, 2018; Gala *et al.*, 2020), German (Klaper *et al.*, 2013; Säuberli *et al.*, 2020), Italian (Brunato *et al.*, 2016; Tonelli *et al.*, 2016), Japanese (Goto *et al.*, 2015), Portuguese (Aluisio *et al.*, 2008; Caseli *et al.*, 2009), Russian (Dmitrieva & Tiedemann, 2018), Spanish (Collados, 2013; Bott *et al.*, 2014). These corpora mainly contain parallel and aligned sentences from existing sources or are crafted manually. Some corpora also indicate transformations due to the simplification types (Brunato *et al.*, 2014; Koptient *et al.*, 2019). The content and quality of these corpora are not assessed.

4 Domain of source documents

The domain of source documents has impact on the simplification result and its evaluation. For instance, documents from general and specialized languages do not require the same simplification transformations or resources. The particularity of specialized languages is double : (1) In some domains (medicine, biology, physics...), they convey specific terminology, which requires intense simplification at the lexical and semantic levels. (2) In other domains (legal and administrative), they have specific syntactic structures, which requires their syntactic simplification. This situation may increase distance between the source and simplified texts, which causes lower scores for some evaluation measures. In the following example from the biomedical domain, the two sentences are semantically identical yet they are very distant lexically and syntactically :

> *Medication inhibiting the peristalsis are counter-indicated in this situation.*
> *In this case, do not take medication for stopping or decreasing the intestinal transit.*

When simplification with synonyms or equivalent expressions is impossible, it becomes necessary to use more general terms or add explanations, which also leads to an increasing lexical and syntactic distance between source and simplified sentences and documents.

5 Evaluation measures

The purpose of evaluation measures is to assess the quality in simplification research. This is a sensitive issue but still under-studied. We present and discuss several evaluation measures, which can

be divided in two types : human (Section 5.1) and automatic (Section 5.2) evaluation.

5.1 Human evaluation

When human judgment is required about the simplification output, three criteria are commonly used (Nisioi *et al.*, 2017; Cardon, 2021) : *semantics* (or *adequacy*) to state whether the meaning is preserved further to the simplification ; *grammaticality* (or *fluency*) to state whether the simplified text remains grammatical and understandable ; and *simplicity* to state whether the simplified text is simpler than the source text. These criteria are assessed without reference data. Evaluation is usually done with grids, in which 1 corresponds to minimal quality (not simple, no semantic relatedness, not grammatical) and 5 to maximal quality. The evaluators have to assess each criteria choosing between maximal to minimal values. This kind of evaluation is subjective. It depends on individual background of annotators and their feeling about the simplicity, semantics and grammaticality. This means that the reproducibility from one annotator to another may be low. It has also been argued that the evaluation guidelines are usually vague (Stodden, 2021). Hence, to reach better objectivity in manual evaluation, the use of checklists has been proposed (Cumbicus-Pineda *et al.*, 2021).

5.2 Automatic measures

Automatic evaluation measures usually compare the simplification output against the reference data. These measures are expected to be more objective. Some metrics are intended to evaluate lexical simplification (like those computing the similarity) and others the whole process. Yet, different metrics used for the evaluation of simplification show limitations and cover the three evaluation criteria (semantics, grammaticality and simplicity) only partially. In what follows, we present and discuss several series of measures used for the evaluation of simplification results.

Standard evaluation metrics like *precision* and *accuracy* have been used in some experiments for measuring lexical simplification (Horn *et al.*, 2014). Such metrics compare the really obtained simplification with the reference data. Higher values indicate that simplification is better because it is closer to the reference data. Such metrics have been developed for more factual and consensual NLP tasks, like information retrieval and extraction, and are not suitable for the evaluation of simplification because its output is not factual.

Other measures are borrowed from the textual similarity task (Levenshtein, 1966; Vásquez-Rodríguez *et al.*, 2021). The original string edit distance corresponds to the minimal number of single-character edits (insert, delete or substitute) required to change one word into the other, or one sentence into the other. Besides, two adaptations have been proposed : (1) *EditNTS* (Dong *et al.*, 2019) which allows to detect and predict three operations (insert, delete, keep), (2) SeqLabel (Alva-Manchego *et al.*, 2020a), which purpose is to automatically identify transformation operations in the original parallel corpus. String edit measures are not really suitable for the evaluation of simplification : simplification implies lexical and syntactic transformations of the source text, which increases the distance. Besides, these transformations are not consensual. Nevertheless, such measures can be used to pre-annotate some simplification-induced transformations.

Not surprisingly, machine translation area proposes evaluation measures which can be exploited for the evaluation of simplification. Indeed, simplification can be considered as monolingual translation of documents from original to simplified languages :

— *BLEU* (*bilingual evaluation understudy*) (Papineni *et al.*, 2002) is an adaptation of precision and takes also into account the word order (n-grams). Like with precision, the evaluation principle is *the higher the better*, which means that when the simplification output is closer to the reference data it has better quality. Some existing work in simplification noticed that this metric is correlated with grammaticality (Wubben *et al.*, 2012; Martin *et al.*, 2018) and semantics (Martin *et al.*, 2018), but not with simplicity ;
— *TERp* (*Translation Edit Rate plus*) (Snover *et al.*, 2009) computes the number of edition operations (insert, delete, substitute, inverse) when transforming one sentence into the other. The principle is *the lower, the better*, which means that when less transformations are required to fit the reference sentence the quality of simplification is better ;
— *OOV* (*out of vocabulary*) is the rate of words missing from the reference vocabulary (Vu *et al.*, 2014). In relation with simplification, if the number of out of vocabulary words is lower the readability of the text is better. Often, the *Basic English list* (Ogden, 1930) is exploited to compute the OOV rate.

In relation with the *OOV* metric, several measures are borrowed from the readability domain. Readability assesses the difficulty of the text. For this, *classical readability scores* (Flesch, 1948; Gunning, 1973; Björnsson & Härd af Segerstad, 1979) are exploited. Their value ranges depend on scores : for some of them low values indicate that the text is simple, while for others high values indicate that the text is simple. According to existing work, these metrics are correlated with syntactic simplicity (Vu *et al.*, 2014). Yet, as simplification often outputs longer sentences such metrics become less suitable (Wubben *et al.*, 2012). Other works observe that such measures are not correlated with simplicity (Woodsend & Lapata, 2011; Wubben *et al.*, 2012; Zheng & Yu, 2017; Tanprasert & Kauchak, 2021).

Finally, there are few new evaluation measures proposed explicitely for simplification : (1) *changed* (Horn *et al.*, 2014), which is the percentage of the test examples where the system suggested some changes, be they correct or not, with the objective to produce the highest number of changes ; (2) *potential* (Paetzold & Specia, 2016), which is the rate of instances among which at least one proposed candidate is in the reference data, with the objective to propose the highest number of such candidates ; (3) *SARI* (Xu *et al.*, 2016), which performs a comparison with reference and source data, like BLEU. Higher values indicate a better simplification. It is considered that this metric is more reliable when several reference datasets are available (Alva-Manchego *et al.*, 2020c; Zhang & Lapata, 2017b). Current work also argues that this metric has no correlation with simplicity (Alva-Manchego *et al.*, 2020b; Cardon & Grabar, 2020) but may be correlated with lexical similarity instead.

Currently, several of these metrics are commonly used for the evaluation of simplification. As discussed, there is no consensus on use of these metrics for the evaluation of simplification. Besides, some metrics may be correlated with some of the three criteria (semantic, grammaticality and simplicity) but none of them covers all the criteria. Hence, it can be necessary to combine certain metrics for a more precise evaluation. Let's also add that the main work on evaluation is done on data in English, while several of these measures are language-dependent, and that evaluation is usually performed at the level of sentence and rare work go beyond the sentence (Todirascu *et al.*, 2013).

6 Conclusion

Evaluation of text simplification systems is still an understudied area. It is heavily based upon evaluation approaches exploited in other NLP tasks. Yet, the simplification task is different from these

other NLP tasks, mainly because its output is subjective. As noticed, this is due to the facts that native simplified-language speaker does not exist, simplification guidelines are vague, there is little consensus on simplification output, and simplification needs depend on the population aimed. This introduces inherent difficulty when creating the reference data and evaluating the simplification, especially for languages other than English, as several of the evaluation measures are language-dependent. In future work, it is necessary to research further the evaluation of simplification : to propose clearer principles for manual evaluation, to define a stronger association between the evaluation criteria and measures, and to propose new and more flexible evaluation metrics. Besides, if the simplification systems are designed for specific target population, then this population should be involved in the development of the solution and the evaluation of the results.

Acknowledgments

Our work is partly supported by the project Context-aware Multilingual Text Simplification (ConMuTeS) PID2019-109066GB-I00/AEI/10.13039/501100011033 awarded by Ministerio de Ciencia, Innovación y Universidades (MCIU), by Agencia Estatal de Investigación (AEI) of Spain, and by the French National Agency for Research (ANR) as part of the CLEAR project (Communication, Literacy, Education, Accessibility, Readability), ANR-17-CE19-0016-01.

Références

ALUISIO S., SPECIA L., PARDO T., MAZIERO E. & DE MATTOS FORTES R. (2008). Towards Brazilian Portoguese automatic text simplification systems. In *Proc of the eighth ACM symp on Document engineering*, p. 240–248.

ALVA-MANCHEGO F., BINGEL J., PAETZOLD G., SCARTON C. & SPECIA L. (2020a). Learning how to simplify from explicit labeling of complex-simplified text pairs. In *Proc of the Eighth Inter Joint Conf on Natural Language Processing*, p. 295–305, Taipei, Taiwan.

ALVA-MANCHEGO F., MARTIN L., SCARTON A. B. C., SAGOT B. & SPECIA L. (2020b). ASSET : A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations. In *Proc of the 58th Annual Meeting of the Assoc for Comp Linguistics*, p. 4668–4679.

ALVA-MANCHEGO F., SCARTON C. & SPECIA L. (2020c). Data-driven sentence simplification : Survey and benchmark. *Computational Linguistics*, p. 1–87.

ARYA D. J., HIEBERT E. H. & PEARSON P. D. (2011). The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *Int Electronic Journal of Elementary Education*, **4**(1), 107–125.

AUDIAU A. (2009). *L'information pour tous. Règles européennes pour une information facile à lire et à comprendre*. Rapport interne, Nous aussi, UNAPEI.

BEIGMAN KLEBANOV B., KNIGHT K. & MARCU D. (2004). Text simplification for information-seeking applications. In R. MEERSMAN & Z. TARI, Éds., *On the Move to Meaningful Internet Systems 2004 : CoopIS, DOA, and ODBASE*. Berlin, Heidelberg : Springer, LNCS vol 3290.

BERNÈCHE F. & PERRON B. (2006). *Développer nos compétences en littératie : un défi porteur d'avenir. Enquête internationale sur l'alphabétisation et les compétences des adultes*. Rapport interne, Institut de la statistique du Québec, Canada.

BJÖRNSSON H. & HÄRD AF SEGERSTAD B. (1979). Lix pä franska och tio andra språk. *Stockholm : Pedagogiskt centrum, Stockholms skolförvaltning*.

BLAKE C., KAMPOV J., ORPHANIDES A., WEST D. & LOWN C. (2007). Query expansion, lexical simplification, and sentence selection strategies for multi-document summarization. In *DUC*.

BOTT S., SAGGION H. & MILLE S. (2014). Text simplification tools for Spanish. In *LREC 2014*, p. 1–7.

BRUNATO D., CIMINO A., DELL'ORLETTA F. & VENTURI G. (2016). PaCCSS–IT : A parallel corpus of complex–simple sentences for automatic text simplification. In *Proc of Conf on Empirical Methods in Natural Language Processing (EMNLP 2016)*, p. 351–361, Austin, Texas, USA.

BRUNATO D., DELL'ORLETTA F., VENTURI G. & MONTEMAGNI S. (2014). Defining an annotation scheme with a view to automatic text simplification. In *CLICIT*, p. 87–92.

CARDON R. (2021). *Simplification Automatique de Textes Techniques et Spécialisés*. Phd, Université de Lille, Lille, France.

CARDON R. & GRABAR N. (2020). French biomedical text simplification : When small and precise helps. In *COLING 2020*, p. 1–8.

CASELI H. M., PEREIRA T. F., SPECIA L., PARDO T. A. S., GASPERIN C. & ALUISIO S. M. (2009). Building a Brazilian Portuguese parallel corpus of original and simplified texts. In *CICLING*, p. 1–12.

CHANDRASEKAR R. & SRINIVAS B. (1997). Automatic induction of rules for text simplification. *Knowledge Based Systems*, **10**(3), 183–190.

CHEN P., ROCHFORD J., KENNEDY D. N., DJAMASBI S., FAY P. & SCOTT W. (2016). Automatic text simplification for people with intellectual disabilities. In *AIST*, p. 1–9.

CLERCQ O. D., HOSTE V., DESMET B., VAN OOSTEN P., COCK M. D. & MACKEN L. (2014). Using the crowd for readability prediction. *Natural Language Engineering*, **20**, 293–325.

COLLADOS J. (2013). Splitting complex sentences for natural language processing applications : Building a simplified Spanish corpus. *Procedia-Social and Behavioral Sciences*, **95**, 464–472.

CUMBICUS-PINEDA O. M., GONZALEZ-DIOS I. & SOROA A. (2021). Linguistic capabilities for a checklist-based evaluation in automatic text simplification. In *Proc of the First Workshop on Current Trends in Text Simplification*, p. 70–83.

DAELEMANS W., HÖTHKER A. & KIM SANG E. T. (2004). Automatic sentence simplification for subtitling in Dutch and English. In *LREC*, p. 1045–1048.

DE BELDER J. & MOENS M.-F. (2010). Text simplification for children. In *Workshop on Accessible Search Systems of SIGIR*, p. 1–8.

DMITRIEVA A. & TIEDEMANN J. (2018). Creating an aligned Russian text simplification dataset from language learner data. In *Proc of the 8th Workshop on Balto-Slavic Natural Language Processing*, p. 73–79, Kiyv, Ukraine.

DONG Y., LI Z., REZAGHOLIZADEH M. & CHEUNG J. (2019). Editnts : An neural programmer-interpreter model for sentence simplification through explicit editing. In *Proc of the Eighth Inter Joint Conf on Natural Language Processing*, p. 3393–3402, Florence, Italy.

FLESCH R. (1948). A new readability yardstick. *Journ Appl Psychol*, **23**, 221–233.

FRANÇOIS T. & FAIRON C. (2013). Les apports du TAL à la lisibilité du français langue étrangère. *TAL*, **54**(1), 171–202.

GALA N., FRANÇOIS T. & FAIRON C. (2013). Towards a French lexicon with difficulty measures : NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In *eLEX-2013*.

GALA N., TACK A., FRANÇOIS L. J.-D. T. & ZIEGLER J. (2020). Alector : A parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers. In *Language Resources and Evaluation for Language Technologies (LREC)*, Marseille, France.

GONZALEZ-DIOS I., ARANZABE M. J. & DÍAZ DE ILARRAZA A. (2018). The corpus of basque simplified texts (cbst). *Language Resources and Evaluation*, **52**, 217–247.

GOTO I., TANAKA H. & KUMANO T. (2015). Japanese news simplification : Task design, data set construction, and analysis of simplified text. In *Proc of MT Summit XV*, p. 17–31, Miami, USA.

GRABAR N. & CARDON R. (2018). Clear – simple corpus for medical French. In *Workshop on Automatic Text Adaption (ATA)*, p. 1–11.

GRABAR N., FARCE E. & SPARROW L. (2018). Study of readability of health documents with eye-tracking and machine learning approaches. In *Int Conf on Healthcare Informatics (ICHI)*, p. 1–2. Poster.

GRABAR N. & HAMON T. (2016). A large rated lexicon with French medical words. In *LREC (Language Resources and Evaluation Conference)*, p. 1–12.

GRABAR N. & HAMON T. (2017). Understanding of unknown medical words. In *BIONLP workshop at RANLP*, p. 1–10.

GUNNING R. (1973). *The art of clear writing*. New York, NY : McGraw Hill.

HORN C., MANDUCA C. & KAUCHAK D. (2014). Learning a lexical simplifier using Wikipedia. In *Annual Meeting of the Association for Computational Linguistics*, p. 458–463.

JONNALAGADDA S., TARI L., HAKENBERG J., BARAL C. & GONZALEZ G. (2009). Towards effective sentence simplification for automatic processing of biomedical text. In *NAACL HLT 2009*, p. 177–180.

KLAPER D., EBLING S. & VOLK M. (2013). Building a German/simple German parallel corpus for automatic text simplification. In *Proc of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations*, p. 11–19, Sofia, Bulgaria.

KLERKE S. & SØGAARD A. (2012). DSim, a Danish parallel corpus for text simplification. In *LREC*, p. 4015–4018.

KOPTIENT A., CARDON R. & GRABAR N. (2019). Simplification-induced transformations : typology and some characteristics. In *Proc of the 18th BioNLP Workshop and Shared Task*, p. 309–318, Florence, Italy.

LEE S.-Y. D., STUCKY B. D., LEE J. Y., ROZIER R. G. & BENDER D. E. (2010). Short assessment of health literacy-Spanish and English : A comparable test of health literacy for Spanish and English speakers. *Health Serv Res*, **45**(4), 1105–20.

LEROY G., KAUCHAK D. & MOURADI O. (2013). A user-study measuring the effects of lexical simplification and coherence enhancement on perceived and actual text difficulty. *Int J Med Inform*, **82**(8), 717–730.

LEVENSHTEIN V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics. Doklady*, **707**(10).

LIFFRAN C. (2015). *Prise en charge institutionnelle et handicap communicationnel. Adaptation de grilles d'entretien pour accompagner les échanges avec l'adulte porteur d'une déficience intellectuelle lors des différentes étapes de son projet personnalisé*. Thèse de doctorat, Institut d'Orthophonie Gabriel Decroix, Université Lille 2, Lille.

MANCUSO J. (2009). Assessment and measurement of health literacy : An integrative review of the literature. *Nurs Health Sci*, **11**, 77–89.

MARTIN L., HUMEAN S., MAZARÉ P.-E., BORDES A., DE LA CLERGERIE E. & SAGOT B. (2018). Reference-less quality estimation of text simplification systems. In *ATA workshop*, p. 1–10.

NISIOI S., STAJNER S., PONZETTO S. P. & DINU L. P. (2017). Exploring neural text simplification models. In *Ann Meeting of the Assoc for Comp Linguistics*, p. 85–91.

OCDE (2015). *Guide de style de l'OCDE Troisième édition : Troisième édition*. OECD Publishing.

OECD (2019). *Skills matter*. Rapport interne, PIAAC, OECD.

OGDEN C. K. (1930). *Basic English : A General Introduction with Rules and Grammar*. London : Paul Treber.

PAETZOLD G. H. & SPECIA L. (2016). Benchmarking lexical simplification systems. In *LREC*, p. 3074–3080.

PAPINENI K., ROUKOS S., WARD T., HENDERSON J. & REEDER F. (2002). BLEU : a method for automatic evaluation of machine translation. In *Annual Meeting of the Association for Computational Linguistics*, p. 311–318.

PETERSEN S. & OSTENDORF M. (2007). Text simplification for language learners : A corpus analysis. In *Speech and Language Technology for Education Workshop (SLaTE)*, p. 69–72.

ROUQUETTE A., NADOT T., LABITRIE P., VAN DEN BROUCKE S., MANCINI J., RIGAL L. & RINGA V. (2018). Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. *PLoS One*, **13**(12), 1–15.

RUEL J., KASSI B., MOREAU A. & MBIDA-MBALLA S. (2011). *Guide de rédaction pour une information accessible*. Gatineau : Pavillon du Parc.

SAGGION H. (2017). *Automatic Text Simplification*, volume 32 de *Synthesis Lectures on Human Language Technologies*. University of Toronto : Morgan & Claypool.

SIDDHARTHAN A. (2014). A survey of research on text simplification. *Int J of Applied Linguistics*, **165**(2), 259–298.

SNOVER M. G., MADNANI N., DORR B. & SCHWARTZ R. (2009). TER-Plus : paraphrase, semantic, and alignment enhancements to translation edit rate. *Machine Translation*, **23**(2-3), 117–127.

SON J. Y., SMITH L. B. & GOLDSTONE R. L. (2008). Simplicity and generalization : Short-cutting abstraction in children's object categorizations. *Cognition*, **108**, 626–638.

SPECIA L., JAUHAR S. & MIHALCEA R. (2012). Semeval-2012 task 1 : English lexical simplification. In **SEM 2012*, p. 347–355.

STODDEN R. (2021). When the scale is unclear – analysis of the interpretation of rating scales in human evaluation of text simplification. In *Proc of the First Workshop on Current Trends in Text Simplification (CTTS 2021)*, p. 1–12.

STYMNE S., TIEDEMANN J., HARDMEIER C. & NIVRE J. (2013). Statistical machine translation with readability constraints. In *NODALIDA*, p. 1–12.

SÄUBERLI A., EBLING S. & VOLK M. (2020). Benchmarking data-driven automatic text simplification for German. In *Proc of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI2020)*, p. 41–48.

TANPRASERT T. & KAUCHAK D. (2021). Flesch-kincaid is not a text simplification evaluation metric. In A. FOR COMPUTATIONAL LINGUISTICS, Éd., *Proc of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, p. 1–14.

TODIRASCU A., FRANÇOIS T., GALA N., FAIRON C., LIGOZAT A.-L. & BERNHARD D. (2013). Coherence and cohesion for the assessment of text readability. In *Proc of 10th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2013)*, p. 11–19, Marseille, France.

TONELLI S., PALMERO APROSIO A. & SALTORI F. (2016). SIMPITIKI : a simplification corpus for Italian. In *Proc of Third Italian Conf on Comp Linguistics (CLiC-it 2016)*, p. 1–6.

UNAPEI (2019). *L'information pour tous*. UNAPEI.

VICKREY D. & KOLLER D. (2008). Sentence simplification for semantic role labeling. In *Annual Meeting of the Association for Computational Linguistics-HLT*, p. 344–352.

ŠTAJNER S. & POPOVIĆ M. (2016). Can text simplification help machine translation ? *Baltic J. Modern Computing*, **4**(2), 230–242.

VU T. T., TRAN G. B. & PHAM S. B. (2014). Learning to simplify children stories with limited data. In L. . SPRINGER, Éd., *Intelligent Information and Database Systems*, p. 31–41.

VÁSQUEZ-RODRÍGUEZ L., SHARDLOW M., PRZYBYŁA P. & ANANIADOU S. (2021). The role of text simplification operations in evaluation. In *Proc of the First Workshop on Current Trends in Text Simplification (CTTS 2021)*, p. 1–13.

WEI C.-H., LEAMAN R. & LU Z. (2014). SimConcept : A hybrid approach for simplifying composite named entities in biomedicine. In *BCB '14*, p. 138–146.

WOODSEND K. & LAPATA M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *EMNLP*, p. 409–420.

WUBBEN S., VAN DEN BOSCH A. & KRAHMER E. (2012). Sentence simplification by monolingual machine translation. In *Annual Meeting of the Association for Computational Linguistics*, p. 1015–1024.

XU W., NAPOLES C., PAVLICK E., CHEN Q. & CALLISON-BURCH C. (2016). Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, **4**, 401–415.

YANEVA V., TEMNIKOVA I. & MITKOV R. (2015). Accessible texts for autism : An eye-tracking study. In ACM, Éd., *Int ACM SIGACCESS Conference on Computers & Accessibility*, p. 49–57.

ZHANG X. & LAPATA M. (2017a). Sentence simplification with deep reinforcement learning. In *Conference on Empirical Methods in Natural Language Processing*, p. 584–594. DOI : 10.18653/v1/D17-1062.

ZHANG X. & LAPATA M. (2017b). Sentence simplification with deep reinforcement learning. In ACL, Éd., *Proc of the Conf on Empirical Methods in Natural Language Processing*, p. 584–594, Copenhagen, Denmark.

ZHENG J. & YU H. (2017). Readability formulas and user perceptions of electronic health records difficulty : A corpus study. *Journal of Medical Internet Research*, **19**(3), 1–15.