HECTOR: A Hybrid TExt SimplifiCation TOol for Raw Texts in French

Amalia Todirascu[1], Rodrigo Wilkens[2], Eva Rolin[2], Thomas François[2], Delphine Bernhard[1], Núria Gala[3]

[1] Université de Strasbourg, LiLPa UR 1339, F-67000 Strasbourg, France
[2] CENTAL, Université catholique de Louvain, Belgium
[3] Aix-Marseille Université, Laboratoire Parole et Langage, LPL CNRS (UMR 7309), France

{todiras, dbernhard}@unistra.fr, {rodrigo.wilkens,eva.rolin,thomas.francois}@uclouvain.be,
nuria.gala@univ-amu.fr

Abstract

Reducing the complexity of texts by applying an Automatic Text Simplification (ATS) system has been sparking interest in the area of Natural Language Processing (NLP) for several years and a number of methods and evaluation campaigns have emerged targeting lexical and syntactic transformations. In recent years, several studies exploit deep learning techniques based on very large comparable corpora. Yet the lack of large amounts of corpora (original-simplified) for French has been hindering the development of an ATS tool for this language. In this paper, we present our system, which is based on a combination of methods relying on word embeddings for lexical simplification and rule-based strategies for syntax and discourse adaptations. We present an evaluation of the lexical, syntactic and discourse-level simplifications according to automatic and human evaluations. We discuss the performances of our system at the lexical, syntactic, and discourse levels.

Keywords: Automatic Text Simplification, hybrid architecture, French corpora, evaluation metrics.

1. Introduction

Automatic Text Simplification (ATS) is a challenging Natural Language Processing (NLP) task aiming at making content more readable and understandable for specific target readers (e.g. people with learning or cognitive disorders, language learners and people with low literacy). Several systems have been proposed for English (Carroll et al., 1998) or Ibero-Romance languages (Ferrés et al., 2017) focusing on lexical and syntactic transformations. However, fewer systems yet propose discourse-level simplifications (Siddharthan, 2014; Siddharthan, 2006) and no specific end-to-end multilevel simplification tool is available for French, apart from MUSS (Martin et al., 2020b), a multilingual language model trained using sentence-level paraphrases for several languages including French.

There has been much work on writing rules for domain-specific ATS over the last two decades but the field has significantly changed during the last few years due to deep learning techniques (Al-Thanyyan and Azmi, 2021). Whereas rule-based systems require time-consuming, hand-made rules based on the study of corpora, ML approaches – and, in particular, deep learning approaches – consider the simplification as a monolingual variant of a machine translation (MT) task: simplification operations are learned from complex-simple sentence pairs, automatically extracted from available corpora (Zhu et al., 2010). However, the lack of large available parallel corpora has been hindering the development of MT-ATS systems in various languages, including French. Faced with this challenge, NLP developers have investigated two main paths for French:

- to develop a MT system by automatically translating existing corpora such as Newsela from English to French and then applying the MT system, for instance the work by Abdul Rauf et al. (2020).
- to define a rule-based system, by collecting and studying a manually simplified existing corpus for French and developing rules based on parser information and linguistic knowledge for syntax transformation (Brouwers et al., 2014b) or for pronoun resolution (Quiniou and Daille, 2018).

The first solution did not seem appropriate for our purposes: experiments with an automatically translated corpus (Newsela) showed that the BLEU and SARI metrics significantly decrease in comparison with results obtained for English (Abdul Rauf et al., 2020) and that the linguistic quality of the automatically simplified corpus is extremely degraded. We thus decided to develop a hybrid system, capitalizing on lexical resources available, and to build linguistically grounded rules for syntactic and discourse transformations. Although this approach is time-consuming, it has the advantage of being modular and interpretable, allowing us to evaluate the simplification process step by step.

With the aim to provide tools and resources for children struggling with reading in French L1 (ALECTOR project[1]), we have developed a hybrid French ATS system for raw corpora which transforms texts at three levels (lexical, syntactic and discursive), named HECTOR. HECTOR is original in several ways. We apply syntactic simplifications on dependency trees and not on constituent trees as proposed in the literature for French (Brouwers et al., 2014a). We propose a meta-language for describing the syntax transformation rules and a specific interface to develop these rules. Our

[1] https://alectorsite.wordpress.com/

discourse-level simplification module is based on coreference chains (Schnedecker, 1997) aiming to maintain text cohesion during and after simplification, following (Siddharthan, 2011). Lastly, we apply a lexical simplification module using word embeddings. As far as we know, no other end-to-end ATS system operating at all levels (lexicon, syntax and discourse) is available for French for the time being. Additionally, our system is designed for people with dyslexia, but the rule-based modules could be easily adapted to another audience: new rules might be added, some rules might be disabled. In the following sections, we first describe related work in ATS. We then present the architecture of the system and address the issue of automatic evaluation of the simplifications (we use BLEU compared to a manually simplified corpus). We also describe the evaluation results, as regards to grammaticality (fluency) (Štajner and Popović, 2019), meaning preservation and simplicity of the output, provided by human judgements. We finally propose some concluding remarks on our system.

2. Related Work

The last two decades have witnessed the publication of numerous studies within the field of ATS, reviewed by Siddharthan (2014), Shardlow (2014), Saggion (2017) and Al-Thanyyan and Azmi (2021). In brief, the field has mainly focused on developing methods to automatically simplify complex words (lexical simplification) and/or complex syntactic structures (syntactic simplification). Historically, ATS systems have first relied on rule-based approaches (Chandrasekar et al., 1996; Siddharthan, 2011) in which a text is automatically parsed before simplification rules defined by experts are applied. Later, ATS has been assimilated to a translation task (the original version is translated into a simplified version) and addressed with statistical translation systems (Specia, 2010; Zhu et al., 2010; Woodsend and Lapata, 2011). As neural machine translation has emerged under the impulse of deep learning, the Seq2Seq model has also become prevalent for ATS (Nisioi et al., 2017; Zhang and Lapata, 2017). Besides, hybrid methods which combine several approaches have also been reported in (Siddharthan and Angrosh, 2014; Narayan and Gardent, 2014; Maddela et al., 2021).

Although ATS has become a rather active field, as in other many NLP fields most of the work has been focused on English, especially because of the amount of available resources, most notably the Simple English Wikipedia. While some work has also been developed for languages like Spanish, Portuguese, Basque or even Japanese (see the above mentioned surveys), French has been hardly researched. Recently, one end-to-end system have been designed: MUSS (Martin et al., 2020b) for poor readers and some experiments dedicated to medical texts Cardon and Grabar (2020). However, these systems and experiments still show limitations for operational purposes and have trouble producing good quality French texts to be read by struggling readers or to be used for other NLP applications.

Work on ATS in the last decade is also not immune to criticism from the research community. The way simplifications are traditionally evaluated, using automatic metrics such as BLEU or the Flesch-Kincaid Reading Grade Level, has started to be questioned (Sulem et al., 2018; Tanprasert and Kauchak, 2021; Alva-Manchego et al., 2021). The purposes of ATS research have also been challenged: Stajner (2021) discusses the lack of consideration for the characteristics of potential target populations and calls for the development of more modular ATS systems, which can be customised for specific populations. Some of the most recent approaches propose systems able to control specific properties of the simplified output, such as Maddela et al. (2021) or Sheang and Saggion (2021) for English and Martin et al. (2020a) for French. However, being based on the Seq2Seq model, such systems do not allow the same amount of control than a rule-based system. In the following section, we present the architecture of HECTOR, a hybrid ATS system for French, and its customisation for French-speaking children.

3. HECTOR's Architecture

HECTOR's architecture is composed of four modules (preprocessing, syntax, discourse and lexical simplification), as illustrated in Figure 1. These modules are based on the contrastive study of original and manually simplified pairs of French texts targeting dyslexic and poor readers (Wilkens and Todirascu, 2020).

The preprocessing module aims to provide the additional information required by the syntax, discourse and lexical simplification modules. This is done by using the Stanza dependency parser[2] (Qi et al., 2018) and CoFR[3] (Wilkens et al., 2020a), a coreference annotation system for French (operating at the discourse level). Moving forward in the pipeline, the syntax simplification module aims to suppress secondary information and standardises the sentences into Subject-Verb-Object (SVO) structures (thus suppressing long introductory participial clauses). It also implements standard transformations (e.g., sentence splitting with relative pronouns and coordination conjunctions, deletion of initial modifiers to warrant SVO order, deletion of appositive or parenthetical constituents, etc.) as detailed in Gala et al. (2020b). This module uses 10 syntax transformation rules. The third module focuses on cohesion text markers such as coreference chains, as proposed by Wilkens and Todirascu (2020)[4]. It also modifies the output of the syntax simplification module, applying 4 discourse transformation rules. At the

[2]In this work, we use the Stanza version 1.2.0 and the model trained on version 2.7 of UD_French-GSD (Guillaume et al., 2019).

[3]https://github.com/boberle/cofr

[4]Syntax and discourse modules are available : https://github.com/rswilkens/text-rewrite

word level, the lexical simplification module aims to identify and replace complex words, while the morphological simplifications[5] change "unusual" verb tenses into more common ones whenever is possible (i.e. past into present tense in descriptive parts of narrative texts). Whereas the lexical simplification module takes advantage of pre-trained models (Section 3.3), the rules implemented in the syntactic (Section 3.1) and discourse modules (Section 3.2) were based on the study of a corpus. It is composed of original and simplified narrative texts addressed to dyslexic children (Methodolodys[6]), detailed in (Wilkens and Todirascu, 2020). In this development corpus, we manually identified the simplification operations at the syntactic and discourse levels. For the evaluation, we compared the system's output with a sample of sentences from ALECTOR, a manually simplified corpus for French containing original literary (tales and stories) and scientific (documentary) texts along with their simplified equivalents (Gala et al., 2020a). The texts are graded according to three levels related to their use in 2nd (CE1), 3rd (CE2) and 4th (CM1) grades of primary school (7 to 10 years old). We selected the 2nd and the 4th levels to have simple and complex texts in the reference corpus (Section 4).

3.1. Syntactic Simplification Module

The syntactic adaptations presented by Gala et al. (2020b) may be grouped into three categories:

Secondary information suppression: Aiming to reduce the length of the sentences. The adverbial, past and present participle clauses are removed.
e.g. *Perdus dans la nature, les enfants pleuraient.* → *Les enfants pleuraient.*
In English, *Lost in the nature, the children wept.* → *The children wept.*

Sentence structure adjustments: Transforming the passive and cleaved sentences into standard SVO order sentences. In particular, pronouns and auxiliary verbs are cut out in the cleaved sentences. The passive voice is transformed into active voice form – the agent becomes the subject and the verb form changes:
e.g. *Les brebis ont été mangées par le loup.* → *Le loup a mangé les brebis.*
In English, *The sheep have been eaten by the wolf.* → *The wolf ate the sheep.*

Sentence splitting: Relative clauses are extracted and transformed into main clauses while the phrases linked by conjunctions are split.
e.g. *Les enfants jouaient et se moquaient du loup.* → *Les enfants jouaient. Les enfants se moquaient du loup.*
In English, *The children played and made fun of the wolf.* → *The children played. The children made fun of the wolf.*

[5]Lexical and morphosyntactic levels are presented together in Figure 1.

[6]https://methodolodys.ch/

The 10 syntactic transformation rules were developed using SemGrex (Levy and Andrew, 2006) and an extension of it based on Tsurgeon targeting text simplification (Wilkens and Todirascu, 2020). In addition, we extended this tool including a high-level transformation language, which simplifies the design of the rules. Each rule checks for the dependency and the morphosyntactic information match (expressed by a Semgrex pattern), and operates the modifications on the dependency trees (e.g. deleting nodes, moving subtrees, inserting new nodes, splitting trees, and replacing nodes and tag information).

An example of the included high-level transformation language developed is presented in Figure 2. The example shows a rewriting rule for a cleaved sentence, with a Semgrex expression checking the matching with the actual text. The activation condition (line 2) identified several dependencies from the root node (`{$}=root`), governor of a *cop* dependency, of an auxiliary *être 'to be'* and of the subject (`>nsubj`) (a pronoun (`{tag:PRON=ce}` which is labelled *"ce"*), and yet, there is a secondary clause linked by *aclrel* dependency from the object of the main verb and containing a relative pronoun *que 'that'*. In Figure 2, `{tag:VERB}=verb >>obj {word:que}=pron` means that the *verb* node is the ancestor of the node *pron* and at least one dependency is labelled *obj*. When a sentence matches with this pattern, the system deletes the auxiliary and pronouns (line 3), then the second verb is used to split the initial sentence (line 4). Furthermore, it is possible to conditionally apply transformations based on the occurrence of elements indicated as optional in the search patterns (line 6 - 8) as well as to replace specific parts of the tag information, as indicated in line 9 where the dependency tag is replaced by the tag *root* and in line 10 where the dependency tag is replaced by *verb*.

3.2. Discourse Simplification

Syntactic simplifications might suppress important information for textual cohesion: suppressing pronouns or some secondary clauses might cut or mix up the coreference chains. Coreference chains are composed of referring expressions (or mentions) – proper nouns, noun phrases (NP) and pronouns (Schnedecker, 1997). Splitting sentences or other syntactic transformations might add extra pronouns or noun phrases which makes the task of decoding the text more difficult. Moreover, ambiguous pronouns require an important amount of inferences from the reader to identity the correct discourse entity. Hence, we carefully designed the system to adjust the coreference chains during the syntactic transformations. In order to reduce the amount of these inferences, we apply discourse simplification rules maintaining the structure of coreference chains.

The discourse simplification module described in this work was originally presented by Wilkens and Todirascu (2020), and it is based on the Accessibility theory

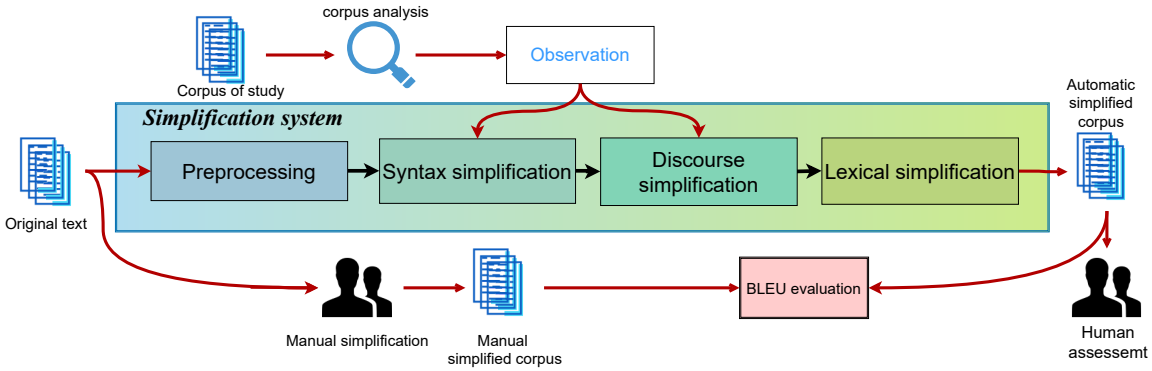

Figure 1: The simplification methodology and the proposed architecture

```
operation cleavedQUE begin
search sent: "{$}=root ?>det{}=det >cop {}=etre >
    nsubj {tag:PRON}=ce >/acl:relcl/ ({tag:VERB}=
    verb >>obj {word:que}=pron)"
sent = delete ce, etre, pron from sent
sent1, sent2 = split verb from sent
sent1 = insert root -0* sent2 from sent1
if det begin
  sent1 = replace det (FORM, LEMMA, FEATS) '1;ce;ce
      ;;PronType=Dem;0;' from sent1
end
root = replace root(DEPENDENCY) '1;;;;;0;obj' from
    root
verb = replace verb(DEPENDENCY) '1;;;;;0;root' from
    verb
save sent1
operation end
```

Input: C'est Marie que nous visitons aujourd'hui. (This is Mary that we visit today.)
Output: Nous visitons Marie aujourd'hui. (We visit Mary today.)

Figure 2: Example of a syntactic simplification rule

(Ariel, 1990). This theory classifies the accessibility of referring expressions (i.e., proper nouns used to introduce new entities into discourse) from low to high using the inference required for anaphora resolution. This property is used to propose the replacement of highly accessible mentions (such as pronouns) with low accessible referring expressions (proper nouns). The 4 simplification rules are grouped in three main categories and are applied in the following order:

Replace new or repeated entities: Explicit referents replace the ambiguous pronouns or successive pronouns in the same chain (2 rules). This operation reduces the quantity of processing inferences done by the reader to identify the correct referent.
e.g. *Le loup* va et vient. *Il* fixe le garçon. → *Le loup* va et vient. *Le loup* fixe le garçon.
In English, *The wolf* goes back and forth. *It* stares at the boy. → *The wolf* goes back and forth. *The wolf* stares at the boy.
In this example, the ambiguous pronoun ("il", *it* in English) is replaced by its antecedent (*le loup*, *the wolf* in English).

Specify entities: New entities are introduced by either an indefinite noun phrase (NP) or a proper noun while definite noun phrases (containing a definite article or a demonstrative determiner) refer to known entities. The change of determiner for a more highly accessible one modifies the accessibility of the referring expression. According to Ariel's accessibility theory (Ariel, 1990), demonstrative NPs are more accessible than definite or indefinite articles, so we replace the demonstrative determiner by a definite one in order to have a less accessible referring expression.
e.g. *ce hérisson* → *le hérisson*.
In English, this hedgehog → the hedgehog

Make NP more accessible: Possessive NPs are replaced by their explicit referent, such as a proper noun or a complete definite NP (automatically computed by CoFR (Wilkens et al., 2020b)).
e.g. *Sa grand-mère* → *la grand-mère du Chaperon rouge.*
In English, *Her grandmother* → *The grandmother of Little Red Riding Hood.*

These 4 rules (implemented in Java) check the cohesion errors due to syntactic simplification and solve the referential ambiguity created in the previous simplification step, before the lexical simplification applies.

3.3. Lexical Simplification

The lexical simplification module, whose purpose is to replace words identified as complex with simpler equivalents, consists of 4 steps, as proposed by Shardlow (2014). For our lexical simplification pipeline, we follow Rolin et al. (2021) who proposed a benchmarking of techniques for lexical simplification for French. We used some of these methods and resources according to the benchmarking results, but also to our needs.

Complex Word identification: In Rolin et al. (2021), no detection of complex words was performed, but other studies have investigated various criteria to detect them. In older work, it can simply be a threshold applied on word frequencies (Biran

et al., 2011) or the fact that the word is present in a list of complex words (Chen et al., 2016; Deléger and Zweigenbaum, 2009). In more recent work, dedicated complex word identification systems have been trained (Billami et al., 2018; Yimam et al., 2018; Alarcon et al., 2019), sometimes relying on ensemble-based models (Malmasi and Zampieri, 2016; Paetzold and Specia, 2016) or on deep learning and transformers (Yaseen et al., 2021; Rao et al., 2021). In our case, as it is hard to find available training data for French, we used a combination of various well-established word features. Precisely, a word is considered as complex if (1) its frequency is lower than 5 per million, based on Lexique3 (New et al., 2007) AND (2) it also meets at least one of the following criteria:

- it is missing from the Manulex list of simple words (Lété et al., 2004);
- word length > 7;
- an etymological letter is detected (as defined in Gala and Ziegler (2016));
- a complex inconsistent letter is detected (as defined in Gala and Ziegler (2016));
- the distance between phonemic and orthographic forms is higher than 2. Phonemic forms were obtained from Lexique3 (New et al., 2007) or, for missing entries, computed by `espeak`.[7]

Substitution Generation: This step consists in generating synonyms for the word identified as complex in the previous step. There are several ways to generate synonyms: using lexical resources, word-embeddings or language models. Recently, Rolin et al. (2021) have shown that using a lexical resource provides better results than other approaches if we look at lemma level.

However, in this research, our aim is to provide a simplified and grammatical sentence, i.e. to take into account the inflection of nouns, adjectives and verbs when replacing the complex word. If we use a lexical resource such as *Resyf* (Billami et al., 2018) we only obtain the lemmas, so we need to include an inflection step in addition to the other steps: unlike English which has several inflection modules (LemmInflect,[8] MorphAdorner,[9] inflection,[10] etc.), we have not found any modules offering the same kind of service for French.

Therefore, we use *FastText* (Bojanowski et al., 2017) which allows us to propose synonyms which are already inflected as shown in Glavaš and Štajner (2015). We use character n-gram embeddings to get a vector representation even for a word that does not exist in the training corpus. Thanks to this technique, we return, for any given complex word, its k-nearest semantic neighbors based on cosine similarity (Rolin et al., 2021).

Substitution Selection: The third step is the disambiguation step, in which we choose the synonyms that are suitable in context. We already provide disambiguated synonyms since we choose the closest semantic neighbours of the complex word. However, we have chosen to perform a second filtering by retaining only those synonyms which have the same part of speech as the word to be substituted. We obtained the parts of speech from the *Delaf* dictionary (Courtois, 1990).

Substitution Ranking: It is the final stage of lexical simplification which ranks synonyms according to their difficulty. We chose to use the frequencies provided by *Lexique3* (New, 2006) as an indicator of difficulty: the more frequent a word is in a corpus, the simpler it is because it is supposed to be known by the reader.

4. Evaluation

In this section, we first assess our system using automated measures, following the method in Alva-Manchego et al. (2020), and then describe the results of the human evaluation.

4.1. Automatic Evaluation

To automatically evaluate our system, we have used the standard measure BLEU (Papineni et al., 2002) that relies on the proportion of n-gram matches between a system's output and the reference corpus ALECTOR. We compared the BLEU score for the complete texts (all the sentences in a document are used to compute the BLEU score) and for the pairs of original and simplified sentences[11] as shown in Table 1.

The results of the evaluation show that BLEU scores are higher for CE1 (2nd grade) than for CM1 (4th grade). Generally, the texts for the 2nd grade are already quite simple and few transformations are applied. Therefore, the BLEU score is higher in part due to the smaller number of changes required. The texts for the 4th grade are more complex, the sentences are longer and they present more varied syntactic structures: splitting sentences or replacing the pronouns is frequent and the transformations may result into a slightly different output as regards to the original input, which may explain lower BLEU scores.

4.2. Human Evaluation: Methodology

As BLEU scores provide few clues about the actual behavior of ATS systems, we also run a human evaluation campaign to analyze the results obtained by the modules presented in Section 3. Each level (i.e., syntax, discourse and lexical simplification) was evaluated sepa-

[7] `http://espeak.sourceforge.net/`

[8] `https://github.com/bjascob/LemmInflect`

[9] `http://morphadorner.northwestern.edu/morphadorner/`

[10] `https://pypi.org/project/inflection/`

[11] Sentences without transformation in both manual and automatic simplified versions are ignored.

Corpus	Genre	Doc. level	Sent. level
CM1 (4th)	SCI	.54 (.11)	.51 (.24)
	LIT	.64 (.10)	.60 (.28)
CE1 (2nd)	SCI	.76 (.06)	.78 (.24)
	LIT	.73 (.03)	.64 (.29)
All	-	.67 (.12)	.62 (.28)

Table 1: The average BLEU score and standard deviation by grade and genre at document/sentence levels.

rately by a group of experts. At each step, we evaluated the specific transformations performed by the system.
The corpus used for the manual evaluation was extracted from a sample of original texts (207 sentences) from ALECTOR . We evaluated only the simplified sentences for each level, ignoring those without changes: 109 sentences for syntax transformations, 41 sentences for the discourse level and 48 sentences for lexical transformations. Each output was evaluated by several experts and a measure of inter-annotator agreement was calculated for every level. The errors were annotated according to guidelines based on three standard criteria in ATS: *fluency* (the sentence is grammatically correct), *meaning preservation* (the original information is kept in the simplified sentence) and *simplicity* (the generated output is simpler than the original). The syntax and discourse transformations were compared with the original sentences: discourse simplification was applied on the output of the syntax module. The lexical transformations were compared to the discourse output, to allow annotators to better focus only on vocabulary adaptations. We used a Likert-scale going from 1 (the worst score) to 5 (the best score) for the annotations. In the evaluation guidelines, we defined the number of errors corresponding to a score in the Likert-scale, e.g. a score of 4 was given to a sentence with only one grammatical error (sentence length of 15 words in average) and 3 if there were up to three errors. *Simplicity* was assessed as follows: 1 when there were several transformations that made the sentence more complex, 2 for one single transformation which added complexity, 3 if the sentence showed the same complexity as the original, 4 if a single transformation simplified the sentence and 5 when several transformations simplified the original input.

4.3. Human Evaluation: Results

We computed inter-annotator agreement with Krippendorf's α for the three criteria and linguistic levels (our three linguistic modules). The syntax and discourse outputs were annotated by three experts and lexical output by two experts.
While *fluency* shows better agreement scores (syntax: 0.738, discourse: 0.632 and lexical: 0.457) , not surprisingly, the results are lower for *meaning preservation* (syntax: 0.58, discourse: 0.265 and lexical: 0.455) and *simplicity* (syntax: 0.485, discourse: 0.29 and lexical: 0.369). *Meaning preservation* and *simplicity* are subjective tasks, opinions may be more often different among experts. It comes also without surprise that the best scores are obtained for syntax (which is the first module applied for the transformations) while the agreement scores decrease for the other levels (the errors from one level influence the quality of the simplification of the following levels). The output from the discourse and lexical module shows the lowest scores for *meaning preservation* and *simplicity*. The experts frequently disagree for the determiner transformations (*'un'* replaced by *'ce'* or *'le'*, 'a' by 'this' ot 'the') thus considering there is a loss of information. For *simplicity*, experts often disagree by considering the determiner changes as 2 (more complex) or 3 (as complex as the original).
We have also computed the average of the Likert-scale grades given by the experts for each evaluation criterion. Based on these averages, we have focused on the proportion of successful transformations that preserve meaning, fluency and make the text simpler (score $\geq$3.5). We represent the scores for the three criteria in Tables 2 to 4 (to improve table readability, we represent only the categories that indicate improvement). Overall, the experts have given higher scores for the three criteria, which means that they considered the fluency and the meaning preserved, and the output sentences simpler than the original ones. We discuss the results obtained for each criterion.
Meaning preservation (Table 2). For syntax simplification, we obtained 70.6 % of sentences graded with a score greater or equal to 3.5 for *meaning preservation*. The results obtained for discourse are better (with a total of 83.34 %), except for the *CE1 Litt* texts. For syntax and discourse, texts from the 4th grade obtain a larger percentage of correctly simplified sentences than the 2nd grade. The lexical level obtained the worst results, especially for documentary texts, and for 4th grade (see Table 2).

	Syntax	Discourse	Lexical
CE1 Litt (2nd)	80.95%	70.00%	66.67%
CE1 Sci (2nd)	50.00%	93.33%	46.00%
CM1 Litt (4th)	73.68%	88.23%	38.88%
CM1 Sci (4th)	77.77%	88.23%	40.62%
Total	70.60%	83.34%	48.04%

Table 2: *Meaning preservation* (score $\geq$ 3.5) for syntax, discourse and lexical simplification.

Fluency (grammaticality) being an objective criterion, the results are higher than for the two other criteria, for all the levels (see Table 3). 87.13 % of sentences were considered grammatically correct or almost correct (score greater or equal to 3.5) at lexical level, followed by 85.19 % for syntax and 82.31 % for discourse.
For **simplicity** (Table 4), 79.36% of sentences are simpler than the original (score $\geq$ 3.5) for syntax, but the results are dramatically decreasing to 59.29 % for lexi-

	Syntax	Discourse	Lexical
CE1 Litt (2nd)	91.89%	87.50%	86.60%
CE1 Sci (2nd)	69.23%	73.33%	73.73%
CM1 Litt (4th)	90.00%	100%	94.44%
CM1 Sci (4th)	89.65%	68.42%	93.75%
Total	85.19%	82.31%	87.13%

Table 3: *Fluency* (score $\geq$ 3.5) for syntax, discourse and lexical simplification.

cal and to 50.90 % for discourse level. For both literary texts, we obtained 84.21% sentences with a good score ($\geq$ 3.5) for syntax and at least 70% for discourse. For the documentary texts of 2nd grade, the results are the worst (50% of sentences are considered simpler than the original for syntax but only 27.27% for discourse). For documentary texts at discourse level, we obtained only 27.27 % and respectively, 23.53 % of sentences considered as simpler than the original. This may be a consequence of some coreference errors when replacing pronouns or definite noun clauses or some determiner changes. Not surprisingly, low results were obtained at the lexical level (40.66%, and 46.66% respectively), in particular for documentary texts where the terminology is very accurate and lexical substitution is not always possible.

	Syntax	Discourse	Lexical
CE1 Litt (2nd)	84.21%	70.00%	86.00%
CE1 Sci (2nd)	50.00%	27.27%	46.66%
CM1 Litt (4th)	84.21%	86.66%	66.66%
CM1 Sci (4th)	82.35%	23.53%	40.62%
Total	79.36%	50.90%	59.29%

Table 4: *Simplicity* (score $\geq$ 3.5) for syntax, discourse and lexical simplification.

We illustrate the variation of *simplicity* across level and genre in Figure 3. For syntax, the average range value is between 4 and 5, which means that most of the simplified sentences are considered by the experts simpler than the original (except for documentary texts of the 2nd grade). For discourse level, more than 50% of the values are included in a range between 3 et 4 and the literature obtain better score than documentary genre (the values are around 3). For the lexical level, most of the values are between 3 and 4 (with an exception for CM1 4th documentary texts), the best scores are obtained for the literary corpora. The simplification of documentary texts is often misleading (replacement of a term by an inappropriate one, changes of the determiner which may result in a more complex clause than the original.)

Finally, we evaluate the sentences satisfying simultaneously the three criteria for each level. We obtain the best results for syntax (except for 2nd CE1 Sci): 73.68 % (2nd CE1 Litt), 12.5% (2nd CE1 Sci), 63.15 % (4th CM1 Litt), and 76.47% (4th CM1 Sci). The percentages are lower for discourse: 25.00 % (CE1 Litt), 27.27% (CE1 Sci), 84.61 % (CM1 Litt), and 23.53% (CM1 Sci). For lexical simplification, we obtain 60.00 % (CE1 Litt), but only 20.00 % (CE1 Sci), 27.27 % (CM1 Litt), and 21.87 % (CM1 Sci). The results are very heterogeneous: we observe important variation between genre or grade for the same level. The best results and the worst results are not always those expected (for discourse, the best result is obtained for CM1 Litt), with complex texts.

4.4. Error Analysis

In this final part of our assessment, we discuss some specific errors made by the system at each level.

Syntax simplification errors. Among the errors that we have identified, the most frequent one concerns wrong word order. In the following example, the adverb (*aussitôt 'soon'*) and the object are misplaced, the latter appears before the main verb in the second sentence (*de ces fruits sublimes/'with the sublime fruits'*):
e.g. *Aussitôt les villageois se précipitent vers l'arbre et se gorgent de ces fruits sublimes. 'Immediately the villagers rush to the tree and stuff themselves with the sublime fruits.' → *Les villageois Aussitôt se précipitent vers l'arbre. De ces fruits sublimes Les villageois se gorgent. 'The villagers immediately rush to the tree. *with the sublime fruits the villagers stuff themselves .'*
In some cases, several syntactic transformations may produce a cascade of errors. In the following example, the relative clause is extracted and the passive voice of the main verb (*était réduit/'is reduced'*) is transformed into the active voice, yet the verb form is incorrect (wrong number agreement).
e.g. *Le salsola kali, plante de la steppe, était réduit en cendres par les bédouins qui en assuraient le transport jusqu'aux savonneries. /'The salsola kali, a steppe plant, was reduced to ashes by the Bedouins who transported it to the soap factories.' → *Les bédouins Le salsola kali, plante de la steppe, réduit en cendres. '*The Bedouins The salsola kali, a plant of the steppe, reduced to ashes.'*

Discourse errors. The replacement of determiners (e.g., *notre → le*) might in some cases change the meaning of the sentence (or the referent) or introduce a grammar error:
e.g. *Les algues produisent plus de la moitié de l'oxygène de notre air./'Algae produce more than half of the oxygen in our air.' → Les algues produisent plus de la moitié de l'oxygène **du** air. /'Algae produce more than half of the oxygen in the air.'*
In this example, there is also a grammatical error (the correct form would be **de l'air**).
Sometimes, the pronoun is replaced by the referent computed by the coreference identification module CoFR. In some contexts, the definite NP or the proper noun replace the pronoun, but the result is ungrammat-

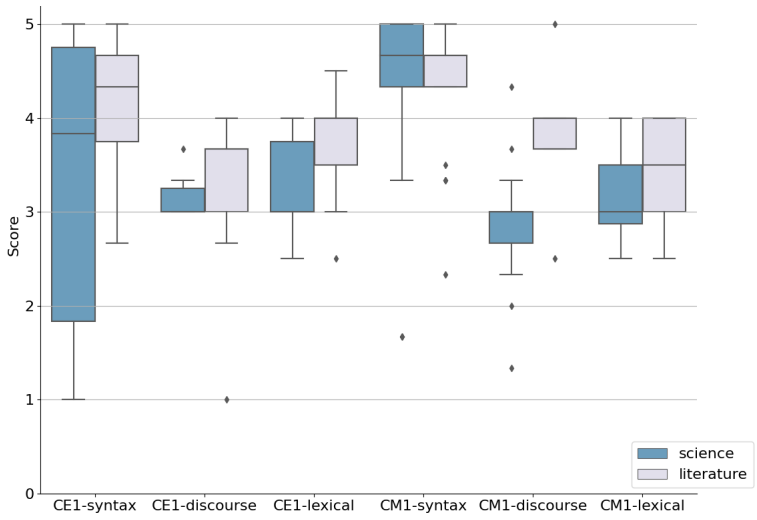

Figure 3: A comparison of the simplicity across levels and genre

ical (the indirect object *lui/'her'* should be replaced in this context by *à la grand-mère/'to the grandmother'*): e.g. *Il lui demanda où elle allait;/'He asked her where she went;'* → **Il la mère-grand demanda où la mère-grand allait ;/*'He asked the grand mother where the grand mother went;'*
Moreover, in this last example, the referent is not correctly identified (the referent is Little Red Riding Hood, not her grandmother).

Lexical errors. Our system produces erroneous outputs when dealing with multi-word expressions. In the following example, *'tomber à la renverse'* (fall backwards) cannot be simplified by substituting only one word: the whole expression has to be replaced:
e.g. *Hugo eut à peine le temps de bondir hors du lit qu'il tomba à la renverse en criant de frayeur (...). /'Hugo barely had time to leap out of bed before he fell backwards screaming with fright (...).'* → **Hugo eut à peine le temps de bondir hors de le lit qu'il tomba à la tombe en criant de frayeur (...) ./'Hugo barely had time to leap out of bed before he fell to the grave screaming with fright (...)'*
Besides, as mentioned before, lexical substitution in documentary (scientific) texts entails important loss of meaning (scientific text use specific terminology that cannot easily be replaced by synonyms or hyperonyms): *Le gypse est une roche tendre./'Gypsum is a soft stone'* → **Un plâtre est une roche tendre./'A plaster is a soft stone)'*
Finally, we can find antonyms among the synonym candidates. This error is explained by the fact that two words in the vector space are two words that share a similar context, but they are not necessarily words that share a meaning as in the following example: *Il est omnivore même s' il a une réputation d' être essentiellement charognard. 'It is omnivorous, although it has a reputation for being mainly scavengers.'*→ **Il est végétarien même s' il a une réputation d' être essentiellement charognard. 'It is a vegetarian although it has a reputation for being primarily scavenger.'*

5. Conclusion

In this paper, we present HECTOR, an end-to-end ATS system for French that combines a rule-based approach for syntax and discourse simplification and an embedding-based substitution component for lexical simplification. Although inter-rater agreement scores highlight the difficulty of the annotation task (e.g., subjectivity of the experts and difficulties to separate the criteria), we were able to show that syntactic transformation made by HECTOR produce good simplifications, while the results decrease at discourse and lexical levels.
In future work, we plan to improve the system taking into account the feedback provided by the fine-grained error analysis performed during the annotation campaign. In particular, we will focus on word order preservation and multi-word integration. As the evaluation was carried out on texts addressed to learners of French (first grades of elementary schools), another objective is to evaluate our improved system on French texts for other target readers, but also for poor readers such as people with dyslexia.

6. Acknowledgements

This research has been funded by the ALECTOR project (Reading tools for dyslexic children and poor readers), from the French National Research Agency (ANR) (ANR-16-CE28-0005), and by the Federation Wallonie-Bruxelles in relation with the AMesure project. We thank Léonie Bouchwalder, Karim El-Haff, Albin Brogialdi, Lara Dunuan, and Yaya Sy for their collaboration on the syntactic and discourse modules. We also thank Rita Hijazi and Nina Cornet Fontaine for collaborating with the authors on the annotation campaign.

7. Bibliographical References

Abdul Rauf, S., Ligozat, A.-L., Yvon, F., Illouz, G., and Hamon, T. (2020). Simplification automatique de texte dans un contexte de faibles ressources. In Christophe Benzitoun, et al., editors, *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)*, pages 332–341, Nancy, France. ATALA.

Al-Thanyyan, S. and Azmi, A. (2021). Automated text simplification: A survey. *Computational Surv.*, 54(2):36.

Alarcon, R., Moreno, L., Segura-Bedmar, I., and Martinez, P. (2019). Lexical simplification approach using easy-to-read resources. *Procesamiento de Lenguaje Natural*, 63:95–102, 10.

Alva-Manchego, F., Scarton, C., and Specia, L. (2020). Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, 46(1):135–187.

Alva-Manchego, F., Scarton, C., and Specia, L. (2021). The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification. *Computational Linguistics*, 47(4):861–889, December.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge.

Billami, M., François, T., and Gala, N. (2018). ReSyf: a French lexicon with ranked synonyms. In *Proceedings of COLING 2018*, pages 2570–2581, Santa Fe, New Mexico, USA.

Biran, O., Brody, S., and Elhadad, N. (2011). Putting it simply: a context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 496–501. Association for Computational Linguistics.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Brouwers, L., Bernhard, D., Ligozat, A.-L., and François, T. (2014a). Syntactic sentence simplification for french. In *Proceedings of the 3rd PITR Workshop)*.

Brouwers, L., Bernhard, D., Ligozat, A.-L., and François, T. (2014b). Syntactic sentence simplification for french. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 47–56.

Cardon, R. and Grabar, N. (2020). French biomedical text simplification: When small and precise helps. In *The 28th International Conference on Computational Linguistics*.

Carroll, J., Minnen, G., Canning, Y., Devlin, S., and Tait, J. (1998). Practical simplification of english newspaper text to assist aphasic readers. In *Proceedings of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10.

Chandrasekar, R., Doran, C., and Srinivas, B. (1996). Motivations and methods for text simplification. In *Proceedings of the 16th conference on Computational Linguistics*, volume 2, pages 1041–1044.

Chen, J., Zheng, J., and Yu, H. (2016). Finding important terms for patients in their electronic health records: A learning-to-rank approach using expert annotations. *JMIR Medical Informatics*, 4:e40, 11.

Courtois, B. (1990). Un système de dictionnaires électroniques pour les mots simples du français.

Deléger, L. and Zweigenbaum, P. (2009). Extracting lay paraphrases of specialized expressions from monolingual comparable medical corpora. In *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora (BUCC)*, pages 2–10.

Ferrés, D., Saggion, H., and Gómez Guinovart, X. (2017). An adaptable lexical simplification architecture for major ibero-romance languages. In *Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems*, pages 40–47.

Gala, N. and Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pages 59–66.

Gala, N., Tack, A., Javourey-Drevet, L., and François, T. (2020a). ALECTOR: A parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers. In *Language Resources and Evaluation Conference (LREC 2020)*, Proceedings of the 12th Edition of the Language Resources and Evaluation Conference (LREC 2020), page 1353-1361, Marseille, France.

Gala, N., Todirascu, A., Bernhard, D., Wilkens, R., and Meyer, J.-P. (2020b). Transformations syntaxiques pour une aide à l'apprentissage de la lecture : typologie, adéquation et corpus adaptés. In *Congrès Mondial de Linguistique Française (CMLF 2020)*, Montpellier, France.

Glavaš, G. and Štajner, S. (2015). Simplifying lexical simplification: Do we need simplified corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 63–68, Beijing, China, July. Association for Computational Linguistics.

Guillaume, B., de Marneffe, M.-C., and Perrier, G. (2019). Conversion et améliorations de corpus du français annotés en universal dependencies. *Traitement Automatique des Langues*, 60(2):71–95.

Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). Manulex : A grade-level lexical database

from French elementary-school readers. *Behavior Research Methods, Instruments and Computers*, 36:156–166.

Levy, R. and Andrew, G. (2006). Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *LREC*, pages 2231–2234.

Maddela, M., Alva-Manchego, F., and Xu, W. (2021). Controllable Text Simplification with Explicit Paraphrasing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3536–3553, Online. Association for Computational Linguistics.

Malmasi, S. and Zampieri, M. (2016). Maza at semeval-2016 task 11: Detecting lexical complexity using a decision stump meta-classifier. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 991–995.

Martin, L., De La Clergerie, É. V., Sagot, B., and Bordes, A. (2020a). Controllable sentence simplification. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4689–4698.

Martin, L., Fan, A., de la Clergerie, É., Bordes, A., and Sagot, B. (2020b). Muss: Multilingual unsupervised sentence simplification by mining paraphrases. *arXiv preprint arXiv:2005.00352*.

Narayan, S. and Gardent, C. (2014). Hybrid simplification using deep semantics and machine translation. In *Proceedings of ACL 2014*, pages 435–445.

New, B., Brysbaert, M., Veronis, J., and Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(04):661–677.

New, B. (2006). Lexique 3: Une nouvelle base de données lexicales. *TALN Conference*, pages 892–900, 01.

Nisioi, S., Štajner, S., Ponzetto, S. P., and Dinu, L. (2017). Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: ShortPapers)*, pages 85—-91.

Paetzold, G. and Specia, L. (2016). Sv000gg at semeval-2016 task 11: Heavy gauge complex word identification with system voting. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 969–974.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Qi, P., Dozat, T., Zhang, Y., and Manning, C. D. (2018). Universal dependency parsing from scratch. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 160–170, Brussels, Belgium, October. Association for Computational Linguistics.

Quiniou, S. and Daille, B. (2018). Towards a Diagnosis of Textual Difficulties for Children with Dyslexia. In *11th International Conference on Language Resources and Evaluation (LREC)*, Miyazaki, Japan, May.

Rao, G., Li, M., Hou, X., Jiang, L., Mo, Y., and Shen, J. (2021). Rg pa at semeval-2021 task 1: A contextual attention-based model with roberta for lexical complexity prediction. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 623–626.

Rolin, E., Langlois, Q., Watrin, P., and François, T. (2021). FrenLyS: A tool for the automatic simplification of French general language texts. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1196–1205, Held Online, September. INCOMA Ltd.

Saggion, H. (2017). Automatic text simplification. *Synthesis Lectures on Human Language Technologies*, 10(1):1–137.

Schnedecker, C. (1997). *Nom propre et chaînes de référence*. Centre d'Études Linguistiques des Textes et Discours de l'Université de Metz.

Shardlow, M. (2014). A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1):58–70.

Sheang, K. C. and Saggion, H. (2021). Controllable Sentence Simplification with a Unified Text-to-Text Transfer Transformer. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 341–352.

Siddharthan, A. and Angrosh, M. (2014). Hybrid text simplification using synchronous dependency grammars with hand-written and automatically harvested rules. In *Proceedings of EACL 2014*, pages 722–731.

Siddharthan, A. (2006). Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1):77–109.

Siddharthan, A. (2011). Text simplification using typed dependencies: A comparison of the robustness of different generation strategies. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 2–11.

Siddharthan, A. (2014). A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.

Specia, L. (2010). Translating from Complex to Simplified Sentences. In *Proceedings of Propor 2010.*, pages 30–39.

Štajner, S. and Popović, M. (2019). Automated text simplification as a preprocessing step for machine translation into an under-resourced language. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1141–1150.

Stajner, S. (2021). Automatic Text Simplification for

Social Good: Progress and Challenges. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2637–2652, Online. Association for Computational Linguistics.

Sulem, E., Abend, O., and Rappoport, A. (2018). BLEU is Not Suitable for the Evaluation of Text Simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 738–744, Brussels, Belgium, October. Association for Computational Linguistics.

Tanprasert, T. and Kauchak, D. (2021). Flesch-Kincaid is Not a Text Simplification Evaluation Metric. In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 1–14.

Wilkens, R. and Todirascu, A. (2020). Simplifying coreference chains for dyslexic children. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1142–1151, Marseille, France, May. European Language Resources Association.

Wilkens, R., Oberle, B., Landragin, F., and Todirascu, A. (2020a). French coreference for spoken and written language. In *Language Resources and Evaluation Conference (LREC 2020)*, Proceedings of the 12th Edition of the Language Resources and Evaluation Conference (LREC 2020), pages 80–89, Marseille, France.

Wilkens, R., Oberle, B., and Todirascu, A. (2020b). Coreference-based text simplification. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*, pages 93–100.

Woodsend, K. and Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of EMNLP 2011*, pages 409–420. Association for Computational Linguistics.

Yaseen, T. B., Ismail, Q., Al-Omari, S., Al-Sobh, E., and Abdullah, M. (2021). Just-blue at semeval-2021 task 1: Predicting lexical complexity using bert and roberta pre-trained language models. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 661–666.

Yimam, S. M., Biemann, C., Malmasi, S., Paetzold, G., Specia, L., Tack, A., and Zampieri, M. (2018). A report on the complex word identification shared task 2018. In *Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications (NAACL 2018 Workshops)*, pages 66–78. Association for Computational Linguistics; Stroudsburg,.

Zhang, X. and Lapata, M. (2017). Sentence simplification with deep reinforcement learning. In *Proceedings of EMNLP 2017*, pages 584–594.

Zhu, Z., Bernhard, D., and Gurevych, I. (2010). A mono-lingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd international con-ference on computational linguistics*, page 1353–1361. Association for Computational Linguistics.