Evaluating Lexical Simplification and Vocabulary Knowledge for Learners of French: Possibilities of Using the FLELex Resource

Anaïs Tack[1], Thomas François[1,3], Anne-Laure Ligozat[2], Cédrick Fairon[1]
[1]Centre for Natural Language Processing (CENTAL), Institute for Language & Communication (IL&C), Place Blaise Pascal 1, University of Louvain, B-1348 Louvain-la-Neuve, Belgium
[2]LIMSI, CNRS and ENSIIE, rue John Von Neumann, 91400 Orsay, France
[3]FNRS Postdoctoral Researcher
anais.tack@uclouvain.be, thomas.francois@uclouvain.be, annlor@limsi.fr, cedrick.fairon@uclouvain.be

Abstract

This study examines two possibilities of using the FLELex graded lexicon for the automated assessment of text complexity in French as a foreign language learning. From the lexical frequency distributions described in FLELex, we derive a single level of difficulty for each word in a parallel corpus of original and simplified texts. We then use this data to automatically address the lexical complexity of texts in two ways. On the one hand, we evaluate the degree of lexical simplification in manually simplified texts with respect to their original version. Our results show a significant simplification effect, both in the case of French narratives simplified for non-native readers and in the case of simplified Wikipedia texts. On the other hand, we define a predictive model which identifies the number of words in a text that are expected to be known at a particular learning level. We assess the accuracy with which these predictions are able to capture actual word knowledge as reported by Dutch-speaking learners of French. Our study shows that although the predictions seem relatively accurate in general (87.4% to 92.3%), they do not yet seem to cover the learners' lack of knowledge very well.

Keywords: FLELex, lexical simplification, vocabulary knowledge prediction

1. Introduction

In the area of second language acquisition, it is argued that mastering vocabulary is essential for skilfully practising a second or foreign language (L2). For L2 reading in particular, a number of studies have stressed the importance of knowing a sufficient amount of words in a text – at least 95% of a text's vocabulary to be precise – in order to achieve adequate reading comprehension (Laufer and Ravenhorst-Kalovski, 2010). It is therefore important to present a learner with suitable reading material, which includes vocabulary that is tailored to the learner's level of proficiency. In this view, recent advances in natural language processing (NLP) have led to the development of tools that automatically adapt the textual content to the learner (Burstein, 2009).

One way such content adaptation could be achieved is via automatic text simplification (ATS), as it alters the text to adjust its difficulty to the learner's level. ATS is commonly carried out through syntactic transformations (e.g. sentence splitting, sentence deletion, etc.) and lexical substitutions. Both can be induced from a parallel corpus of original and simplified texts (Woodsend and Lapata, 2011). However, the availability of such corpora is an issue, especially for French and even more for French L2. Indeed, the commonly used corpus constitutes an aligned version of Wikipedia and Simple Wikipedia (Zhu et al., 2010) and is thus mainly intended for native (L1) English readers. As a result, the lack of available corpora calls for the use of word frequency lists in the lexical substitution task (Devlin and Tait, 1998; Shardlow, 2014).

However, as the commonly used frequency lists are mainly L1-focused, we believe that they do not fit the L2 context very well. We therefore claim that a graded lexical resource intended for L2 learners is a better way to detect a learner's lexical difficulties and to identify the vocabulary to be automatically simplified. Such a resource has already been developed for French as a foreign language, viz. the FLELex graded lexical resource (François et al., 2014). Nevertheless, graded lexical resources have neither yet been examined as regards their usefulness for lexical simplification, nor have they been used to predict the difficult (or unknown) words for actual non-native speakers. These are therefore the goals of our study.

1.1. Previous work

A number of NLP studies have shown interest in the development of tools for dealing with lexical complexity in texts. One approach involves automatic lexical simplification, which aims at substituting difficult words with easier synonyms while preserving their original meaning. Most studies on word substitution have relied upon the combined use of a synonym database and a strategy to rank substitution candidates by difficulty. These substitution candidates are often ranked either according to their frequency of use – which might be obtained from word frequency lists (Devlin and Tait, 1998) or estimated from simple texts (Ligozat et al., 2012) – or according to a combination of word length and frequency measures (Bott et al., 2012), or even according to a combination of word difficulty features within a classifier (Shardlow, 2013; Gala et al., 2014; Jauhar and Specia, 2012). However, these strategies have not proven very successful so far. Shardlow (2014) stressed several shortcomings of a frequency-based approach, whereas the feature-based approaches were found struggling to outperform a frequency-based baseline, as has been shown in several studies (Shardlow, 2013; Specia et al., 2012; Gala et al., 2014).

Word frequency lists present other shortcomings in the context of L2 learning. They approximate the use of native speakers, but do not provide any information about the frequency of words within the different stages of the L2 curriculum. Thus, when substituting a word, it would be unfruitful to select a more frequent synonym based on L1 data, which has, nevertheless, not yet been introduced in a given L2 learner's curriculum.

For this reason, graded lexical resources offer several advantages over frequency lists such as those used by Devlin and Tait (1998). Not only do they contain information on the complexity of words relative to each other, but also information on what words are taught or learnt at a specific proficiency level. For French, two graded lexicons have been developed for L1 and for L2 respectively: the Manulex (Lété et al., 2004) and the FLELex (François et al., 2014) resources. As we are interested in the evaluation of lexical complexity for L2 speakers, our work will focus on the latter resource. FLELex includes lexical entries for single and multiword expressions linked with their frequency distribution across the six levels of the Common European Framework of Reference (CEFR) (Council of Europe, 2001). These frequency distributions are derived from textbooks used for teaching French L2.

1.2. Objectives

This study examines how to use a specialised lexicon such as FLELex in order to automatically identify those words in a text that are be difficult for and unknown to a non-native speaker and which might subsequently serve as potential candidates for lexical substitution. In particular, we investigate two possibilities of using the FLELex resource. On the one hand, we use FLELex to identify easy and complex words in a text through the annotation of a word's level of difficulty. By doing so, we aim to analyse whether FLELex can be used to detect a significant lexical simplification effect in manually simplified texts with respect to their original version. On the other hand, we use FLELex to predict the words in a text that are unknown to a learner of a given CEFR proficiency level. Our aim is to determine whether FLELex can be used to accurately predict known and unknown words in a text, by comparing these predictions to gold-standard learner annotations of vocabulary knowledge.

In the next section, we present the methodology adopted for collecting and for annotating the textual data used in our two experiments (Section 2.). We first present how we defined a corpus of original and simplified texts (Section 2.1.). We then describe how we automatically annotated this corpus in reference to FLELex and how we obtained a number of gold-standard learner annotations (Section 2.2.). The two subsequent sections then address the results of our analyses. We first assess the extent to which simplified texts contain easier words (i.e. having a lower difficulty level on the CEFR scale) compared to their original version (Section 3.). We then evaluate the relevance of using FLELex to predict L2 vocabulary knowledge (Section 4.). Finally, we present some concluding remarks and some future perspectives to enhance the automatic identification of lexical complexity with FLELex (Section 5.).

2. Method

2.1. Corpus definition

To investigate the issue of lexical complexity in French texts, we combined two aligned corpora of authentic and simplified texts previously collected by Brouwers et al. (2014), each of which pertained to a specific text genre, viz. the *Tales* corpus and the *Wiki* corpus. The *Tales* corpus constitutes a parallel corpus of 32 French narratives simplified for non-native speakers of French, whereas the *Wiki* corpus is a comparable corpus of 13,638 informative texts sourced from the French Wikipedia and from its Vikidia counterpart (i.e. a simplified Wikipedia for young native speakers of French). However, we should note that we only made use of a subcollection of the original *Wiki* corpus, containing all 388 wiki pages starting with the letter A. In **Table 1**, we report the number of texts analysed per corpus.

Study	*Tales*	*Wiki*	Total
Lexical simplification	32	388	410
Vocabulary knowledge	6	45	51

Table 1: The number of texts used per study and per corpus.

While we used the entire corpus in our first study on lexical simplification, we only used a part of it in the second one, where we compare FLELex's predictions of vocabulary knowledge to non-native speaker data. Indeed, as the combined corpus counted over 400 texts, we considered it too extensive to be read and annotated entirely by non-native speakers. We therefore reduced the initial corpus to a smaller sample of texts, while ensuring that the texts' vocabulary remained as much varied as possible. To this end, we used a greedy selection algorithm which retrieved from the combined corpus a subset of texts that had the best possible lexical diversity. Given a limitation to the number of lexical units allowed in the final subset, the algorithm searches in the space of all possible subsets to identify a subset of texts that portrays the least lexical overlap. Following the similarity measures proposed by Weeds et al. (2004), we compute the overlap between the vocabulary V of each pair of texts i and j using the Jaccard coefficient **(Equation 1)**. Each vocabulary V is defined as a set of unique *(lemma, POS-tag)* combinations.

$$overlap(V_i, V_j) = \frac{|V_i \cap V_j|}{|V_i \cup V_j|} \quad (1)$$

The algorithm uses each document d in the corpus as a starting point to build a new subset T and then iteratively integrates into T a new document t that shares the least lexical units with the documents already included. After having constructed a complete subset T, the algorithm compares this subset to the previously constructed subsets in order to select the one with the least overall lexical overlap. In the event that there are two documents that present the same degree of overlap, the algorithm chooses the one that portrays the richest vocabulary using the Standardised Type/Token Ratio **(Equation 2)** (Scott, 1996).

$$STTR = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{types_i}{tokens_i}\right) \times 100 \qquad (2)$$

After the selection procedure, we obtained an optimal subset of 51 texts, counting 21,047 lexical units in total.

2.2. Corpus annotation

After having defined the corpus of texts to be used in each part of our study, we proceeded to the annotation of the texts' lexical units. To this aim, we automatically lemmatised and part-of-speech tagged each text using TreeTagger (Schmid, 1994). We also identified multiword expressions (MWE) using the version of the FLELex resource including MWEs identified by means of a Conditional Random Field tagger. We then generated two sets of lexical annotations. On the one hand, we automatically annotated each lexical unit in the global corpus with its level of difficulty based on FLELex (see Section 2.2.1.). On the other hand, we asked a group of Dutch-speaking learners of French to manually annotate the 51-text sample according to their vocabulary knowledge (see Section 2.2.2.).

2.2.1. FLELex annotations

We created a system that automatically annotates the level of complexity of each lexical unit in our corpus using the lexical frequency distributions observed in the FLELex resource **(Table 2)**. Following Gala et al. (2014), we defined the level of complexity of a of a lexical unit as the CEFR level where it occurs for the first time. In this way, the adjective *épais* ("thick") received the lowest difficulty level (i.e. A1), whereas the noun *épanchement* ("outpouring") received the highest difficulty level (i.e. C2). All lexical units that were absent from the resource received a *NA* value. **Listing 1** gives an example of a FLELex-annotated text.

```
<p>
 <span id="4" lemma="le" pos="DET" level="
  A1">Le</span>
 <span id="5" lemma="petit" pos="NOM"
  level="A2">petit</span>
 <span id="6" lemma="se" pos="PRO" level="
  A1">s'</span>
 <span id="7" lemma="appeler" pos="VER"
  level="A1">appelle</span>
 <span id="8" lemma="le" pos="DET" level="
  A1">l'</span>
 <span id="9" lemma="aiglon" pos="NOM"
  level="mot_absent">aiglon</span>.
</p>
```

Listing 1: Example FLELex annotation (excerpt from the *Wiki* corpus, text *Aigle royal*, simplified).

We manually verified the accuracy of our annotation system on the *Tales* corpus and achieved a F_1 measure of .99, due to some lemmatisation and tagging errors made by TreeTagger. In addition, we made the system freely available as a tool on the FLELex website[1]. The tool enables a user to analyse the lexical complexity of a text for a specific complexity level (according to the CEFR scale) by highlighting the words that have a complexity level beyond a user-defined threshold **(Figure 1)**.

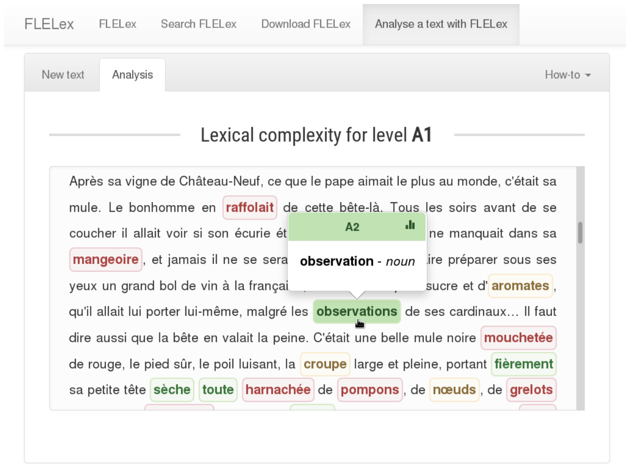

Figure 1: Example output of our lexical complexity annotation tool, highlighting the words that receive a difficulty level higher than the A1 level according to FLELex.

2.2.2. Learner annotations

We enriched our 51-text sample with learner annotations defining which words are (un)known to a learner having a particular CEFR proficiency level. In order to obtain such learner data, we conducted a reading experiment with four Dutch-speaking learners of French, amongst whom were two learners having attained the A2 proficiency level (one having finished the second form of secondary school and one the third form, hereinafter the learners A2-2 and A2-3) and two having attained the B1 proficiency level (one having finished the fourth form and one the first year of university, hereinafter the learners B1-4 and B1-U). Each learner was presented with the 51-text sample via a web interface where they had to identify the words they did not know the meaning of **(Figure 2)**. The interface selected the texts to be annotated in a random order and presented them one sentence at a time in order to reduce the effect of inferring the meaning of unknown words from context. **Listing 2** gives an example of the gold-standard learner annotations obtained via the reading experiment.

```
<p>
 <span id="4">Le</span>
 <span id="5">petit</span>
 <span id="6">s'</span>
 <span id="7">appelle</span>
 <span id="8">l'</span>
 <span id="9" class="unknown">aiglon</span>
 .
</p>
```

Listing 2: Example learner annotation (excerpt from the *Wiki* corpus, text *Aigle royal*, simplified, annotated by learner A2-3).

[1] http://cental.uclouvain.be/flelex/

Lemma	POS-tag	A1	A2	B1	B2	C1	C2	Total
épais	ADJ	0.8346	14.6716	22.0201	13.6059	6.0005	0.0	16.9325
épaisseur	NOM	0.0	0.0	3.5822	6.0629	6.8846	0.0	3.5051
épanchement	NOM	0.0	0.0	0.0	0.0	0.0	19.328	0.1431
épandre	VER	0.0	0.0	0.0	2.2839	6.0005	0.0	0.7085
épanouir	VER	0.0	0.0	0.6378	8.4014	6.0005	12.833	3.5281
épanouissement	NOM	0.0	0.0	0.0	2.2839	19.861	19.328	2.5855
épargne	NOM	0.0	0.0	0.6378	0.0	0.0	0.0	0.0592

Table 2: A fragment of the lexical frequency distributions per lexical item as reported in the FLELex resource.

Figure 2: The web interface used to collect the learner annotations. The words that are unknown to a given learner are highlighted in red.

3. Assessing lexical simplification

In the previous section, we discussed how we annotated the level of lexical difficulty in a corpus of original and simplified texts and how we defined the level of difficulty of a word as its first level of occurrence in the FLELex resource. Using this annotation of a text's lexical complexity, we then proceeded to the evaluation of lexical simplification in manually simplified texts. Our aim was to examine whether the use of FLELex could enable us to distinguish between different levels of textual complexity.

In order to detect a significant lexical simplification effect, we compared each pair of original and simplified texts with respect to the number of lexical units that were attributed to each of the six difficulty levels in FLELex (**Table 3**). However, as the texts included in both the *Tales* and the *Wiki* corpus differed greatly in terms of text length, we normalised the counts to a length of 1,000. Furthermore, due to sparsity issues at the higher difficulty levels, we combined the counts of the elementary levels (i.e. A1 and A2) into one global level and did the same for the intermediate (i.e. B1 and B2) and for the advanced (i.e. C1 and C2) levels.

Table 3 shows that the majority of the words in the original and simplified texts were easy and belonged to the A1 and A2 difficulty levels. This was not surprising as many of the words were grammatical words and most of them were already frequently observed from the A1 level onwards. Furthermore, we also observed that the simplified texts had a simpler vocabulary, with words belonging primarily to the elementary levels, whereas the original tales included more intermediate and advanced vocabulary items. We conducted a MANOVA on the normalised word counts for each of the three difficulty levels, comparing each aligned pair of texts. Our test revealed a significant association between the vocabulary counts per level and the text version (original vs. simplified), both for the *Tales* corpus ($F(3,28)$ = 26.337; p <.001) as for the *Wiki* corpus ($F(3,384)$ = 16.939; p <.001).

These results indicate that FLELex could effectively be used to evaluate the degree of lexical simplification in a text for non-native readers. Indeed, from the examples **(1)** and **(2)**, we can see that the verb *geindre* ("to moan"), which appears in FLELex from the advanced C1 level onwards, is substituted in the simplified version by its easier synonym *gémir* ("to moan"), which appears in FLELex from the elementary A2 level onwards. We can thus safely state that a graded lexical resource such as FLELex could be used to correctly distinguish between the complexity of two synonyms and hence to automatically substitute complex words with their easier synonyms observed in the resource. However, we should note that our results remain inconclusive as to the effectiveness of using FLELex in an automatic text simplification system.

(1) Le paysan cessa un instant de **geindre** pour répondre.
A1 A1 A2 A1 A1 A1 **C1** A1 A1
"The peasant stopped **moaning** for an instant to answer."
(*Tales* corpus, *La Bête à Maître Belhomme*, original)

(2) Le paysan cesse de **gémir** pour répondre.
A1 A1 A2 A1 **A2** A1 A1
"The peasant stopped **moaning** to answer."
(*Tales* corpus, *La Bête à Maître Belhomme*, simplified)

4. Predicting a learner's lexical knowledge

We also investigated the possibility of using FLELex's CEFR level frequencies to predict which words are known by a non-native speaker at a given proficiency level. We used FLELex as a predictive model of the learner's lexical knowledge, hereinafter referred to as the *expert model*. This expert model classified all lexical units annotated with a level higher than the learner's level as *unknown*. The words that did not appear in FLELex (i.e. annotated with a *NA* value) were assigned the highest difficulty level, since they were regarded as not having been taught during the learner's curriculum. We then compared the model's predictive accuracy with respect to the gold-standard learner annotations collected for the 51-text sample.

As we recall from the previous section, the majority of the lexical units in our FLELex-annotated corpus belonged to

Corpus	Version	A1-A2		B1-B2		C1-C2	
Tales	original	961.3	(σ = 13.0)	34.8	(σ = 12.3)	4.0	(σ = 1.5)
	simplified	989.6	(σ = 13.9)	10.0	(σ = 13.7)	0.4	(σ = 0.6)
Wiki	adults	927.7	(σ = 23.2)	57.9	(σ = 18.3)	14.4	(σ = 9.7)
	children	949.4	(σ = 36.9)	42.7	(σ = 32.4)	8.3	(σ = 14.1)

Table 3: The mean number of words in a text that belong to a specific difficulty level as observed in FLELex. All word counts have been normalised to a text length of 1,000 words.

the A1 level, which was none different in the 51-text sample **(Figure 3)**. As a consequence, the expert model predicted the majority of the words in the sample as known to the study participants, who had attained a proficiency level higher than the A1 level (i.e. A2/B1).

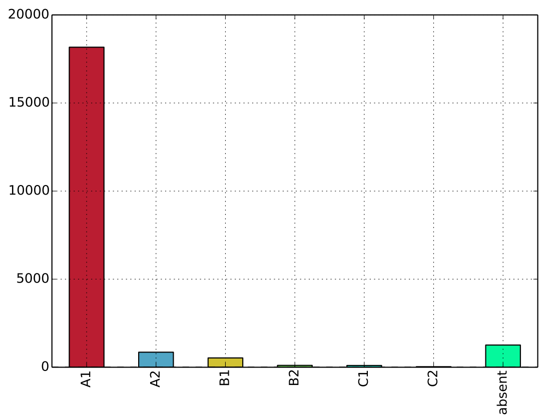

Figure 3: The number of lexical units in the 51-text sample that belong to a specific CEFR level as observed in FLELex.

As for the learner annotations, we observed that the study participants also reported knowing the majority of the words the texts **(Table 4)**, which corresponds to the findings reported in studies examining the percentage of words known to L2 readers (Laufer and Ravenhorst-Kalovski, 2010; Schmitt et al., 2011).

	Known	Unknown
learner A2-2	95.7%	4.3%
learner A2-3	88.1%	11.9%
learner B1-4	97.0%	3.0%
learner B1-U	96.7%	3.3%

Table 4: The percentage of words annotated as *known* and *unknown* according to the study participants.

When comparing the predictions of the expert model to the actual learner annotated data, we noticed that the predictions of our expert model were relatively accurate, ranging from 87.4% to 92.3% correctly classified lexical units as *known* and *unknown* **(Table 5)**. However, we also noticed that the accuracy of our model differed according to the word's part-of-speech category. In fact, it seemed that our predictions of vocabulary knowledge were almost totally accurate in the case of grammatical words, with an accuracy ranging from 99.2% to 99.8%. As for the lexical words, the model was less accurate, having correctly classified between 81.1% and 91.3% of them. This could be related to the fact that the inter-annotator agreement between the learners of the same proficiency level differed more greatly in the case of lexical words **(Table 6)**. Indeed, whereas the learners tended to share the same grammatical knowledge, they did not always seem to know the same lexical words. As a consequence, even though the expert model predicted the grammatical knowledge quite accurately, it did not yet take into account the individual differences that we observed in the learners' lexical knowledge within a given proficiency level.

Compared to	Lexical	Grammatical	Total
learner A2-2	86.6%	99.2%	89.7%
learner A2-3	81.1%	99.2%	87.4%
learner B1-4	91.3%	99.7%	92.3%
learner B1-U	90.8%	99.8%	92.0%

Table 5: The accuracy of the predictions made by the expert model in comparison to the participants' annotations.

Between learners	Lexical	Grammatical	Total
A2-2 and A2-3	81.1%	99.5%	89.3%
B1-4 and B1-U	94.6%	99.9%	97.1%

Table 6: The inter-annotator agreement of word knowledge between learners having the same CEFR proficiency level.

Looking more closely at the model's per-class accuracy, we observed that it captured the known words very accurately, but not the unknown words **(Table 7)**. Indeed, the model did not recall half of the words that were actually unknown to the participants. This could be attributed to the fact that the rule to transform the FLELex distributions into a single level (i.e. using the first level of occurrence) might have fallen short. As a large number of words already occur in the resource from the A1 level onwards, defining a word's difficulty level as the first level of occurrence might have led us to consider some words too readily as known. In order to solve this classification problem, we need to explore other possible rules for transforming the FLELex distributions into a single difficulty level. In particular, we propose two possible solutions. On the one hand, we could explore the across-textbook frequency distributions in more detail in order to evaluate whether a word really belongs to the core vocabulary taught at a given level. The core vocabulary per level includes those words that are observed in most textbooks. In contrast, the peripheral vocabulary per level includes those words that tend to appear in only few textbooks and which are hence not indicative of the vocabulary commonly targeted at a given proficiency level. On the other hand, we could collect more extensive learner annotations for all six proficiency levels in order to automatically learn a new transformation rule that defines vocabulary knowledge similarly to the learners' annotations.

	Known		Unknown	
	precision	*recall*	*precision*	*recall*
learner A2-2	0.97	0.92	0.19	0.42
learner A2-3	0.92	0.94	0.47	0.38
learner B1-4	0.98	0.94	0.17	0.40
learner B1-U	0.98	0.94	0.17	0.37

Table 7: The expert model's precision and recall of known and unknown words.

5. Conclusion and future perspectives

This study investigated two possibilities of using the FLELex resource: to analyse lexical simplification in parallel corpora and to predict the vocabulary knowledge of learners at a given proficiency level. On the one hand, we observed that defining the word's CEFR level on the basis of its first occurrence in FLELex enabled us to detect a significant simplification effect in manually simplified texts. On the other hand, we observed that an expert model predicting the learner's actual vocabulary knowledge based on the word's CEFR level was relatively accurate, but that it presented some errors with respect to the recall of unknown words. As a consequence, we can safely state that a graded lexical resource such as FLELex offers a promising solution to the automated assessment of a text's lexical complexity for non-native readers.

However, in order for FLELex to be successfully integrated in a system that analyses whether and how a text's vocabulary should be simplified for a non-native reader, we need to continue improving our method. In particular, we need to establish a better rule to transform the word's FLELex distributions into a single CEFR level, distinguishing a level's core vocabulary from its peripheral vocabulary. By doing so, we will be able to better pinpoint the words in a text that are unknown to a learner in order to correctly identify the ones that should be simplified for enhancing the text's readability.

6. Acknowledgements

We would like to thank the study participants for the time they spent giving us an essential feedback on their lexical competence, which enabled us to gain some valuable insights. We also thank the French as a foreign language teacher who helped us get in contact with her students.

7. Bibliographical References

Bott, S., Rello, L., Drndarevic, B., and Saggion, H. (2012). Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *Proceedings of Coling 2012: The 24th International Conference on Computational Linguistics*, pages 357–374.

Brouwers, L., Bernhard, D., Ligozat, A.-L., and François, T. (2014). Syntactic Sentence Simplification for French. In *Proceedings of the 3rd International Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 47–56. Association for Computational Linguistics.

Burstein, J. (2009). Opportunities for Natural Language Processing Research in Education. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, number 5449 in Lecture Notes in Computer Science, pages 6–27. Springer Berlin Heidelberg.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press, Cambridge, UK.

Devlin, S. and Tait, J. (1998). The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic databases*, pages 161–173.

François, T., Gala, N., Watrin, P., and Fairon, C. (2014). FLELex: a graded lexical resource for French foreign learners. In *Proceedings of the International conference on Language Resources and Evaluation*, LREC '14, pages 3766–3773.

Gala, N., François, T., Bernhard, D., and Fairon, C. (2014). Un modèle pour prédire la complexité lexicale et graduer les mots. In *Actes de la 21e conférence sur le Traitement Automatique des Langues Naturelles*, TALN '14, pages 91–102.

Jauhar, S. K. and Specia, L. (2012). UOW-SHEF: SimpLex–lexical simplicity ranking based on contextual and psycholinguistic features. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 477–481.

Laufer, B. and Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, 22(1):15–30.

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.

Ligozat, A.-L., Garcia-Fernandez, A., Grouin, C., and Bernhard, D. (2012). ANNLOR: a naïve notation-system for lexical outputs ranking. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 487–492.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*, volume 12, pages 44–49.

Schmitt, N., Jiang, X., and Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1):26–43.

Scott, M. (1996). *Wordsmith tools*. Oxford: Oxford University Press.

Shardlow, M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. In *ACL Student Research Workshop*, pages 103–109.

Shardlow, M. (2014). Out in the Open: Finding and Categorising Errors in the Lexical Simplification Pipeline. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, LREC '14, pages 1583–1590. European Language Resources Association (ELRA).

Specia, L., Jauhar, S. K., and Mihalcea, R. (2012). Semeval-2012 task 1: English lexical simplification. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, SemEval '12, pages 347–355.

Weeds, J., Weir, D., and McCarthy, D. (2004). Characterising Measures of Lexical Distributional Similarity. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Woodsend, K. and Lapata, M. (2011). Learning to Simplify Sentences with Quasi-synchronous Grammar and Integer Programming. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 409–420. Association for Computational Linguistics.

Zhu, Z., Bernhard, D., and Gurevych, I. (2010). A Monolingual Tree-based Translation Model for Sentence Simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 1353–1361. Association for Computational Linguistics.