Exploration of Spanish Word Embeddings for Lexical Simplification

Rodrigo Alarcón[1], Lourdes Moreno[1] and Paloma Martínez[1]

[1]Computer Science and Engineering Department, Universidad Carlos III de Madrid, 28911 Madrid, Spain

Abstract

Lexical simplification systems rely heavily on handcrafted databases or parallel corpora which represents a high cost of production. In this paper we present alternatives for every step in the lexical simplification process individually for the Spanish language by exploring the potential that word embeddings can offer. This study covers the entire pipeline in lexical simplification, from the task of complex word identification (CWI) to substitute generation, selection and ranking (SG/SS/SR). Taking advantage of the different applications of BERT models, we fine-tune two pre-trained models to detect unusual words with the help of available Spanish datasets. Next, we compare features that different types of embedding can give to find the best candidate for replacement for a target word. The resulting models in the CWI step show a fair result compared to other systems that used the same datasets. Also, we found better results than previous works by analyzing the similarity of words in context when evaluating embedding models.

Keywords

lexical simplification, word embedding, BERT, Spanish

1. Introduction

Lexical simplification (LS) is an important subfield of text simplification that gives attention to the complexity of words, and particularly how to measure readability and reduce the complexity using alternative replacements. Most current approaches to LS heavily rely on corpus statistics and surface-level features, such as word length and corpus-based word frequencies [1]. The most popular LS systems still predominantly use a set of rules for substituting complex words with their frequent synonyms from carefully handcrafted databases or automatically induced from parallel corpora. However, language resources are scarce or expensive to produce, such as WordNet and Simple Wikipedia. Also, the scarcity of these resources may be greater according to the language, as it is the case of Spanish versus English.

In order to find alternative solutions to these costly procedures, recent works use word embeddings to extract important information from a text with less effort [2]. This information can help in various aspects of the simplification process, such as extracting word vectors to detect unusual words [3] or determining the similarity of words in a given context.

Proceedings of the First Workshop on Current Trends in Text Simplification (CTTS 2021), co-located with SEPLN 2021. September 21st, 2021 (Online). Saggion, H., Štajner, S. and Ferrés, D. (Eds).
ralarcon@inf.uc3m.es (R. Alarcón); lmoreno@inf.uc3m.es (L. Moreno); pmf@inf.uc3m.es (P. Martínez)
0000-0003-3296-6873 (R. Alarcón); 0000-0002-9021-2546 (L. Moreno); 0000-0003-3013-3771 (P. Martínez)

In this paper, we explore applications that word embedding can offer in each of the stages of the LS process individually. We present a simple way of fine-tuning Spanish and multilingual BERT models for the detection of unusual or complex words. As a next step, to provide replacements that match the context of the original word, we use the similarity information between the word vectors of different types of embeddings. Finally, we combine this information to rank words in terms of simplicity.

The remainder of this paper is organized as follows. In Section 2, we review the work related to the text simplification process. In Section 3, we describe the datasets to train and evaluate the procedures proposed in each step of LS. Sections 4, 5, 6 and 7 provide the procedures and evaluation to the complex word identification (CWI), substitute generation/selection/ranking (SG/SS/SR) modules. Finally, Section 8 offers conclusions and future work.

2. Related Work

Natural language processing is a discipline dedicated to developing technology capable of understanding natural language in a way similar to human beings. One area in which this could be applied is the development of technology that improves accessibility for individuals with disabilities. LS aims at replacing complex words with simpler alternatives which can help various groups of people, including people with autism [4] [5] [6], aphasia [7] [8], low vision [9], dyslexia [10] [11] or people with intellectual disabilities [3] [12] [13] [14]. According to studies [2], LS is an essential task because a person needs to know 95% of the vocabulary to understand a text at a basic level. Therefore, this suggests that replacing words that are unusual for a person can improve the accessibility of a given text.

There are three approaches to LS, from supervised machine learning algorithms to unsupervised algorithms or even hybrid approaches which combine the advantages of both. Paetzold [2] proposed four stages to achieve LS, which are: CWI, SG, SS and SR. This paper follows these stages and explores whether embeddings can help at each stage.

CWI aims to select the complex words candidates to be simplified in a given text. Several ways to accomplish this task have been proposed, however, approaches based on machine learning have proven to be the most suitable. Shardlow [15] compared binary support vector machine (SVM), threshold-based, and "Simplify Everything" approaches, where in the latter, it is assumed that all words in a sentence can be simplified. The results demonstrated that the SVM approach outperforms the others in terms of precision. This is confirmed in competitions focused on this task, such as the BEA workshop (2018) [16], where machine learning-based approaches proved to have the best F1 scores [17]. In this research work, we exploit the versatility of BERT and perform fine-tuning of a model with the help of information from two Spanish datasets (see Section 3) with the aim of performing Named Entity Recognition (NER) to detect complex and simple words in a given text described at Section 4.

Moving on to the next step, SG involves producing substitute candidates for the complex words detected. Two approaches have been propose, which are linguistic database querying and automatic generation [2]. The first one obtains candidates from databases manually constructed by professionals, thus providing reliable data on words associated with their candidates [18] [13]. Nevertheless, it has the disadvantages of being a time-consuming task and not having a wide coverage, especially in Spanish. Automatic generation focuses on overcoming this disadvantage and seeks to gather extracted candidates from less expensive resources. For example, the simple multilingual Paraphrase Database (PPDB) [19] where they initially annotated a paraphrase dataset to train a model to classify simplified paraphrases resulting in a database of more than a billion of paraphrases for different languages. Due to the limited amount of resources available for the Spanish language, in this work we explore another source that has the ability to provide replacements for a target word, such as the case of word embeddings. In Section 5, we compare different models to evaluate which type of embedding performs better at this step.

In the third step (SS), in which a substitute is selected from the set of synonyms extracted from the previous step, the most suitable synonym is chosen according to its simplicity and context. In recent years, several strategies have been proposed, for example, works that took this step as a task of Word Sense Disambiguation (WSD) [20]. Moreover, in languages where WSD resources are sparse or unavailable, Part-of-speech (POS) strategies were proposed, as in [21], where the words are filtered using a set of rules, including among others, the POS tag of the candidate. Unfortunately, this approach showed poor results when dealing with highly ambiguous words. Therefore, to address these problems, recent works incorporate similarity metrics in the selectors where authors picked out the final synonym using the cosine distance in a word embedding model. Given a word to be simplified, the word with the closest vector based on cosine similarity was chosen [22]. In Section 6 we perform a similar procedure and explore which type of embedding has better results when selecting which words are more appropriate for a given context.

Finally, the SR step consists in deciding which of the candidate substitutions that fit the context of a complex word is the simplest. One of the most commonly used and simplest strategies for dealing with this task is frequency-based procedures. These suggest that the more a word is used, the more familiar it is to a user; these word frequencies can be extracted from very large corpora [23] and can be quite effective in front of other approaches. As in the other steps, machine learning assisted approaches have been adopted lately. For example, a support vector machine accompanied with additional metrics to sort words according to their simplicity [24]. Later, more sophisticated works such as neural approaches were presented [1], such is the case of [25], where a supervised neural ranking model is presented. This ranker receives a set of features (n-gram probabilities) for a pair of candidates as input, and produces as output the simplicity difference between them. Recently, some works have combined resources obtained from the strategies described above. Such is the case of [26], which uses a weighting system where it takes features extracted from word embeddings, language models and word frequencies. In this work, we follow this idea by incorporating our own features adapted to Spanish, extracted from word embeddings and a frequency dictionary (described at Section 7).

3. Datasets

For experimentation, different datasets for training and testing are used. These datasets are described below.

3.1. BEA DATASET

The dataset is composed of annotated Spanish Wikipedia pages proposed in the BEA Workshop 2018 for the CWI [1] task. As shown in Table 3, a total of 17603 instances were annotated by 54 Spanish speakers, most of whom were native. Each instance contains a uniword/multiword target which is selected by annotators. A target is marked as complex if at least one annotator designates it as complex.

3.2. EASIER DATASETS

These datasets [27] are part of the EASIER corpus [2] [3], which were developed by the authors of this work to offer evaluation support for CWI tasks and fitting SG/SS tasks contextually. A linguist expert in easy-to-read and plain language guidelines annotated 260 news documents. Later on, two additional experts and a target audience analysed the resulting corpus to assure the quality of the data provided.

Table 1 and 2, show examples of instances found in the CWI and substitute datasets, respectively. In the dataset for CWI, information such as a sentence, a target word, the word offset and the gold-standard label for the binary task can be found. While, for the substitutes dataset, a sentence, a target word and proposed substitutes can be found.

Sentence	Start offset	End offset	Word	Label
La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	3	14	Importancia (importance)	0
La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	18	22	Leer (reading)	0
La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	31	41	Etiquetado (labelling)	1
La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	51	58	Comprar (purchasing)	0
La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	62	70	Alimento (foodstuffs)	0

Table 1
CWI dataset sample of EASIER corpus

Word	Sentence	Proposed Substitutions
Etiquetado (labelling)	La importancia de leer bien el etiquetado antes de comprar un alimento. (The importance of carefully reading the labelling before purchasing foodstuffs.)	Letrero (sign), inscripción (inscription), rótulo (banner)

Table 2
Substitutes dataset sample of EASIER corpus

Also to complement the above information, Table 3 shows additional information on the size of the resources. The BEA dataset contains more than 17,000 instances where more than 7,000 complex words are found. Whereas, the CWI dataset of the EASIER corpus contains more than

[1] google.com/view/cwisharedtask2018
[2] https://data.mendeley.com/datasets/ywhmbnzvmx/2
[3] github.com/LURMORENO/EASIER_CORPUS

44,000 instances where more than 8,000 complex words are found. Similarly, the EASIER corpus substitutes dataset contains 5,130 instances resulting in more than 7,000 proposed substitutes.

Table 3
CWI-SG/SS datasets distribution

	Instances	Complex	Proposed synonyms
BEA	17,603	7,015	-
EASIER-CWI	44,975	8,155	-
EASIER-SG/SS	5,130	-	7,892

4. Complex Word Identification

In this stage, we need to distinguish which words are complex and which are not for a certain audience.

We propose BERT [28] for NER because is a powerful NLP model but using it for NER without fine-tuning it on NER dataset won't give good results. In this work, we fine-tune two models with the help of the datasets described at Section 3 to perform the CWI task: a Google´s multilingual BERT pre-trained model (mBERT) [4] [28] and an Spanish BERT pre-trained model (BETO)[5] [29]. We took the original idea from an implementation [6] for CoNLL-2003 [30] and then modified it so that the model can predict the entities "COMPLEX" and "SIMPLE" in a given text. The default parameters for the fine-tuning are the following:

- train_batch_size: 32
- max_seq_length: 128
- learning_rate: 2e-5
- num_train_epochs: 4.0
- do_lower_case: False
- Crf: True

Table 4 shows the results for CWI task and also shows the results of the models trained with the combination of the data of the Spanish datasets. Also, we compare the results with a traditional machine learning approach, presented in [3], where a SVM is trained with the BEA dataset. This linear SVM was trained with word length, morphological, easy-to-read content and embedding model features (Word2vec, BERT). Concerning the typical metrics for this task, we use precision, recall and the F1-score.

The traditional machine learning approach still shows better results on these datasets by achieving an F-1 score of 0.792 versus 0.694 for the mBERT model trained and tested with EASIER data. Also by combining datasets both of the models obtained an F-1 score of 0.685 on the EASIER test dataset.

[4] https://github.com/google-research/bert/blob/master/multilingual.md
[5] https://github.com/dccuchile/beto
[6] github.com/kyzhouhzau/BERT-NER

Table 4
Results for BERT models on CWI task where the structure is Model_TrainDataset_TestDataset

	Precision	Recall	F1
mBERT_EASIER_EASIER-test	0.695	0.694	0.694
BETO_EASIER_EASIER-test	0.696	0.691	0.693
mBERT_BEA_BEA-test	0.669	0.628	0.643
BETO_BEA_BEA-test	0.653	0.640	0.640
mBERT_EASIER-BEA_BEA-test	0.676	0.675	0.674
BETO_EASIER-BEA_BEA-test	0.639	0.603	0.598
mBERT_EASIER-BEA_EASIER-test	0.685	0.687	0.685
BETO_EASIER-BEA_EASIER-test	0.695	0.677	0.685
SVM approach_BEA_BEA-test	**0.80**	**0.79**	**0.792**

Also, by analyzing errors, we noticed that many of the false positives were instances where a multiword was the target, however, the data for training/testing used in fine-tuning, only correspond to instances that have uniwords, while SVM classifies instances where the target is uniwords and multiwords. We believe that by incorporating these instances in the classification of the BERT model, the score can be improved and compared to the BEA score.

5. Substitute Generation

The SG stage generates substitution candidates for complex words, considering all the contexts in which they may appear. We tested the performance of different embedding models by extracting and evaluating the nearest neighbors of each target word (top-50 neighbors). The tested models are the following:

- **Word2vec** model: pre-trained on The Spanish Billion Words Corpus [7].
- **Sense2Vec** model: Since there are no Spanish Sense2Vec models [31]. We created a model trained on The Spanish Billion Words Corpus [32]. A sense is a word combined with a label that represents the context of a word (in this case we use the POS tag as a label). The main diference of Sense2Vec and Word2Vec vectors is that the latter fail to encode the context by assigning a single key regardless of the context in which it appears. This does not happen in a Sense2vec model, because it generates vectors of words with contextual keys (i.e., one vector for each sense of the word).
- **FastText** model: pre-trained on Wikipedia with the FastText tool with character n-grams of length 5 [8] [33].
- **BERT** model: Pytorch BETO model described at Section 4.

The gold set is part of the EASIER corpus, which is represented by 575 instances in which for each instance a target word has three proposed substitutes. For this evaluation the first 500 instances are taken for the test set. In addition, we compare the results of the models

[7] https://crscardellino.ar/SBWCE/
[8] https://fasttext.cc/docs/en/crawl-vectors.html

with a previous approach which performs a linguistic database strategy proposed in [3] where we developed the same task by extracting replacements for a target word from Babelnet [34], Thesaurus [9] and PPDB [19].

The evaluation metrics used are those found in the work of Paetzold [2], which are as follows:

- **Potential:** The proportion of instances for which at least one of the candidates generated is contained within the gold standard.
- **Precision:** The proportion of generated substitutions that are contained within the gold standard.
- **Recall:** The proportion of gold-standard substitutions that are among the generated substitutions.
- **F-1:** The harmonic average between precision and recall.

Table 5 includes the results obtained for this step. At this stage, potential and recall are important measures because, according to its definition, it is required to obtain the widest coverage in the contexts in which a word may appear. The approach developed in [3] showed a higher performance than the result of the embedding models, obtaining a potential and recall of 0.898 and 0.597 respectively, versus the second best being the Sense2Vec model with a potential of 0.506 and a recall of 0.298. When analyzing the negative results we found cases in which the output was a repeated candidate but in different grammatical forms. In turn, because these models provide semantic similarity of words, in many cases, apart from synonyms, antonyms were found in the lists.

Table 5
Results for SG

	Potential	Precision	Recall	F-1
Word2vec	0.358	0.0191	0.188	0.034
FastText	0.464	0.0294	0.289	0.053
Sense2Vec	0.506	**0.056**	0.298	**0.095**
BERT	0.348	0.030	0.282	0.054
Easier (Previous approach)	**0.898**	0.043	**0.597**	0.080

6. Substitute Selection

The SS stage takes the list of synonyms extracted from the previous step and selects the most suitable synonym according to its simplicity and context. As the core resource in this step, we use different types of word embedding models, from static to contextualized. We use the same embedding models as in Section 5. These models allow us to calculate the cosine distance between word vectors to perform the following procedures:

- **No selections :** selects all candidates.

[9] thesaurus.altervista.org

- **Lexical window :** obtains three similarity values (candidate and target word, candidate and target word's context words in the sentence (previous and subsequent words)). Next, these values are added and stored. Finally, this process is repeated for every candidate, and the selector picks the three candidates with the highest values.

For the evaluation of this stage, we use the same data set and metrics as in Section 5. On the selector to evaluate, each selector needed candidates to rank, so we use the generator with the best potential ranking from the previous step described in Section 5 and then randomly insert the correct substitutes for each of the instances from the gold set. Furthermore, in this evaluation each selector had to propose the top 3 substitutes per instance. We believe that in this way we can easily determine the effectiveness of the selector based on which selector yields the highest number of potential correct answers.

Table 6 illustrates the results. Unlike the previous step, in the substitute selection a higher precision is pursued. As expected, performing no selection results in high potential and recall, however, the precision is very low. With the closest score in potential is the FastText model, which showed a precision of 0.364 being higher than the Word2Vec model used in previous works [3] with a precision score of 0.315. We assume that this higher score was obtained because the FastText model provides char and ngrams embeddings to face the problem of OOV (Out-of-vocabulary) words.

On the other hand, about the results of the BERT model, it is worth noting that word-level similarity comparisons are not appropriate with BERT embeddings because these embeddings are contextually dependent, meaning that the word vector changes depending on the sentence it appears in. A better intuition for this stage with the BERT model would be to evaluate the similarity between the sentences in which a candidate is found.

Table 6
Results for SS

	Potential	Precision	Recall	F-1
No Selection	**1.0**	0.098	**1.0**	0.178
Word2vec	0.618	0.315	0.315	0.315
FastText	0.674	**0.364**	0.364	**0.364**
Sense2Vec	0.62	0.312	0.312	0.312
BERT	0.192	0.108	0.108	0.108

7. Substitute Ranking

The SR stage takes the list of synonyms extracted from the previous step and chooses which candidate that fits the context is the simplest, taking into account the target user. At this stage, a combination of frequency-based and machine learning-assisted strategies has been implemented by developing a weighting module that uses the different features to rank a word.

Table 7 shows the results of different combinations of these features. To our knowledge there are no datasets in Spanish to evaluate this procedure, therefore, the decision to adapt these procedures to evaluate it with English language datasets was made, specifically, datasets from the English Lexical Simplification task of SemEval 2012 [35]. The trial set is composed of 300 instances, and the test set, 1, 710 instances. Each instance contains a sentence, a target complex word, and candidates ranked by their simplicity.

The evaluation metric is the TRank measure, proposed in the shared task. This metric calculates the proportion of instances for which the highest ranked candidate produced by a ranker is the same as the one in the gold-standard. In addition, the Table 7 also shows results of the best ranker presented in the shared task. The ranker must make the decision to choose the simplest candidate based on the candidates that obtained the best results in each of the following features:

- **BERT prediction:** Probability distribution of the candidate. This can be obtained from the vocabulary corresponding to the mask word. The higher the probability, the more relevant the candidate for the original sentence. The BETO model described above is used for Spanish and a multilingual DistilBERT [36] model is used for English.
- **Semantic similarity:** Cosine distance between the original word vectors and the candidate vectors in the list. The shorter the distance, the more similar the two words. To extract these vectors we test different embedding models. For the Spanish language, the classic embedding models described above are used and for English, pre-trained models for that language are used [10] [11] [12].
- **Frequency Feature:** Because frequency-based approaches have shown good results at this stage, the decision was made to incorporate it as a feature in the ranker. The more frequent a word is, we assume that a word is simpler. For Spanish, we used a dictionary of the Real Academia de la Lengua Española (RAE)[13] to extract the frequency of each candidate, which is made up of 10000 terms ordered by their frequency. As for the English language, a portion of about 5000 instances of the Corpus of Contemporary American English (COCA)[14] has been used.

As shown in Table 7, the frequency-based approach alone obtained good results with a TRank of 0.513, outperforming a strong baseline with TRank of 0.454 and being close in TRank to the best team (UOW-SHEF-SimpLex) presented in the task which developed a supervised approach with contextual and psycholinguistic features. On the other hand, the proposed embedding approaches did not show good results in detecting the simplicity of the words, moreover, when combined with the other features, they showed a lower TRank score than the individual frequency feature (0.37). When analyzing errors, problems were detected with the classification of multiwords, because the classical embedding models receive uniwords as inputs, they did not assign a weight to the multiwords, consequently, classifying it as the most complex term in the list and therefore, obtaining wrong results in many cases. In the case of the results for the BERT model, we believe that by performing a fine-tuning process as was done in the CWI stage, it could improve the results in this task.

[10] https://mccormickml.com/2016/04/12/googles-pretrained-word2vec-model-in-python/
[11] https://fasttext.cc/docs/en/crawl-vectors.html
[12] https://github.com/explosion/sense2vec
[13] http://corpus.rae.es/lfrecuencias.html
[14] https://www.english-corpora.org/coca/

Table 7
Results for SR on English test dataset

	TRank
Baseline-L-Sub Gold (SemEval-2012 approach)	0.454
Frequency Feature	**0.513**
Word2vec (Semantic Similarity)	0.168
FastText (Semantic Similarity)	0.1882
Sense2Vec (Semantic Similarity)	0.142
BERT prediction	0.177
Frequency-BERT-FastText (Combined features)	0.37
UOW-SHEF-SimpLex (SemEval-2012 approach)	**0.602**

8. Conclusions and Future Work

The main objective of this work is to explore the possible uses of recent Spanish word embeddings for each of the stages of the LS task in the Spanish language, which has limited resources.

To achieve this goal, as a first step, we fine-tuned BETO and mBERT models to perform the NER task to discern between complex and simple words. The experiments showed a fair result against other supervised approaches, however, there is room for improvement. On the generator side, the performance of different types of embeddings was explored. By analyzing the results, we can understand that embeddings are not recommended for this stage, due to the presence of antonymy in the near neighbors of a target word. In contrast, when evaluating the similarity between words, the embeddings models showed better results in the selectors, such is the case of the FastText model that obtained a higher precision than in previous works. Finally, in the SR stage, a weighting system using information extracted from frequency dictionaries and embedding models was proposed to choose the simplest candidate. When adapted and evaluated in English, the best results were obtained for the frequency features and showed room for improvement with the embedding features.

As future work, the incorporation of multiwords in the fine-tuning process should be contemplated for the task of CWI and SR, because they were the main cause of the drop in the respective scores for each task. Also, a round-trip evaluation is necessary to determine the results of a complete system.

Acknowledgments

This work has been supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M17), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

References

[1] M. Maddela, W. Xu, A word-complexity lexicon and a neural readability ranking model for lexical simplification, arXiv preprint arXiv:1810.05754 (2018).

[2] G. H. Paetzold, L. Specia, A survey on lexical simplification, Journal of Artificial Intelligence Research 60 (2017) 549–593.

[3] R. Alarcon, L. Moreno, P. Martínez, Lexical simplification system to improve web accessibility, IEEE Access 9 (2021) 58755–58767. doi:10.1109/ACCESS.2021.3072697.

[4] E. Barbu, M. T. Martín-Valdivia, L. A. U. Lopez, Open book: a tool for helping asd users' semantic comprehension, in: Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility, 2013, pp. 11–19.

[5] R. Evans, C. Orasan, I. Dornescu, An evaluation of syntactic simplification rules for people with autism, Association for Computational Linguistics, 2014.

[6] E. Barbu, M. T. Martín-Valdivia, E. Martínez-Cámara, L. A. Urena-López, Language technologies applied to document simplification for helping autistic people, Expert Systems with Applications 42 (2015) 5076–5086.

[7] J. Carroll, G. Minnen, Y. Canning, S. Devlin, J. Tait, Practical simplification of english newspaper text to assist aphasic readers, in: Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology, 1998, pp. 7–10.

[8] S. Devlin, G. Unthank, Helping aphasic people process online information, in: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, 2006, pp. 225–226.

[9] L. Sauvan, N. Stolowy, C. Aguilar, T. François, N. Gala, F. Matonti, E. Castet, A. Calabrese, Text simplification to help individuals with low vision read more fluently, in: LREC 2020-Language Resources and Evaluation Conference, 2020, pp. 11–16.

[10] R. Wilkens, B. Oberle, A. Todirascu, Coreference-based text simplification, in: Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI), 2020, pp. 93–100.

[11] L. Rello, R. Baeza-Yates, S. Bott, H. Saggion, Simplify or help? text simplification strategies for people with dyslexia, in: Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, 2013, pp. 1–10.

[12] H. Saggion, E. Gómez-Martínez, E. Etayo, A. Anula, L. Bourg, Text simplification in simplext: Making texts more accessible, Procesamiento del lenguaje natural (2011) 341–342.

[13] S. Bott, L. Rello, B. Drndarević, H. Saggion, Can spanish be simpler? lexsis: Lexical simplification for spanish, in: Proceedings of COLING 2012, 2012, pp. 357–374.

[14] H. Saggion, Automatic text simplification: Synthesis lectures on human language technologies, vol. 10 (1), California, Morgan & Claypool Publishers (2017).

[15] M. Shardlow, A comparison of techniques to automatically identify complex words., in: 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop, 2013, pp. 103–109.

[16] S. M. Yimam, S. Stajner, M. Riedl, C. Biemann, Multilingual and cross-lingual complex word identification., in: RANLP, 2017, pp. 813–822.

[17] S. M. Yimam, C. Biemann, S. Malmasi, G. H. Paetzold, L. Specia, S. Štajner, A. Tack,

M. Zampieri, A report on the complex word identification shared task 2018, arXiv preprint arXiv:1804.09132 (2018).
[18] J. Burstein, J. Shore, J. Sabatini, Y.-W. Lee, M. Ventura, The automated text adaptation tool, in: Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2007, pp. 3–4.
[19] E. Pavlick, C. Callison-Burch, Simple ppdb: A paraphrase database for simplification, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2016, pp. 143–148.
[20] S. R. Thomas, S. Anderson, Wordnet-based lexical simplification of a document., in: KONVENS, 2012, pp. 80–88.
[21] G. Paetzold, L. Specia, Text simplification as tree transduction, in: Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, 2013.
[22] G. Paetzold, L. Specia, Lexenstein: A framework for lexical simplification, in: Proceedings of ACL-IJCNLP 2015 System Demonstrations, 2015, pp. 85–90.
[23] R. Baeza-Yates, L. Rello, J. Dembowski, Cassa: A context-aware synonym simplification algorithm, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 1380–1385.
[24] C. Horn, C. Manduca, D. Kauchak, Learning a lexical simplifier using wikipedia, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2014, pp. 458–463.
[25] G. Paetzold, L. Specia, Lexical simplification with neural ranking, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017, pp. 34–40.
[26] J. Qiang, Y. Li, Y. Zhu, Y. Yuan, X. Wu, Lexical simplification with pretrained encoders, 2020. arXiv:1907.06226.
[27] R. Alarcon, Dataset of sentences annotated with complex words and their synonyms to support lexical simplification, 2021. URL: https://data.mendeley.com/datasets/ywhmbnzvmx/2.
[28] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[29] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020.
[30] E. F. Tjong Kim Sang, F. De Meulder, Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition, in: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 2003, pp. 142–147. URL: https://www.aclweb.org/anthology/W03-0419.
[31] A. Trask, P. Michalak, J. Liu, sense2vec - a fast and accurate method for word sense disambiguation in neural word embeddings, 2015. arXiv:1511.06388.
[32] C. Cardellino, Spanish Billion Words Corpus and Embeddings, 2019. URL: https://crscardellino.github.io/SBWCE/.
[33] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, 2018. arXiv:1802.06893.
[34] R. Navigli, S. P. Ponzetto, Babelnet: Building a very large multilingual semantic network,

in: Proceedings of the 48th annual meeting of the association for computational linguistics, 2010, pp. 216–225.

[35] L. Specia, S. K. Jauhar, R. Mihalcea, SemEval-2012 task 1: English lexical simplification, in: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), Association for Computational Linguistics, Montréal, Canada, 2012, pp. 347–355. URL: https://aclanthology.org/S12-1046.

[36] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, ArXiv abs/1910.01108 (2019).