Reliable Lexical Simplification for Non-Native Speakers

Gustavo Henrique Paetzold
Department of Computer Science
University of Sheffield
Sheffield, United Kingdom
ghpaetzold1@sheffield.ac.uk

Abstract

Lexical Simplification is the task of modifying the lexical content of complex sentences in order to make them simpler. Due to the lack of reliable resources available for the task, most existing approaches have difficulties producing simplifications which are grammatical and that preserve the meaning of the original text. In order to improve on the state-of-the-art of this task, we propose user studies with non-native speakers, which will result in new, sizeable datasets, as well as novel ways of performing Lexical Simplification. The results of our first experiments show that new types of classifiers, along with the use of additional resources such as spoken text language models, produce the state-of-the-art results for the Lexical Simplification task of SemEval-2012.

1 Introduction

Lexical Simplification (LS) is often perceived as the simplest of all Text Simplification sub-tasks. Its goal is to replace the complex words and expressions of a given sentence with simpler alternatives of equivalent meaning. However, this is a very challenging task as the substitution must preserve both original meaning and grammaticality of the sentence being simplified.

However, this is a very challenging task as the substitution needs to ensure grammaticality and meaning preservation. Most LS strategies in the literature are structured according to the pipeline illustrated in Figure 1, which is an adaptation of the one proposed by (Shardlow, 2014).

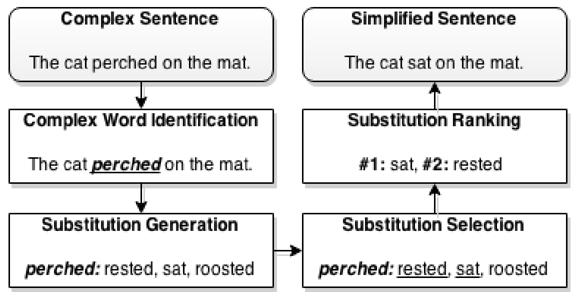

Figure 1: Lexical Simplification pipeline

In this thesis, we intend to identify and address the major limitations of the approaches in the literature with respect to each step of the LS pipeline of Figure 1. In an effort to create new reliable datasets for LS and to unveil information about the needs of those who can most benefit from Text Simplification, we propose new user studies with non-native speakers. We also present novel modelling strategies for each step of the LS pipeline with respect to the limitations of the approaches in the literature.

2 Lexical Simplification: A Survey

To our knowledge, there are no examples of studies which compare the performance of LS approaches in their entirety. For this reason, we choose instead to discuss the merits and limitations of strategies used by authors to address each step of the LS pipeline.

2.1 Complex Word Identification

The goal of Complex Word Identification (CWI) is to identify which words in a given sentence need to be simplified. Some authors, such as (Devlin and Tait, 1998), (Carroll et al., 1998) and (Carroll et al.,

Proceedings of NAACL-HLT 2015 Student Research Workshop (SRW), pages 9–16,
Denver, Colorado, June 1, 2015.

1999) choose to not address this task, but as shown in (Paetzold and Specia, 2013), this can lead to the production of incoherent and/or ungrammatical sentences. Several categories of CWI strategies can be found in literature:

Lexicon-Based Explore the hypothesis that, if a word w is part of a lexicon L of complex/simple words, then it does/does not need to be simplified. While (Watanabe and Junior, 2009) and (Aluisio and Gasperin, 2010) use as lexicons books for children, (Elhadad and Sutaria, 2007), (Deléger and Zweigenbaum, 2009) and (Elhadad, 2006) use a database of complex medical terms. Acquiring lexicons can be easy, but they must correlate with the needs of the target audience in question.

Threshold-Based Explore the hypothesis that a threshold t over a word metric $M(w)$ can separate complex from simple words. The most frequently used metrics are word frequency (Bott et al., 2012), (Leroy et al., 2013) and word length (Keskisärkkä, 2012). However, the corpus evaluation of (Bott et al., 2012) shows that determining such threshold t is impractical.

User-Driven Such approaches allow the users themselves to select which words are complex, and simplify them on demand. Although the results obtained by (Devlin and Unthank, 2006) and (Rello et al., 2013) show that this is a very effective strategy, it might be difficult for it to be used in smaller devices, such as phones.

Classification Methods Train classifiers which discriminate between complex and simple words. For English, the SVM approach of (Shardlow, 2013a) is the only example in literature. Although their study shows that their SVM is not able to outperform neither a threshold-based approach or a "simplify everything" method, we believe the results obtained are controversial.

In another study conducted by the same author (Shardlow, 2014) it was found that replacing words which do not need simplification is one of the most frequent mistakes made by naive LS approaches, and hence we believe the results obtained by (Shardlow, 2013a) do not reveal the potential of classification methods in CWI. Also, the dataset used the experiments of (Shardlow, 2013a) was created automatically and did not attempt to model the needs of any particular target audience. A more substantial comparative study between multiple distinct machine learning methods over a more carefully crafted corpus could be a major milestone in the development of more efficient CWI approaches.

2.2 Substitution Generation

The Substitution Generation (SG) task consists in acquiring candidate substitutions for the complex words in a sentence. This task have been approached by authors in two different ways:

Querying Linguistic Databases Resources such as WordNet (Fellbaum, 1998) and UMLS (Bodenreider, 2004) provide large word ontologies, and have been largely used even in modern contributions. The approaches of (Devlin and Tait, 1998), (Sinha, 2012), (Leroy et al., 2013), (Chen et al., 2012), (Elhadad, 2006) and (Nunes et al., 2013) are some examples. The study of (Shardlow, 2014), however, shows that over 42% of the mistakes made by the approach of (Carroll et al., 1998) are caused by WordNet not having simpler synonyms for complex words. Using such resources also limits the cross-lingual capabilities of the approach, since most of those resources are restricted to one or very few languages.

Automatic Generation Consists in automatically generating pairs of related words and paraphrases. The works of (Elhadad and Sutaria, 2007), (Kauchak and Barzilay, 2006) and (Deléger and Zweigenbaum, 2009) focus on extracting paraphrases from comparable documents. The methods of (Paetzold and Specia, 2013), (Feblowitz and Kauchak, 2013), and(Horn et al., 2014) extract pairs of similar expressions from a aligned sentences from Wikipedia and Simple Wikipedia. But although such approaches do not need linguistic databases, they require for other resources, such as parallel corpora, which are also scarce. They can also suffer for extracting too many meaningless substitutions, such as observed in (Paetzold and Specia, 2013).

In order to solve the cross-lingual problem, an SG approach would have to be able to find substitutions by exploiting only resources which are either abundant in most languages or easy to produce. In Section 3 we discuss how we attempt to address this problem.

2.3 Substitution Selection

Substitution Selection (SS) is the task of determining which substitutions fit the context in which a complex word appears, and hence ensuring meaning preservation. SS have been addressed by authors in three ways:

Word Sense Disambiguation Determine the sense of a complex word in a target sentence, and then filter substitutions which do not share such sense. The approaches of (Sedding and Kazakov, 2004) and (Nunes et al., 2013) have proven to be successful in SS alone, but have not been evaluated in practice. The main limitation of this strategy is that it relies on manually constructed sense databases, which are scarce.

Adapted Disambiguation Use surrogate classes to discriminate between the meanings of an ambiguous word. The words' POS tags are used in the works of (Aluisio and Gasperin, 2010), (Yamamoto, 2013) and (Paetzold and Specia, 2013). While using POS tags may help with words of more than one grammatical type, it does not solve the problem of highly ambiguous words.

Semantic Similarity Estimate the semantic similarity between words and verify if they are replaceable. In (Keskisärkkä, 2012) is employed a simple approach: if a pair of words has a synonymy coefficient higher than a threshold, they are replaceable. This approach, however, requires for a database of synonymy levels. The approach of (Biran et al., 2011) solves that by representing the semantic context of words with word vectors estimated over large corpora, then using the cosine distance between vectors as its semantic dissimilarity.

We did not find mentions of Machine Learning methods being applied to SS. Such methods have been used to produce state-of-the-art results in many classification tasks, and hence modelling SS as a classification problem can be a promising strategy.

2.4 Substitution Ranking

Consists in deciding which substitution is the simplest of the ones available. The LS task of SemEval 2012 brought a lot of visibility to the task, and many authors still visit this subject to this day. The three most efficient strategies found in literature are:

Frequency-based Explore the intuition that the more frequently a word is used, the simpler it is. Most authors use raw frequencies from large corpora (Keskisärkkä, 2012), (Leroy et al., 2013), (Aluisio and Gasperin, 2010), (Nunes et al., 2013) or the Kucera-Francis coefficient (Rudell, 1993), (Devlin and Tait, 1998), (Carroll et al., 1998). Although (Brysbaert and New, 2009) points out several issues with the Kucera-Francis coefficient, the results of SemEval 2012 (Specia et al., 2012) show that raw frequencies from the Google 1T corpus outperform almost all other approaches.

Measuring Simplicity Elaborate metrics to represent the simplicity of a word. The metric of (Sinha, 2012) considers the word's length, number of senses and frequency, and have tied in 2nd place in SemEval 2012 with the Google 1T baseline. The other examples in literature, (Biran et al., 2011) and (Bott et al., 2012), were published before SemEval 2012, and hence have not yet been compared to other approaches.

Linear Scoring Functions Rank candidates based on a linear scoring function over various metrics, such as frequency and word length. This strategy is used by the approach that placed 1st in SemEval 2012 (Jauhar and Specia, 2012).

In (Shardlow, 2014) it is shown that word frequencies from spoken text corpora have great potential in SR. In Section 3.4 we describe an experiment which reveals the potential of such resources.

3 Planning and Preliminary Results

In the following Sections, we discuss which challenges we aim to address in the near future, and briefly describe the solutions we intend explore.

3.1 User Studies and Datasets

As pointed out in Section 2, the scarcity of user studies about audiences that may benefit from LS compel authors to treat simplification as a generalised process, forcing them to use datasets such as the Simple Wikipedia, which can be edited by anyone.

Since we do not believe this ideal, we intend to conduct an array of user studies with non-native speakers. We chose such audience because of three main reasons:

Demand Unfamiliarity with a language is not a medical condition that can be cured, and hence such audience is not likely to disappear in the near future.

Convenience Conducting studies with ill or young subjects needs to be done within various ethical constraints, and can be both expensive and time consuming. Although the needs of these audiences should also be addressed, hiring non-native speakers is much easier, and we believe they fit best our time and resource constraints.

Diversity Statistics show that there is a lot of age, nationality and education level diversity among the non-native speakers (Austin et al., 2006). Such diversity allows for us to investigate several interesting hypothesis regarding possible correlations between the subjects' characteristics and difficulty with certain types of words.

We propose two initial user studies:

Identifying Complex Words In this user study, subjects select which words from a given sentence they do not understand the meaning of. From this study we hope to better understand what types of words are challenging for non-native speakers.

It is very important for a reliable Complex Word Identification dataset to be made available in literature. To our knowledge, there is only one contribution in literature that compares different CWI approaches (Shardlow, 2013a), and since the dataset used was not created with respect to the needs of a specific target audience, the results obtained are not very informative.

This study is already being conducted. Several volunteers of various nationalities were asked to select which words they find complex in 40 English sentences each, of which 10 are part of a set which overlaps between 5 volunteers and 30 are unique. The sentences vary between 20 and 40 words in length, and were extracted from 3 distinct sources: the CW corpus (Shardlow, 2013b), the LexMturk corpus (Horn et al., 2014) and Wikipedia (Kauchak, 2013). From the CW and LexMturk corpora were extracted 231 and 269 non-spurious sentences, respectively, of which exactly 1 word is deemed complex by an anonymous annotator (more specifically, a Wikipedia editor). From Wikipedia were extracted 11945 sentences which were aligned to an identical sentence from Simple Wikipedia. By selecting such sentences, we hope to be able to judge whether or not those resources can be reliably used for the training of Lexical Simplification approaches for non-native speakers.

So far, 51 volunteers participated, who annotated a total of $2,040$ sentences. A total of $1,261$ distinct complex words ($1,597$ total) were identified, 12% of $10,650$ distinct words ($53,125$ total). The volunteers have distinct education levels (8% High School, 57% Undergraduate and 35% Postgraduate), English proficiency levels (11% Advanced, 18% Pre-Advanced, 18% Upper-Intermediate, 37% Intermediate, 14% Pre-Intermediate, 2% Elementary), and have ages varying between 17 and 38 years old (averaging 24 years old).

Selecting the Simplest Candidate We intend to find out what are the key features taken into consideration by non-native speakers on determining which is the simplest word that fits a given context. Just like in the case of Complex Word Identification, we believe that the creation of a reliable dataset for Substitution Ranking is very important.

The only dataset developed specifically for this purpose is the one presented in SemEval 2012. But since the rankings were produced by only 5 non-native annotators, there are a various examples of ties between two candidate substitutions. Also, all subjects were skilled speakers of the English language, which means that, at best, the dataset captures the LS needs of an audience which may not need LS at all. With a larger dataset annotated by more subjects of the same target audience, we will be able to have a more reliable resource to create novel Substitution Ranking approaches.

3.2 Complex Word Identification Methods

We intend to, based on the new datasets produced in our user studies, propose and evaluate the efficiency of multiple different methods of Complex Word Identification. The methods we intend to evaluate are:

Lexicon-Based Approaches We will compile a selection of corpora and see whether or not we can build lexicons from them which separate complex from simple words. The Simple Wikipedia (Horn et al., 2014) and the SUBTLEX corpus (Brysbaert and New, 2009) are some examples.

Threshold-Based Approaches There are multiple metrics which we plan to use in order to train a threshold-based complex word identifier, some of them are: word frequency in a given corpus, word length, number of syllables, familiarity and age of acquisition.

Machine Learning Assisted By combining metrics and lexicons, we can train many different classification systems by using Machine Learning methods. Support Vector Machines, Gaussian Processes and Decision Trees are some Machine Learning methods which we intend to test on Complex Word Identification.

3.3 Substitution Generation and Selection

We propose an entirely new setup for joint modelling Substitution Generation and Selection. Our approach consists in training classifiers capable of deciding which words w_s of a vocabulary V can replace a target word w_c in a sentence s.

Although this seems like a very challenging task, such an approach could be a very powerful tool for LS. It could possibly dismiss entirely the need of using parallel corpora or linguistic databases for such tasks, and hence provide a cost-effective strategy for LS approaches to be ported to multiple languages. We suggest a two-step solution for this task:

1. Define a set $G \subseteq V$ composed by all words w_s from vocabulary V that can replace a word w_c in sentence s without compromising its grammaticality.

2. Define a set $M \subseteq V$ composed by all words w_s from set G that express the same meaning of w_c in sentence s.

Once set M is determined, one can then use a Substitution Ranking method to select which one of them is the simplest. To create a dataset for this task, we plan to hire volunteer native speakers of the English language to manually judge which words can be part of G and M for a large array of different contexts. The user study data will be composed by several automatically generated substitutions for a set of 50 complex words manually selected from the ones produced in the Complex Word Identification study.

3.4 Substitution Ranking

The findings of the Lexical Simplification Task of SemEval 2012 (Specia et al., 2012) have shown that ranking substitution candidates with respect to their simplicity is not an easy task. In order to improve on the state-of-the-art of Substitution Ranking, we intend to explore the usage of spoken textual content. As discussed in (Brysbaert and New, 2009), frequencies extracted from corpora of spoken text, such as subtitles, tend to correlate better with word familiarity than frequencies of other sources, given that the text in subtitles is mostly composed of speech excerpts from character interactions similar to the ones that frequently occur in real life. In order to evaluate their potential, we conducted a preliminary experiment.

Goal In this experiment, we aim to answer the following question: *Can a language model of spoken text be used to outperform state-of-the-art Substitution Ranking approaches?*

Datasets To build a corpus of spoken text, we have parsed 13 HTML lists of movies and series for children created by IMDB[1] users. A total of $1,793$ IMDB IDs of distinct movies and series were gathered. We then used such IDs to query the OpenSubtitles[2] API in search of subtitles for them. Since their API imposes a limit of 100 downloads per day, so far we were only able to collect subtitles of 163 movies and series. By removing the annotations from the files downloaded, we compiled a corpus of $2,103,237$ sentences. For testing, we chose the SemEval $2,012$ corpus, which contains 300 training instances and $1,710$ test instances. Each instance is composed of a sentence, a target word to be simplified, and a list of candidate substitutions.

Approach To rank the candidate substitutions, we propose a novel binary classification setup for the task. For each training instance, we assign the label

[1]http://www.imdb.com

[2]http://www.opensubtitles.org

1 to the highest ranked candidate, and 0 to the remaining ones. We then train a linear classifier over the data to learn ranking weights for the selected features. In testing, we rank substitution candidates according to their distance to the decision boundary: the furthest they are from the "negative" region, the simpler they are.

Our feature set is composed by 9 different collocational features. Each collocational feature of a candidate substitution c in context s is the log probability produced by KenLM (Heafield et al., 2013), given the language model of a certain corpus, of an n-gram $s_{i-l}^{i-1}\ c\ s_{i+1}^{i+r}$, where i is the position of the target complex word in s, and both l and r are token windows in the interval $[0:2]$. If l and r are 0, then the collocational feature says respect to the probability of candidate c independent of context s.

Evaluation Metrics We have chosen the TRnk and *recall-at-n* measures proposed by (Specia et al., 2012) to estimate the performance of our approach. The TRnk calculates the ratio with which a given approach has correctly ranked at least one of the highest ranked substitutions on the gold-standard, while *recall-at-n* measures the coverage of correctly ranked candidates until position $1 \leq n \leq 3$. The reason for using such metrics instead of a ranking score is that we believe they best represent the goal of the task in practice, which is selecting the simplest substitution possible for a complex word.

Results Table 1 shows a performance comparison between the highest ranking approach of SemEval 2012 and our novel strategy trained with 10-fold cross validation over the training set. We extract collocational features from 4 distinct corpora: our corpus of IMDB subtitles (SubIMDB), the Simple Wikipedia corpus (Horn et al., 2014), composed of $505,254$ sentences, the SUBTLEX corpus (Brysbaert and New, 2009), composed of $6,043,188$ sentences taken from assorted subtitles, and the concatenation of SubIMDB and SUBTLEX.

The results show that our strategy outperforms the former state-of-the-art approach of SemEval 2012 by around 5% in TRank and 3% in *recall-at*-1. The *recall-at*-2 and 3 results, although lower than SemEval's best, showcase not a limitation, but rather an advantage of our binary classification setup: by focusing on the task's goal in practice, we are able

Table 1: TRank and *recall-at-n* results obtained

Corpus	TRnk	n=1	n=2	n=3
Best SemEval	0.602	0.575	**0.689**	**0.769**
IMDB+LEX	**0.654**	**0.607**	0.594	0.658
SUBTLEX	0.638	0.592	0.584	0.658
SubIMDB	0.628	0.583	0.578	0.637
Simple Wiki	0.601	0.558	0.571	0.645

to optimize not the correlation between the learned rankings and the gold-standard, but instead the likelihood of the best candidate substitution to be ranked first. We can also notice from the results that, when trained with features extracted from the SubIMDB corpus, our approach performs similarly than when trained with the SUBTLEX corpus, which is 3 times larger. This phenomena suggests that restricting the domain of the subtitles selected to that of movies targeting younger audiences may help ranking approaches in capturing word simplicity.

In the future, we want to experiment with other types of language models, and also explore the potential of other types of spoken content, such as song lyrics and online conversations.

4 Final Remarks and Future work

In this paper we described a thesis proposal which focuses in providing studies on the needs of non-native speakers in terms of LS, producing more reliable datasets for various tasks of the LS pipeline, and devising novel solutions to the limitations of modern LS approaches. We have provided a thorough discussion on the state-of-the-art of LS, a detailed plan of the activities to be conducted throughout the doctorate program and the results of our first experiment, in which we managed to achieve state-of-the-art results for the task of Substitution Ranking.

In the future, we intend to study the simplification needs of other target audiences and explore LS strategies that go beyond replacing complex words and expressions for simpler equivalents, such as by removing unimportant information and learning deep simplification rules from parallel corpora by combining constituency and dependency parses.

References

Aluisio, S. and Gasperin, C. (2010). *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, chapter Fostering Digital Inclusion and Accessibility: The PorSimples project for Simplification of Portuguese Texts, pages 46–53. Association for Computational Linguistics.

Austin, M., Paul, B., and Phil, B. (2006). Current state of english-language learners in the u.s. k-12 student population.

Biran, O., Brody, S., and Elhadad, N. (2011). Putting it Simply: a Context-Aware Approach to Lexical Simplification. *Proceedings of the 49th Annual Meeting of the Association for Computaional Linguistics*, pages 496–501.

Bodenreider, O. (2004). The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*.

Bott, S., Rello, L., Drndarevic, B., and Saggion, H. (2012). Can Spanish Be Simpler ? LexSiS : Lexical Simplification for Spanish Puede ser el Español más simple ? LexSiS : Simplificación Léxica en Español.

Brysbaert, M. and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*.

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

Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S., and Tait, J. (1999). Simplifying Text for Language Impaired Readers. *Proceedings of the 9th Conference of the European Chapter of the ACL (EACL '99)*.

Chen, H.-b., Huang, H.-h., Chen, H.-h., and Tan, C.-t. (2012). A Simplification-Translation-Restoration Framework for Cross-Domain SMT Applications.

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

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.

Devlin, S. and Unthank, G. (2006). Helping aphasic people process online information. *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*.

Elhadad, N. (2006). Comprehending Technical Texts : Predicting and Defining Unfamiliar Terms. pages 239–243.

Elhadad, N. and Sutaria, K. (2007). Mining a lexicon of technical terms and lay equivalents. *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*.

Feblowitz, D. and Kauchak, D. (2013). Sentence simplification as tree transduction. *Proc. of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 1–10.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Heafield, K., Pouzyrevsky, I., Clark, J. H., and Koehn, P. (2013). Scalable modified Kneser-Ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 690–696, Sofia, Bulgaria.

Horn, C., Manduca, C., and Kauchak, D. (2014). Learning a Lexical Simplifier Using Wikipedia. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 458–463.

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

Kauchak, D. (2013). Improving text simplification language modeling using unsimplified text data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1537–1546, Sofia, Bulgaria. Association for Computational Linguistics.

Kauchak, D. and Barzilay, R. (2006). Paraphrasing for automatic evaluation. *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*.

Keskisärkkä, R. (2012). Automatic text simplification via synonym replacement.

Leroy, G., Endicott, E. J., Kauchak, D., Mouradi, O., and Just, M. (2013). User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. *Journal of Medical Internet Research (JMIR)*.

Nunes, B. P., Kawase, R., Siehndel, P., Casanova, M. a., and Dietze, S. (2013). As Simple as It Gets - A Sentence Simplifier for Different Learning Levels and Contexts. *2013 IEEE 13th International Conference on Advanced Learning Technologies*.

Paetzold, G. H. and Specia, L. (2013). Text simplification as tree transduction. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.

Rello, L., Baeza-Yates, R., Bott, S., and Saggion, H. (2013). Simplify or help?: text simplification strategies for people with dyslexia. *Proceedings of the 10th W4A*.

Rudell, A. P. (1993). Frequency of word usage and perceived word difficulty: Ratings of Kuera and Francis words. *Behavior Research Methods*.

Sedding, J. and Kazakov, D. (2004). Wordnet-based text document clustering. In *Proceedings of the 3rd Workshop on Robust Methods in Analysis of Natural Language Data*. Association for Computational Linguistics.

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

Shardlow, M. (2013b). *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, chapter The CW Corpus: A New Resource for Evaluating the Identification of Complex Words, pages 69–77. Association for Computational Linguistics.

Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications 2014*, pages 58–70.

Sinha, R. (2012). UNT-S IMPRANK : Systems for Lexical Simplification Ranking. pages 493–496.

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*. Association for Computational Linguistics.

Watanabe, W. and Junior, A. (2009). Facilita: reading assistance for low-literacy readers. *Proceedings of the 2010 international cross-disciplinary workshop on Web accessibility*.

Yamamoto, T. (2013). Selecting Proper Lexical Paraphrase for Children. *Proceedings of the Twenty-Fifth Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*.