Towards A Word Complexity Score for Individuals with Dyslexia

Evelina Rennes and **Daniel Holmer**
Department of Computer and Information Science
Linköping University
firstname.lastname@liu.se

Abstract

Individuals with dyslexia is one often-addressed receiver of automatic text adaptation techniques. However, the methods and techniques developed today are often rather general in their nature, as they are trained on general Easy Language data, or evaluated with general evaluation metrics. In this work, we aim to take one step closer towards a clearer user focus in the adaptation field. By combining what is known about specific reading challenges of individuals with dyslexia with what we learnt from actually testing the audience, we aim to find a baseline ranking of word-level features to use in our work towards a dyslexia-specific word complexity score.

1 Introduction

Automatic Text Adaptation (ATA) denotes the process of adapting a text to a specific reader or reader audience in order to facilitate reading. ATA mainly includes Automatic Text Simplification (ATS) techniques, but may also include other natural language processing techniques (such as automatic text summarisation) or adaptations related to the typography of a text.

Studies of ATA and ATS often target poor readers. Commonly described target audiences include individuals with dyslexia, aphasia, second language (L2) learners of a language, children or individuals with intellectual disability. However, even if the reader audience is mentioned as a possible receiver of the adapted text, the actual readers are seldom included in the studies. The techniques are often built using general resources, such as the standard and simple versions of the English Wikipedia (Zhu et al., 2010; Coster and Kauchak, 2011; Woodsend and Lapata, 2011; Hwang et al., 2015; Zhang and Lapata, 2017), or following general Easy Language guidelines. To evaluate the adapted text against representatives of the targeted audience seems like a logical way of including the reader, but such evaluations are rather time-consuming and cumbersome, and have the disadvantage of being less generalisable than automatic evaluation metrics such as BLEU (Papineni et al., 2002), SARI (Xu et al., 2016) and SAMSA (Sulem et al., 2018). In short, there is a consensus that the reader audience should be addressed to a larger extent than what is done today, but how this should be done remains to be formulated.

The work described in this paper is one step towards a more clear user focus in the ATA field. By combining what is known about specific reading challenges of one of the target audiences of ATA—individuals with dyslexia—with what we learnt from actually testing the audience, we aimed to find a baseline ranking of word-level features to use in our work towards a dyslexia-specific word complexity score.

2 Related Work

This section describes previous work on the related topics.

2.1 Dyslexia

Dyslexia is a specific learning disability commonly addressed in automatic text adaptation research. Dyslexia is believed to be connected to deficits in the phonological processing (Vellutino, 1987), but it has also been suggested as a deficit involving several cognitive abilities, such as working memory and auditory temporal processing (Fostick and Revah, 2018). This view of dyslexia as a multi-deficit disorder could explain the many differences and sub-types described in this group of readers.

Readers with dyslexia often experience difficulties when decoding words, which is connected to problems to establish the grapheme-phoneme correspondences needed for decoding (Vellutino et al., 2004). Even if many individuals with dyslexia learn to decode, the decoding process is demanding in terms of cognitive resources, which makes it difficult to perform other mental operations (Høien and Lundberg, 2013). For instance, individuals with dyslexia have limited ability to make various types of inferences (Simmons and Singleton, 2000). It has been described that individuals with dyslexia often struggle with long and low-frequent words (Rello et al., 2013b), homophones, words that are orthographically similar, new words, and non-words (Rello et al., 2013a).

2.2 Word-level Complexity

The research area of text complexity and readability is extensively studied, and numerous readability formulae and text complexity features have been developed. For English, common readability measures include the *Flesch Reading Ease Score* (Flesch, 1948), the *Flesch-Kincaid Grade Level* (Kincaid et al., 1975) and the *Dale-Chall Readability Formula* (Chall and Dale, 1995) which also considers the ratio of *difficult words*. For Swedish, the traditional LIX (Björnsson, 1968) is commonly used. LIX is given in Equation 1, where the number of words is denoted by *n(w)*, and the number of sentences is denoted by *n(s)*. More recent approaches for Swedish text complexity include the SVIT model of readability Heimann Mühlenbock (2013) and the SCREAM features (Falkenjack, 2018).

$$\text{LIX} = \frac{n(w)}{n(s)} + (\frac{n(words\ >\ 6\ chars)}{n(w)} \times 100) \quad (1)$$

Word-level complexity is, however, a slightly different—and less studied—task than text-level complexity. Regarding the complexity of specific words, it is commonly addressed in the natural language processing task *Complex Word Identification* (CWI), which is one of the steps of the lexical simplification pipeline (Shardlow, 2014; Paetzold and Specia, 2017). CWI has gained some attention in the natural language processing field through the shared tasks on CWI (Paetzold and Specia, 2016; Yimam et al., 2018).

The issue of lexical complexity of words has been thoroughly studied for Swedish second language learners (Alfter, 2021; Alfter and Volodina, 2018), where classification methods were used to reveal feature importance for this audience. Alfter (2021) highlighted the multifaceted nature of lexical complexity; as it may vary along different dimensions, or show somewhat contradictory tendencies. For instance, word frequency is a common measure of word complexity as more frequently used words are considered less complex. However, as pointed out by Alfter (2021), more frequent words also tend to be polysemous to a higher degree, which in turn could indicate a higher level of complexity.

3 Procedure

In order to gain insight into the features that might be important for determining dyslexia-specific complexity, we trained an SVM classifier to distinguish between Easy Language words and standard words. The idea is that the predicting features of a linear classifier would provide useful insights in how the given word-level features should be weighted in a future word-level complexity score. This section describes the features and resources used, as well as the procedure for training and evaluating the classifier.

3.1 Features of Word Complexity

The word-level features used in this study are presented in Table 1. The features word frequency, word length, number of orthographic neighbours and number of homonyms are derived from the theoretical knowledge of what constitutes a difficult word for a person with dyslexia. The rest of the features are derived from a number of comprehension and decoding tests that adolescents with dyslexia conducted as a part of the research project TEXTAD. A qualitative analysis of a survey conducted after the reading comprehension tests revealed some problematic aspects at the word level. These features were the number of occurrences of some specific characters (*Å*, *Z*, and *X*), the number of digraphs or trigraphs that are specific for Swedish and can be spelled in different ways, such as spelling variants of the "sj"-sound, occurrences of two vowels in a row, double consonants, and compound words.

3.2 Resources

In this work, we consulted several different resources. The features were calculated using various sources. For the number of orthographic neighbours, word frequency, and homonyms, we consulted the AFC LIST (Witte et al., 2021; Witte and Köbler, 2019).

Feature	Definition
Word frequency	Absolute word frequencies
Word length	Word length in characters
Number of Orthographic Neighbours	The number of words that differ from the given word by one letter insertion, deletion, or replacement
Number of Homonyms	The sum of possible homographs and homophones to a given word
Number of ÅZX	Occurrences of the characters Å, Z, and X
Number of Special Digraphs or Trigraphs	The number of digraphs or trigraphs that are specific for Swedish and can be spelled in different ways, such as spelling variants of the ”sj”-sound.
Number of Double Vowels	The number of subsequent vowels in a word. For instance, the word ”jour” would score 1 on this feature.
Number of Double Consonants	The number of double occurrences of the same consonant, such as *tt*, *ll*, or *rr* in a word.
Number of Compound Roots*	The number of root morphemes in a compound word.

*This feature was excluded from the classification since it failed to provide a compound analysis to a majority of the words.

Table 1: The dyslexia-specific features used for classification.

For the classification of Easy Language and standard language words, we used NYLLEX (Holmer and Rennes, 2022) and SUC (Ejerhed et al., 2006).

NYLLEX is a lexical resource of 6,668 words compiled from books written in Easy Language, annotated with information about absolute and dispersed frequency over the various readability levels (1–6).

THE STOCKHOLM-UMEÅ CORPUS (or SUC) is a balanced corpus of Swedish texts from the ’90s. In total, the corpus comprises 1,166,593 tokens and 74,245 sentences. However, for our purpose, we filtered SUC on categories to create a better balance between our two resources. Since the Easy Language books used for the assembly of NYLLEX mostly could be described as either fiction, biographies, or nonfiction (historical/scientific), we only included texts from the corresponding categories K (fiction), G (biographies/essays), and F (popular lore) of SUC. The following subcorpus comprises 406,363 tokens in total, distributed over 31,032 entries.

By computing the log-likelihood of each word in one corpus in relation to the other, we could find words that were significantly more frequent in one resource over the other. We could then label the given word to the resource where it had the highest normalised relative frequency count. This gave us a dataset of 2,319 words that were assigned as typical to SUC and 1,030 words that were assigned as typical to NYLLEX.

3.3 Classification

In order to analyse the descriptive power of our identified features of word complexity, we calculated them on the dataset of typical SUC and NYLLEX words and used them to train a linear SVM for classification. Since the dataset was unbalanced (the SUC words overpowered the NYLLEX words in a 2:1 manner), we oversampled the NYLLEX minority class to a near 1:1 ratio between the classes. The SVM was implemented with the SCIKIT-LEARN package (Pedregosa et al., 2011) in Python. All features were normalised and standardised before training a linear SVM model with the default *regularization* parameter. To evaluate the performance of the model, we used a 10-fold stratified cross-validation procedure. Following this, we trained a classifier with the same settings, but on the whole dataset, and extracted its coefficients to perform an initial investigation of the impact of each feature for the classification.

4 Results

Over the 10-fold cross validation, the mean F1-score of the linear SVM classifier was 0.655 with a standard deviation of 0.033.

Figure 1 displays the coefficients of each of the features used for classification in the SVM model. The number of characters of a given word seemed

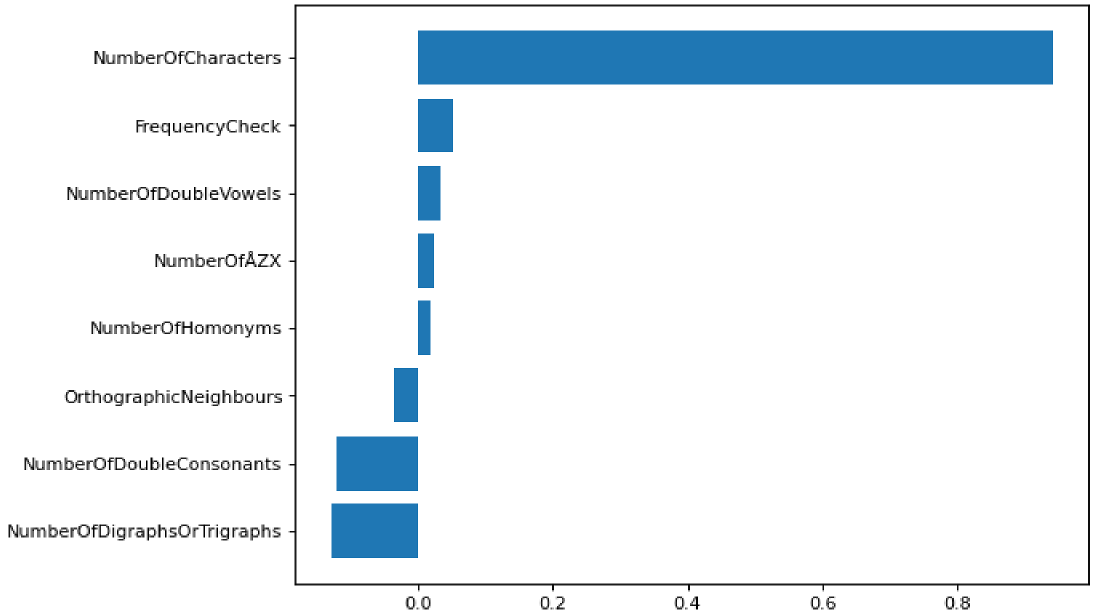

Figure 1: Feature coefficients extracted from the SVM model.

to have the most influence over the prediction for the positive class SUC. The number of double consonants and number of specific digraphs and trigraphs are conversely more involved in the prediction of the negative class NyLLex, although to a lesser extent than the number of characters-feature for the positive class.

5 Discussion

From the results, it was clear that word length in characters seemed to be the most important feature to classify simple and standard words. This is somewhat expected, since the Easy Language resource consists of books where the writers were explicitly instructed to use LIX as a guiding tool. The LIX formula rewards short words and it is, thus, reasonable to think that a lexical resource extracted from such books would exhibit LIX-like characteristics.

The aim of this paper was to find a baseline ranking of word-level features to use in our work towards a dyslexia-specific word complexity score. There are, however, many other feature ranking methods that could have been used, and it is conceivable that other methods might yield slightly varying results. However, the idea is that this baseline setting will be evaluated on readers with dyslexia, in order to see whether or not their experienced difficulty corresponds to the classification results. By revising the feature coefficients accordingly, we hope to get a more fine-tuned word-level complexity of what might constitute a difficult word for a dyslectic person.

In this work, we used a resource which is based on general Easy Language resources. Ideally, we would have had a resource derived from texts targeting individuals with dyslexia. However, this was not possible as no such resource is available today. Similarly, we hypothesised that the SUC words would be representative of standard language. Although this might be true in one sense, SUC might also contain words that are considered "easy", just as the NyLLex dataset might comprise words that are bad representatives of "easiness". However, we believe that at the corpus level, the used resources are sufficient representatives of easy and standard language to work for our purposes.

6 Conclusion

This paper describes on-going work on the development of a dyslexia-specific word-level complexity score. By the use of an SVM classifier trained on easy and standard words, we extracted feature weights to use as coefficients in a baseline score. Future work includes annotation of the words to see whether actual readers agree with the classifier, and a consecutive fine-tuning of the given features. We also aim to provide an analysis of compound words to include into the feature set.

Acknowledgments

This research was supported by Vetenskapsrådet.

References

David Alfter. 2021. *Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective*. Ph.D. thesis, Department of Swedish, University of Gothenburg, Gothenburg, Sweden.

David Alfter and Elena Volodina. 2018. Towards single word lexical complexity prediction. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, pages 79–88.

Carl Hugo Björnsson. 1968. *Läsbarhet*. Liber, Stockholm.

Jeanne S. Chall and Edgar Dale. 1995. *Readability revisited: The new Dale–Chall readability formula*. Brookline Books, Cambride, MA.

William Coster and David Kauchak. 2011. Simple English Wikipedia: A New Text Simplification Task. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 665–669.

Eva Ejerhed, Gunnel Källgren, and Benny Brodda. 2006. Stockholm Umeå Corpus version 2.0.

Johan Falkenjack. 2018. *Towards a model of general text complexity for Swedish*. Licentiate thesis, Linköping University Electronic Press.

Rudolf Flesch. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Leah Fostick and Hadas Revah. 2018. Dyslexia as a multi-deficit disorder: Working memory and auditory temporal processing. *Acta psychologica*, 183:19–28.

Katarina Heimann Mühlenbock. 2013. *I see what you mean. Assessing readability for specific target groups*. Dissertation, Språkbanken, Dept of Swedish, University of Gothenburg.

Torleiv Høien and Ingvar Lundberg. 2013. *Dyslexi - från teori till praktik*, 2., [uppda edition. Natur & Kultur, Stockholm.

Daniel Holmer and Evelina Rennes. 2022. NyLLex: A novel resource of swedish words annotated with reading proficiency level. In *Proceedings of the Language Resources and Evaluation Conference*, pages 1326–1331, Marseille, France. European Language Resources Association.

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning sentences from Standard Wikipedia to Simple Wikipedia. In *HLT-NAACL*, pages 211–217.

J. Peter Kincaid, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. Technical report, U.S. Naval Air Station, Millington, TN.

Gustavo Paetzold and Lucia Specia. 2016. Semeval 2016 task 11: Complex word identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 560–569.

Gustavo Henrique Paetzold and Lucia Specia. 2017. A survey on lexical simplification. *Journal of Artificial Intelligence Research*, 60:549–593.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830.

Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013a. Simplify or help? Text simplification strategies for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 1–10.

Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013b. Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*, pages 203–219. Springer.

Matthew Shardlow. 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)*, pages 1583–1590.

Fiona Simmons and Chris Singleton. 2000. The reading comprehension abilities of dyslexic students in higher education. *Dyslexia*, 6(3):178–192.

Elior Sulem, Omri Abend, and Ari Rappoport. 2018. Semantic structural evaluation for text simplification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 685–696, New Orleans, Louisiana. Association for Computational Linguistics.

Frank R Vellutino. 1987. Dyslexia. *Scientific American*, 256(3):34–41.

Frank R Vellutino, Jack M Fletcher, Margaret J Snowling, and Donna M Scanlon. 2004. Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry*, 45(1):2–40.

Erik Witte, Jens Edlund, Arne Jönsson, and Henrik Danielsson. 2021. Swedish Word Metrics: A Swe-Clarin resource for psycholinguistic research in the Swedish language. In *CLARIN Annual Conference 2021*, page 54.

Erik Witte and Susanne Köbler. 2019. Linguistic materials and metrics for the creation of well-controlled Swedish speech perception tests. *Journal of Speech, Language, and Hearing Research*, 62(7):2280–2294.

Kristian Woodsend and Mirella Lapata. 2011. Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 409–420, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Seid Muhie Yimam, Chris Biemann, Shervin Malmasi, Gustavo H Paetzold, Lucia Specia, Sanja Štajner, Anaïs Tack, and Marcos Zampieri. 2018. A report on the complex word identification shared task 2018. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 66–78.

Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark. Association for Computational Linguistics.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A Monolingual Tree-based Translation Model for Sentence Simplification. *23rd International Conference on Computational Linguistics*, (August):1353–1361.