Evaluating the Readability of Text Simplification Output for Readers with Cognitive Disabilities

Victoria Yaneva*, Irina Temnikova, Ruslan Mitkov***
*Research Institute in Information and Language Processing,University of Wolverhampton,
**Qatar Computing Research Institute,HBKU, Doha, Qatar,
*Wolverhampton,UK, **Doha,Qatar,
V.Yaneva@wlv.ac.uk, itemnikova@qf.org.qa, R.Mitkov@wlv.ac.uk

Abstract

This paper presents an approach for automatic evaluation of the readability of text simplification output for readers with cognitive disabilities. First, we present our work towards the development of the EasyRead corpus, which contains easy-to-read documents created especially for people with cognitive disabilities. We then compare the EasyRead corpus to the simplified output contained in the LocalNews corpus (Feng, 2009), the accessibility of which has been evaluated through reading comprehension experiments including 20 adults with mild intellectual disability. This comparison is made on the basis of 13 disability-specific linguistic features. The comparison reveals that there are no major differences between the two corpora, which shows that the EasyRead corpus is to a similar reading level as the user-evaluated texts. We also discuss the role of Simple Wikipedia (Zhu et al., 2010) as a widely-used accessibility benchmark, in light of our finding that it is significantly more complex than both the EasyRead and the LocalNews corpora.

Keywords: readability, text simplification, cognitive disabilities, autism, easy-to-read

1. Introduction

The main task of automatic text simplification (TS) is to convert texts into a more understandable form for readers with lower than average reading skills, without changing the original meaning of the text. These readers could include children and people with poor literacy (Aluisio et al., 2010), learning disabilities (Feng, 2009), neurodevelopmental disorders (Rello et al., 2013), second language learners (Petersen, 2007), among others.

The evaluation of TS output could be defined through the tasks of quality and accessibility (readability) estimation. Quality estimation is performed at sentence level and its purpose is to establish whether the output sentences are grammatical, meaningful and simpler. Readability estimation on the other hand is done at text level and is concerned with measuring how understandable the output is for the target reader population. In this paper we focus solely on discussing readability evaluation.

Currently, readability of TS output is measured in two ways: automatic evaluation through readability formulae and human evaluation, where sample texts are evaluated through reading comprehension experiments involving target readers. While the latter is certainly the most reliable way to measure text accessibility and harness insights into user preferences, it is very time-consuming and expensive to perform, especially in cases where the target readers are people with cognitive disabilities (intellectual disability, autism, hyperactivity, etc.). To address this issue, research has been focusing on developing automatic methods for readability evaluation, which could account for the specific reading difficulties of various reader populations (Dubay, 2004; Benjamin, 2012), including people with cognitive disabilities (Feng et al., 2010; Yaneva and Evans, 2015). However, relying on readability formulae to measure text accessibility is an approach that has many drawbacks (Benjamin, 2012; Dubay, 2004; Siddharthan, 2004), mainly related to the fact that readability formulae only employ surface text features such as word or sentence length. Thus for example, the Flesch formula (Flesch, 1948) presented below is very widely-used but cannot take into account the difficulty autistic readers have in using abstraction (Minshew et al., 2002) or the difficulty aphasic readers have with passive voice or other syntactic constructions (Berndt et al., 1996).

$$FleschReadingEase = 206.835 - 1.015 \times \frac{words}{sentences} - 84.6 \times \frac{syllables}{words}$$

(Flesch, 1948)

Subtler text characteristics could be accounted for by readability models based on machine learning algorithms. However, developing these models for people with cognitive disabilities is currently not feasible, due to the lack of large enough corpora for model training and evaluation. The question remains of how to automatically evaluate the accessibility of TS output for readers with cognitive disabilities while accounting for the specific reading difficulties they have.

One way to address this issue is to have a collection of texts which are known to be at a suitable level for people with cognitive disabilities. TS output could then be compared to this text collection by employing features which specifically account for the reading difficulties of this population. Exploiting more features than just the ones employed in the readability formulae would allow comparison of different aspects of the text and could cast light on the further types of simplification needed. For example, if there is no statistically significant difference in sentence length between the TS output and the comparison texts but there is difference in word familiarity and age of acquisition, then word substitution with simpler synonyms

would be a more suitable strategy for further TS than sentence splitting.
A suitable source of such texts, known to match the level of reading ability of people with cognitive disabilities, are the so-called easy-to-read documents. These documents are produced by humans following a set of guidelines for accessible writing developed specifically for people with learning difficulties such as the 'Make It Simple' guidelines (Freyhoff et al., 1998) or 'Guidelines for Easy-to-read Materials' (Nomura et al., 2010). The comprehensibility of the easy-to-read documents is further enhanced through the inclusion of images illustrating the main ideas in the text, and is ensured by the evaluation of these documents on a focus group of disabled people (Figure 1). The compilation of a corpus consisting of easy-to-read materials has now become feasible due to the already large number of existing easy-to-read documents produced since the early 2000s, as a result of campaigns promoting accessibility of information.

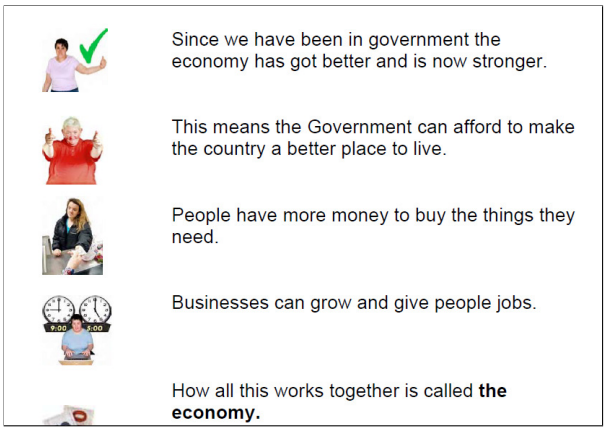

Figure 1: Conservative Party Manifesto Easy-to-Read Version (2015)

With a view to address the problem of evaluating the readability of TS output for people with cognitive disabilities, the main contributions of this paper are as follows:

- *Development and evaluation of the EasyRead corpus* (Section 3.)

- *Matching the main simplification requirements of people with cognitive disabilities as listed in the 'Make It Simple guidelines' (Freyhoff et al., 1998) to linguistic features accounting for these difficulties* (Section 4).

- *Evaluation of the approach by comparing the EasyRead corpus to the simplified documents in the user-evaluated LocalNews corpus (Feng, 2009) and to Simple Wikipedia (Zhu et al., 2010)* (Section 4).

- *Providing further evidence to the argument that Simple Wikipedia is not a suitable gold standard for evaluation of text simplification for people with cognitive disabilities (Xu et al., 2015; Štajner et al., 2012)* (Section 4).

Discussion of the findings is presented in Section 5 and the main conclusions are summarised in Section 6. Section 2 presents related work in the fields of readability, corpora used for TS and easy-to-read documents.

2. Related Work

This section presents related work on readability assessment (Section 2.1), corpora for test simplification (Section 2.2), as well as our previous work on evaluating the understandability of easy-to-read documents for people with autism (Section 2.3).

2.1. Readability

Readability assessment is a construct that takes into account the relationship between specific reader populations, specific texts and the purpose of reading (Pikulski, 1995). Text difficulty is predicted via readability formulae which are equations exploiting surface text features such as word and sentence length, number of suffixes, number of pronouns, etc. (Dubay, 2004). Some of the most widely-used readability formulae are the Flesch Reading Ease formula presented in Section 1 (Flesch, 1948), Flesch-Kincaid Grade Level (Kincaid et al., 1975), Army's Readability Index (ARI) (Senter and Smith, 1967), Fog Index (Gunning, 1952), and the Simple Measure of Gobbledygook (SMOG) (McLaughlin, 1969), etc.
Readability formulae have been shown to be very useful in strengthening trends to bring down the complexity level of texts in textbooks or magazines (Dubay, 2004), but at the same time have been heavily criticised for not taking into account subtler text characteristics such as the organisation of ideas in a text (cohesion) or the cognitive load the text poses on the readers (Benjamin, 2012; Siddharthan, 2006; Dubay, 2004). Addressing these limitations, readability research starts putting more emphasis on the psycholinguistic aspects of the text-reader interaction and this led to the creation of a number of cognitively-motivated features. These features are obtained through human rankings and reflect reader-related aspects of word processing such as familiarity and age of acquisition of common words, as well as their levels of abstractness, concreteness, imagability and meaningfulness (Coltheart, 1981). The common scores for these rankings for 98 538 words are stored in the MRC Psycholinguistic Database (Coltheart, 1981) and have been included in readability assessment software such as CohMetrix (McNamara et al., 2010).
More recent advances in the fields of Natural Language Processing and Artificial Intelligence enable the development of NLP-enhanced features (e.g. n-grams or depth of a parse tree) and their combination into more complex models using machine learning algorithms. These new features and the faster computation of the old features enabled by current computers allow investigation of broader and more novel text types (e.g. web content (Si and Callan, 2001)) and various reader populations. These populations include people with cognitive disabilities such as mild intellectual disability (MID) (Feng, 2009), dyslexia (Rello et al., 2012a; Rello et al., 2012b) or autism (Yaneva and Evans, 2015; Yaneva et al., 2015). (Section 2.3).

For example, people with MID find it difficult to remember within- and between-sentence relations due to the smaller capacity of their working memory (Jansche et al., 2010), which is why suitable readability features reflecting these difficulties were found to be *entity density per sentence* and *lexical chains* (synonymy or hyponymy relations between nouns) (Jansche et al., 2010; Feng, 2009; Huenerfauth et al., 2009). This study on readability assessment for people with MID is currently the only one to employ machine learning readability classifiers for people with cognitive disabilities by using the Weekly Reader corpus (Allen, 2009) as training data and then evaluating on the LocalNews corpus (Feng, 2009). LocalNews consists of 11 original and 11 simplified news stories that have been assessed in terms of their understandability by people with MID. In Section 4 we compare the complexity of this corpus to the complexity of our easy-read corpus.

2.2. Corpora Used for Text Simplification

Currently the output of most text simplification systems is compared to gold standards such as the widely-used Simple English Wikipedia (Zhu et al., 2010) or Encyclopaedia Britannica for Children (Barzilay and Elhadad, 2003), where texts have been manually adapted by humans to match the reading abilities of children and contain two complexity levels:original and simplified documents (Table 1, column 4); in the cases of the Weekly Reader corpus (Allen, 2009) or Literacy Works corpus (Petersen and Ostendorf, 2007) the simplification has been done for second language learners (Table 1).

Corpus	Target readers	Domain	Articles
Newsella	L2 learners	News	1,130 x 4
SimpleWiki	Various	Encylcl	In progress
Britannica	Children	Encylcl	20 x 2
WeeklyReader	L2 learners	News	100 x 2
Literacy works	L2 learners		104 x 2
FIRST	Autism	General	25 x 2
LocalNews	MID	News	11 x 2

Table 1: English-language corpora used in text simplification research

While these corpora are undoubtedly very useful for developing TS systems because they contain pairs of matched original and simplified sentences, they have never been evaluated by the targeted readers (except for LocalNews, Section 2.1). The lack of evaluations casts doubt on the quality of the resource, with evidence showing that Simple Wikipedia is not actually *that* accessible (Štajner et al., 2012). Some researchers appeal for "the community to drop it as the standard benchmark set for simplification" (Xu et al., 2015). In Section 4 of this paper we test this further by comparing a subset of the SimpleWiki corpus to LocalNews and the EasyRead corpora.

In addition, more research is needed to testify whether the quality of texts produced through manual simplification with the primary objective of developing aligned corpora is similar to the quality of accessible documents produced with the reader in mind (e.g. easy-to-read documents). Finally, all of the corpora mentioned above are genre-specific, containing either encyclopedic or newspaper articles, while easy-to-read texts are of a variety of registers such as health, news, lifestyle, politics and other various general topics. In Section 2.3 we discuss previous work on assessing the compliance of these documents collected from the Web with their production guidelines in order to gain insight into their quality, as well as the assessment of the comprehensibility of a few easy-to-read documents for people with autism.

2.3. Easy-to-read Documents and Cognitive Disabilities

In previous work we evaluated the comprehensibility of 7 easy-to-read documents with a group of 20 adults diagnosed with autism and a control group of 20 non-autistic adults matched for age and education level (Yaneva et al., 2015). The results confirmed that the easy-to-read documents were understood well by all participants, but the autistic ones showed more heterogeneous perceptions of their difficulty ranging from very easy (n=54), easy (n=37), medium (n=23) and even difficult (n=4) and very difficult (n=2). The control participants had a predominant rate of very easy (n=117) and easy (n=20) and none of them ranked any text as difficult or very difficult (Yaneva et al., 2015).
We then investigated the compliance of a 150-document sample of easy-to-read documents available on the Web to the official guidelines for easy-to-read document production, affirming that the majority of them meet the required standards and that the Web could be used as a source for obtaining good-quality easy-to-read texts (Yaneva, 2015). Finally, we compared the level of text complexity in the 150-document sample of easy-to-read documents to corpora evaluated on people with mild intellectual disability (LocalNews corpus, (Feng, 2009)) or developed for people with autism (FIRST corpus (Jordanova, 2013), showing that there were no major differences between the text complexity of these corpora (Yaneva, 2015).
Section 3 presents the main characteristics of the EasyRead corpus.

3. The EasyRead Corpus

An initial collection of 372 English-language easy-to-read documents was obtained and their text complexity assessed using the Flesh Reading Ease formula (Flesch, 1948) and the Flesch-Kincaid Grade Level (FKGL) formula (Kincaid et al., 1975). A general rule of thumb is that texts with Flesch index higher than 65 are considered accessible writing (DuBay, 2004), which is why a manual examination of all documents which fell under the threshold of 65 was performed. Thus, certain documents with Flesch index lower than 65 were not discarded even though they included complex terms, because these terms had been adequately explained and the aim was to familiarise the readers with the jargon used in the genre. An example of this type is a document from the General domain called "Terms We Use in This Pack", which has a Flesch score of 53.
After the manual examination of the documents, 19 of them were discarded leaving 353 documents in the corpus.

3.1. Composition

The EasyRead corpus consists of 353 easy-to-read documents and 272,206 words in total, split in three domains: Newspaper articles, Medical documents including healthcare information and medicine leaflets, and General informational articles, which include lifestyle, politics, instructions and other miscellaneous domains (Table 2).

Domain	Files	Words	Sent. Len		Word Len	
			Aver	SD	Aver	SD
News	123	17,449	11.1	4.2	1.13	0.5
Medical	119	105,270	9.7	2.2	1.46	0.5
General	114	149,487	9.97	2.5	1.3	0.5
Total	353	272,206	10.56	2.6	1.32	0.5

Table 2: Composition and basic statistics of the EasyRead corpus, where Sent.Len = sentence length, Word Len = word length, aver = average and SD = standard deviation.

3.2. Sources and Topics

The documents were obtained from a variety of US and UK charity organisation websites (83 documents), government departments (31), healthcare services (119), as well as news websites for people with disabilities (123).

The topics covered in the Medical domain vary from medicine leaflets (34), mental health information (10), advice on making healthcare choices (10), and general medical leaflets from healthcare providers (85). The General domain comprises simple government documents and policies such as party manifestos, guides to voting or the rights of people with disabilities (50), in addition to materials explaining autism-related issues (12), tourist guides (5), alcohol and smoking (5), emotions (7), health tips (10), diet and exercise (11), advice on getting support (4) and other miscellaneous information (10). The News domain contains a set of news from Britain and the world obtained from the year 2014 onwards.

3.3. Corpus Pre-processing

All documents were collected manually and then converted from .pdf to plain text format using an online file converter. After that the plain text files were processed using the Stanford parser (Klein and Manning, 2003), which identified individual words in the texts, their part of speech, syntactic role in the sentence, and anaphoric links within each text.

4. Corpus Characteristics and Comparison to Other Corpora

A set of linguistic features was matched to the rules in one of the manuals for development of easy-to-read content, namely the widely-used "Make It Simple" guidelines (Freyhoff et al., 1998) (Table 3). Each rule was assigned one or more linguistic features as outlined in Table 3. By matching linguistic features to the rules from the "Make It Simple" guidelines, we address those aspects of the text that are particularly relevant to the target population of cognitively disabled readers.

Columns 1 and 2 in Table 3 list the main writing rules from the "Make It Simple" guidelines and their matched linguistic features used in our analysis. Column 3 presents the scores obtained for these features for the EasyRead corpus, divided into two sub-columns presenting their average values and their standard deviations. Columns 5 and 6 show results for the Simple Wikipedia corpus and the LocalNews corpus.

Indicators of text complexity are higher word length, sentence length, age of acquisition, passive voice and negation density, as well as Flesch-Kincaid Grade Level (Kincaid et al., 1975). Indicators of low complexity are high word frequency, familiarity, first and second person pronoun incidence, imagability, concreteness and Flesch Reading Ease (Flesch, 1948).

Some of the rules in the guidelines remained outside of the scope of this study due to lack of relevant linguistic features. These rules were: *Use Practical Examples, Address the Readers in a Respectful Form, Cover Only One Idea per Sentence, Do Not Assume Previous Knowledge, Use Words Consistently, Do Not Use the Subjunctive Tense* and *Be Careful with Metaphors and Figurative Language*.

To test whether the EasyRead corpus could truly be used to evaluate the accessibility of TS output, it should ideally be compared to texts which satisfy the following three conditions:

- Have been produced through TS.
- Have been post-edited by humans so that they are grammatical.
- Have been evaluated on a target group of cognitively disabled individuals.

To the best of our knowledge, the only English-language resource that satisfies all these conditions is the LocalNews corpus by (Feng, 2009), which is why we have chosen it for the evaluation of the EasyRead corpus. We also want to find out how Simple Wikipedia (Zhu et al., 2010), which is the most widely-used corpus in simplification research for English, compares to the user-evaluated LocalNews and to our EasyRead corpus.

The three corpora (the EasyRead corpus, 273 randomly selected articles from Simple Wikipedia and the LocalNews corpus) were assessed using the features presented in Table 3. LocalNews contains 11 original and 11 simplified news stories evaluated on 20 adults with MID (Feng, 2009). In our experiments we use the simplified documents from LocalNews as a gold standard for accessible writing for readers with MID.

4.1. Study Hypotheses

Our study investigates the following hypotheses:

H1: There are no statistically significant differences between the indices obtained for the EasyRead and LocalNews corpora for all linguistic features used in this study.

To strengthen the comparability of topics, for **H1** we only compare the News domain from the ER to LocalNews. In order to find out how SimpleWikipedia compares to our

Writing Rules	Features	EasyRead		SimpleWiki		LocalNews	
		Aver	SD	Aver	SD	Aver	SD
Use short sentences	Av. Sent. Length (words)	10.56	2.57	17.956	3.67	11.454	2.07
Use short, familiar words	Av. Word Length (syllables)	1.32	0 .47	1.558	0.12	1.182	0.4
	Word Frequency	6795	1843	9104	1474	7746	1323
	Age of Acquisition	5.83	0.54	6.584	0.67	6	0.63
	Familiarity	580.43	7.59	568.5	8.59	574.6	10.4
Use active verbs	Agentless Passive Voice Density	11.1	8.8	15.276	8.21	7	9.8
Use positive language	Negation Density	7.37	6.44	4.6	3.81	3.9	3.75
Use many personal words	1st Person Singular Pronoun Inc.	3.14	8.38	.569	2.15	2	5.71
	2nd Person Singular Pronoun Inc.	28.18	25.86	.528	2.5	.55	1.81
Avoid abstract Concepts	Imagability	416.2	25.85	418.05	21.42	432.45	31.23
	Concreteness	387.43	27.13	389.6	24.7	401.6	33.91
Use simple language	Flesch Reading Ease	73.95	8.64	56.77	11.63	72.72	7.95
	Flesch Kincaid Grade Level	5.55	1.48	9.8	2.3	6	1.55

Table 3: Writing rules (Freyhoff et al., 1998) and their corresponding linguistic features for three corpora

user-evaluated corpora we test the following hypothesis:

H2: There are no statistically significant differences between the indices obtained for the SimpleWikipedia and LocalNews corpora for all linguistic features used in this study.

Finally, we want to find out whether the EasyRead corpus is truly simpler than SimpleWikipedia:

H3: There are no statistically significant differences between the indices obtained for the EasyRead and SimpleWikipedia corpora for all linguistic features used in this study.

4.2. Results

A Shapiro-Wilk test showed that the data was non-normally distributed, which is why a nonparametric Wilcoxon Signed Rank test was used in order to compare the differences between the three corpora. Table 4 presents the Z scores of the test, where the values in bold signify statistically significant differences.

As can be seen from Table 4, Simple Wikipedia is significantly more complex than the user-evaluated LocalNews corpus, excluding values related to word familiarity and concreteness, as well as passive voice, negation and personal pronouns. The EasyRead corpus is significantly less complex than Simple Wikipedia according to all values and is much similar in terms of complexity to the user-evaluated gold standard, the only differences being in the age of acquisition of words (higher in the ER sample) and the use of negation (higher in the LocalNews texts).

5. Discussion

The results of the experiments presented above entail several important conclusions. While English Wikipedia and Simple English Wikipedia are valuable in terms of model training due to the paired original and simplified sentences they contain, we show that Simple wikipedia is in fact much more complex than texts which have been understood well by people with mild intellectual disability. The increased complexity lies in the lexical component of the corpus (word length, frequency, age of acquisition and imagability), as well as in the syntactic component (sentence length) and the relation between the two, as evidenced by the Flesch and Flesch-Kincaid indices. These findings dispute the quality of Simple Wikipedia as a gold standard for measuring the accessibility of text simplification output in the context of cognitive disabilities research. This result is yet another argument in favour of using additional criteria for evaluating the accessibility of text simplification output, as previous research (Štajner et al., 2012) has found that Simple Wikipedia is in fact more complex than news texts from the METER corpus (Gaizauskas et al., 2001) and healthcare leaflets for distribution to the general public contained in the British National Corpus (Burnard, 1995).

Comparison between the user-evaluated LocalNews corpus and the EasyRead corpus shows a different pattern. The differences there lie only in the age of acquisition of words and the use of negations, however, their incidences are lower in the EasyRead corpus showing that its complexity with regards to these two phenomena is actually lower. All the rest of the indices show no difference in the overall complexity of the two corpora, supporting the idea that the EasyRead corpus reflects a suitable reading level for people with cognitive disabilities. Finally, a comparison between Simple Wikipedia and the EasyRead corpus reveals that the EasyRead corpus is significantly less complex according to all 13 indices used in this study.

There are several limitations related to this study, one of which is the small size of the corpora. In the case of the EasyRead corpus the small size is due to the relatively small number of easy-to-read documents available due to the relatively recent enforcement of the *accessibility of information* laws, which demanded the creation of such documents and the onerous and expensive process of their writing and evaluation. The small size of the LocalNews corpus is understandably due to the difficulties in recruiting participants from this population and the very high cognitive load that reading comprehension tests impose on participants with

	SL	WL	Freq	AOA	Fam	Concr	Imag	Pass	Neg	1st p.	2nd p.	Flesch	FKRE
ER - LN	-.462	-1.41	-1.96	**-2.0**	-1.74	-1.78	-1.60	-1.89	**-2.7**	-1.34	-.45	-.62	-.66
SW - LN	**-2.93**	**-2.31**	**-2.4**	**-2.22**	-1.07	-1.69	**-2.05**	-1.96	-.89	-1.34	-.45	**-2.49**	**-2.85**
SW - ER	**-5.59**	**-11.7**	**-10.56**	**-11.96**	**-4.74**	**-5.15**	**-6.37**	**-12.1**	**-2.24**	**-2.06**	**-12.4**	**-14**	**-13.93**

Table 4: Wilcoxon test Z scores comparison for the EasyRead (ER), LocalNews (LN) and SimpleWiki (SW) corpora. The values in bold signify statistically significant differences.

cognitive disabilities resulting in a small number of texts being evaluated.

Another limitation of this approach, as well as an avenue for future research, is exploration of the existence of easy-to-read documents in languages other than English. Since easy-to-read documents are a direct product of implementation and enforcement of information accessibility laws and policies, their availability in non-English speaking countries may vary and thus the compilation of such a corpus for some languages may not be feasible.

6. Conclusions

We presented the EasyRead corpus consisting of easy-to-read texts developed specifically for people with cognitive disabilities. Together with a set of features matched to the guidelines for writing texts for this population, the EasyRead corpus can be used as a gold standard for assessment of TS output targeted to people with cognitive disabilities. We also show that Simple Wikipedia is significantly more complex than user-evaluated texts, adding another piece of evidence to the argument against its use as a gold standard for measuring text accessibility.

7. Acknowledgements

The authors would like to thank Dr. Georgiana Marsic for her valuable help with feature extraction.

8. Bibliographical References

Allen, M. L. (2009). Brief report: decoding representations: how children with autism understand drawings. *Journal of autism and developmental disorders*, 39(3):539–43, March.

Aluisio, S., Specia, L., Gasperin, C., and Scarton, C. (2010). Readability assessment for text simplification. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9. Association for Computational Linguistics.

Barzilay, R. and Elhadad, N. (2003). Sentence alignment for monolingual comparable corpora. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, pages 25–32, Stroudsburg, PA, USA. Association for Computational Linguistics.

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24:1–26.

Berndt, R. S., Mitchum, C. C., and Haendiges, A. N. (1996). Comprehension of reversible sentences in agrammatism: a meta-analysis. *Cognition*, 58(3):289 – 308.

Burnard, L. (1995). Users reference guide british national corpus version 1.0.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505.

Dubay, W. H. (2004). *The Principles of Readability*. Impact Information.

Feng, L., Jansche, M., Huenerfauth, M., and Elhadad, N. (2010). A comparison of features for automatic readability assessment. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 276–284, Stroudsburg, PA, USA. Association for Computational Linguistics.

Feng, L. (2009). Automatic readability assessment for people with intellectual disabilities. *SIGACCESS Access. Comput.*, (93):84–91, January.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Freyhoff, G., Hess, G., Kerr, L., Tronbacke, B., and Van Der Veken, K. (1998). Make it Simple. European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability. Technical report, ILSMH European Association.

Gaizauskas, R., Foster, J., Wilks, Y., Arundel, J., Clough, P., and Piao, S. (2001). The meter corpus: a corpus for analysing journalistic text reuse. In *Proceedings of the Corpus Linguistics 2001 Conference*, pages 214–223.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill, New York.

Huenerfauth, M., Feng, L., and Elhadad, N. (2009). Comparing evaluation techniques for text readability software for adults with intellectual disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '09, pages 3–10, New York, NY, USA. ACM.

Jansche, M., Feng, L., and Huenerfauth, M. (2010). Reading difficulty in adults with intellectual disabilities: Analysis with a hierarchical latent trait model. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10, pages 277–278, New York, NY, USA. ACM.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Technical report, CNTECHTRA Research Branch Report.

Klein, D. and Manning, C. D. (2003). Natural language

parsing. In *Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference*, volume 15, page 3. MIT Press.

McLaughlin, H. G. (1969). SMOG grading - a new readability formula. *Journal of Reading*, pages 639–646, May.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., and Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion, May.

Minshew, N. J., Meyer, J., and Goldstein, G. (2002). Abstract reasoning in autism: A disassociation between concept formation and concept identification. *Neuropsychology,*, 16(3):327–334.

Nomura, M., Nielsen, G. S., and Tronbacke, B. (2010). Guidelines for easy-to-read materials/rev. Technical report, International Federation of Library Associations and Institutions, IFLA Headquarters, The Hague.

Petersen, S. E. and Ostendorf, M. (2007). Text simplification for language learners: a corpus analysis. In *SLaTE*, pages 69–72. Citeseer.

Petersen, S. E. (2007). *Natural language processing tools for reading level assessment and text simplification for bilingual education*. Ph.d. thesis, University of Washington, Seattle.

Pikulski, J. J. (1995). Re a d a b i l i t y.

Rello, L., Baeza-yates, R., Dempere-marco, L., and Saggion, H. (2012a). Frequent Words Improve Readability and Shorter Words Improve Understandability for People with Dyslexia. (1):22–24.

Rello, L., Bayarri, C., and Gorriz, A. (2012b). What is wrong with this word? dyseggxia: A game for children with dyslexia. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 219–220, New York, NY, USA. ACM.

Rello, L., Baeza-Yates, R., Dempere-Marco, L., and Saggion, H., (2013). *Human-Computer Interaction – INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV*, chapter Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia, pages 203–219. Springer Berlin Heidelberg, Berlin, Heidelberg.

Senter, R. J. and Smith, E. A. (1967). Automated Readability Index. Technical Report AMRL-TR-6620, Wright-Patterson Air Force Base.

Si, L. and Callan, J. (2001). A statistical model for scientific readability. In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, CIKM '01, pages 574–576, New York, NY, USA. ACM.

Siddharthan, A. (2004). *Syntactic Simplification and Text Cohesion*. Ph.D. thesis, University of Cambridge.

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

Štajner, S., Evans, R., Orasan, C., and Mitkov, R. (2012). What can readability measures really tell us about text complexity? In Luz Rello et al., editors, *Proceedings of the LREC'12 Workshop: Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Yaneva, V. and Evans, R. (2015). Six good predictors of autistic text comprehension. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 697–706, Hissar, Bulgaria, September. INCOMA Ltd. Shoumen, BULGARIA.

Yaneva, V., Temnikova, I., and Mitkov, R. (2015). Accessible texts for autism: An eye-tracking study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '15, pages 49–57, New York, NY, USA. ACM.

Yaneva, V. (2015). Easy-read documents as a gold standard for evaluation of text simplification output. In *Proceedings of the Student Research Workshop*, pages 30–36, Hissar, Bulgaria, September. INCOMA Ltd. Shoumen, BULGARIA.

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, Stroudsburg, PA, USA. Association for Computational Linguistics.