Easy-read Documents as a Gold Standard for Evaluation of Text Simplification Output

Victoria Yaneva
Research Group in Computational Linguistics
University of Wolverhampton
`v.yaneva@wlv.ac.uk`

Abstract

This paper presents initial research into the use of easy-read articles written for people with cognitive disabilities as a gold standard for the evaluation of the output of text simplification systems. We investigate the compliance of the easy-read documents available on the Web with guidelines for development of easy-read material, as well as their suitability as a gold standard for simple documents for two types of populations in particular: adult readers with autism and readers with mild intellectual disability (MID). The results indicate an overall good level of compliance with the guidelines and suggest that easy-read documents are a suitable resource for evaluation of accessible documents produced for adults with autism or MID.

1 Introduction

Common reading materials on the Internet, in newspapers or in textbooks are generally understood by a large part of the population. However, there are readers with disabilities such as intellectual disability, autism, aphasia or dementia, among others, who struggle to comprehend most of these written materials.

To ensure the constitutional right of all individuals to have access to information (WHO, 2011), there is a campaign for the production of "easy-read" documents, which are accessible documents produced by humans, following a set of guidelines for accessible writing, such as the 'Make It Simple' guidelines (Freyhoff, 1998) or 'Guidelines for Easy-to-read Materials' (Nomura et al., 2010). The comprehensibility of the easy-read documents is also ensured by the inclusion of images to illustrate the main ideas in the text, and by the evaluation of these documents on a focus group of disabled people. While many governmental and healthcare organisations within the UK and the USA are required by law to produce accessible versions of important documents (Equality Act 2010), and many charity organisations produce such documents too, their writing and evaluation is time-consuming and expensive.

Another way to make text documents accessible for disabled readers is to convert them using automatic text simplification. Text simplification (TS) is a process which aims to enhance the understandability of a text by performing different linguistic transformations without changing its original meaning (Max, 2006). While automatic TS is promising in terms of time and financial cost, current TS systems are still not advanced enough to replace humans in the production of accessible documents. This problem is partially due to the scarcity of corpora of original and accessible texts with aligned sentences (parallel corpora) on which to train TS systems, as well as the issue of deciding which texts are simple enough for particular groups of disabled readers to be used as a gold standard for the evaluation of the TS output.

This paper describes initial research into the question of whether human-produced easy-read versions of documents could be used as a gold standard for accessible writing for particular user groups, such as readers with autism or intellectual disability. The compilation of such a corpus has now become feasible due to the already large number of existing easy-read documents produced between the early 2000s and now. Thus, for example, the original and easy-read versions of the UK political parties' manifestos from the 2015 elections (Figure 1 and Figure 2) illustrate the progress of the easy-read campaigns in adapting documents from various genres and domains:

Five years ago, Britain was on the brink. As the outgoing Labour Treasury Minister put it with brutal candour, 'there is no money'. Since then, we have turned things around.

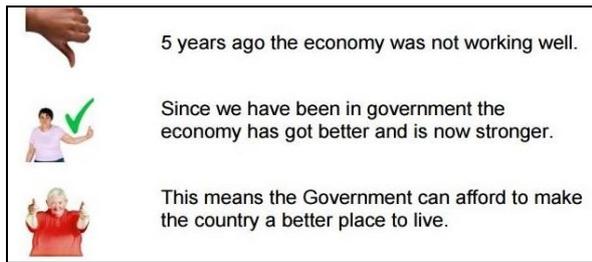Figure 1: Conservative Party Manifesto (2015)

Figure 2: Conservative Party Manifesto Easy-Read Version (2015)

However, before accepting such easy-read documents as the epitome of how an accessible text should look, we should keep in mind the variety of sources they come from, some of which may not be reputable. The current paper investigates two main aspects of easy-read documents on the Web: 1) their compliance with the guidelines according to which they were produced and 2) their suitability for particular target populations such as readers with autism or intellectual disability. In particular, we investigate the following questions:

- Do the easy-read documents available on the Web comply with the rules outlined in the guidelines for creation of easy-read documents?

- Are the easy-read documents available on the Web simple enough (or too simple?) for TS target groups such as people with mild intellectual disability or autism?

The first question is important because no official data for the evaluation of these documents has been published. We aim to assess compliance through, firstly, assigning linguistic features to the rules outlined in the guidelines for production of easy-read material and then analysing a sample of easy-read documents based on these linguistic features (Section 3). This analysis of compliance is going to cast light on the question of whether the majority of the easy-read material coming from a variety of sources on the Web could actually be regarded as such.

The second question of suitability is no less important due to the heterogeneity of conditions entailing reading difficulties. This means that a text at a certain readability level may be suitable for adults with mild intellectual disability (MID), but at the same time it could be too simplistic for adults with autism or too challenging for children with MID. This is an argument against the "one-size-fits-all" approach in creating accessible documents, where no distinction is made between the different levels of ability in cognitively

disabled people, in addition to their individual differences. In the context of this research we define suitability as the appropriate level of difficulty of texts for particular target populations. We investigate the suitability of easy-read documents for adults with autism and MID by comparing them with corpora developed for and evaluated on such readers based on 13 relevant linguistic features (Section 4).

2 Related Work

As mentioned in Section 1, automatic TS systems make use of monolingual corpora, where the text pairs could be an original article and its simplified version (parallel corpus), or two articles with different complexity levels collected based on similar criteria (e.g. topic or timespan) (comparable corpus).

2.1 Existing Corpora

For English there are several comparable corpora, which have been used in TS tasks. Simple English Wikipedia[1] together with English Wikipedia[2] comprise probably the largest resource used in automatic TS. However, the accessibility of articles in Simple Wikipedia has been disputed, with researchers appealing for "the community to drop it as the standard benchmark set for simplification" due to its many drawbacks (Xu et al., 2015). For example, Stajner et al. (2012) compare the corpus to articles from the genres of News, Health and Fiction on the basis of 4 readability formulae and 16 linguistically motivated features and find that the articles in Simple Wikipedia are more complex than the ones in the Fiction genre (Stajner et al., 2012).

Other corpora used for TS include the relatively small EncBrit (Barzilay and Elhadad, 2003), consisting of 20 articles from Encyclopedia Britannica and their manually simplified versions for children from Britannica Elementary. Due to its small size, this corpus has been used as a test set only. WeeklyReader (Allen, 2009) and Literacyworks (Peterson and Ostenforf, 2007) also have manually simplified versions for language learners for, respectively, 100 and 104 of their articles. A larger and more recent TS resource is the parallel Newsella corpus (Xu et al., 2015), which consists of 1,130 news articles, re-written for children at 4 different grade levels.

---

[1] http://simple.wikipedia.org/wiki/Main Page

[2] https://en.wikipedia.org/wiki/English_Wikipedia

Currently there are only two parallel corpora for English, which have been specifically compiled for people with disabilities. The FIRST corpus consists of 25 texts and their simplified versions for people with autism (Jordanova et al, 2013). While the simplification was performed by experts working with autistic people who followed autism-specific simplification guidelines (Martos et al., 2012), the corpus was never actually evaluated by autistic readers. The other corpus, compiled by Feng et al. (2009), is called LocalNews and is comprised of 11 newspaper articles and their simplified versions for adults with mild intellectual disability (ID). Unlike FIRST, Local-News has been evaluated by 20 adults with mild ID.

2.2 Strengths and Limitations of the Existing Corpora

A great advantage of using parallel corpora such as those mentioned above is that the original and simplified sentences are aligned, which allows automatic learning of simplification rules, hence the use of the corpora not only for evaluation but also for development of TS systems. However, currently there is no information as to whether manual simplification done with the primary objective of producing aligned corpora is of a similar quality to accessible documents produced with the reader in mind (e.g. easy-read documents). At present, this question remains an avenue for future research into the quality of resources used in TS. In addition, few of these corpora have actually been evaluated on relevant reader groups and in some cases (e.g. FIRST) the sole fact that the simplification has been done by experts is used to ensure the quality of the resource. Finally, a drawback of all corpora mentioned above is the fact that they all come from one source only (e.g. Wikipedia or Encyclopedia Britannica) and are genre-specific (encyclopedic articles, newspaper articles in the case of Local-News, newspaper and informational articles in the case of FIRST).

Compiling a corpus of easy-read documents has the potential to overcome the issues related to source and genre, because: 1) easy-read texts come from a variety of sources (the credibility of some of them being uncertain, which is why we are first going to assess the compliance of these documents collected from the Web with their production guidelines) and 2) they cover a wider variety of genres such as newspaper articles, general informational articles, healthcare, politics, literature, fun facts, etc. (Section 3). Finally, easy-read documents are widely available and do not require the time-consuming rewriting of original articles.

3 Assessing Compliance of Easy-Read Documents On the Web

As an initial step towards the development of a corpus of easy-read documents, we have first collected a sample of 150 easy-read documents in order to assess their compliance with the guidelines for their production.

3.1 Collecting a Sample of Easy-Read Documents

The sample of 150 documents consists of 78,324 words and 12,692 sentences in total. The average number of sentences per document was 84.05 with standard deviation (SD) of 98.7 and average sentence length in words 6.3 (SD = 2.17). The average number of words per document was 518.7 (SD = 624.18). When collecting this sample, we have tried to make it a balanced representation of sources and genres. The documents included in the sample were obtained from various UK and US charity organisation websites (38 documents), government departments (26 documents), healthcare services (32 documents), as well as demos of adapted books from educational websites (3 documents) and online news websites for people with disabilities (50 documents). All documents were written in English. The topics of the documents were highly dependent on their sources and thus they encompass healthcare, news, literature, politics, policies, and general information for everyday life, which is typically provided by the charity organisations (e.g. how to shop for healthy food or how to make a doctor's appointment). In Section 3.2 we analyse some characteristics of this sample, relevant to the initial guidelines for creating easy-read texts.

3.2 Linguistic Features

There are various guidelines for creating easy-read documents (Freyhoff, 1998; Nomura, Nielsen and Tronbacke, 2010), with some charity organisations creating manuals of their own. In this paper we focus on the linguistic aspect of these documents by summarising the main points of the Make It Simple guidelines (Freyhoff, 1998), as a well-known resource for producing easy-read documents, and by analysing the easy-read sample through identifying and measuring 13 features relevant to the postulates of these

guidelines. Column 1 in Table 1 lists the main recommendations of the writing guidelines, matched with corresponding linguistic features used in our analysis reflecting these recommendations (column 2). Column 3 gives the scores obtained for these features for our sample of 150 documents. The features were obtained with the Coh-Metrix 3.0 system (McNamara, 2013).

Writing Rules	Linguistic Features	Score	SD
Use short sentences	Average Sentence Length in Words	6.3	2.17
Use short words of everyday spoken language	Average Word Length in Syllables	1.44	0.12
	Word Frequency	2.43	0.2
	Age of Acquisition	317.4	35.7
	Familiarity	580.8	7.58
Use active verbs	Agentless Passive Voice Density	7.53	8.36
Use positive language	Negation Density	9.16	8.66
Use many personal words	1st Person Singular Pronoun Incidence	5.34	19.6
	2nd Person Pronoun Incidence	34.24	39.5
Avoid abstract concepts	Imagability	419.8	29.4
	Concreteness	388.9	33.4
Use simple language	Flesch Reading Ease	78.84	10.9
	Flesch-Kincaid Grade Level	3.83	1.75

Table 1: Writing rules (Freyhoff, 1998) and their corresponding linguistic features

As can be seen from Table 1, we have used the **average sentence length in words** feature as a straightforward measure of the *Length of the Sentences.* The use of *Short Words of Everyday Language* we measure through 4 indices: **word length, word frequency, age of acquisition and familiarity**, the latter two being based on norms from the MRC psycholinguistic database (Gilhooly and Logie, 1980) incorporated into the Coh-Metrix 3.0 package (McNamara, 2013). The MRC database is based on human ratings, where a word is assigned low AOA index if most people have rated it as acquired early in childhood, e.g. words such as *milk* or *pony* have a score of 202 and words such as *dogma* or *matrix* have a score of 700. The familiarity index goes into the opposite direction: a high score means that the

word is very familiar for a large part of the population sample. By comparison, the familiarity of the word *milk* has received a score of 588, while *dogma* is 328.

The use of *Active Verbs* and *Positive Language* has been measured through counting the **number of passive voice and negative constructions** respectively, so the lower the scores of these indices are, the higher the readability. Use of *Personal Words* is defined in the guidelines as: "Address your readers in a direct and personal form". To account for this aspect we have included indices such as **first person and second person pronoun incidence**. *Abstractness* is measured through **imagability** and **word concreteness** indices, which aim to identify words that evoke mental images and are thus easier to process, based on human ratings from (Gilhooly and Logie, 1980). Finally, the general *Simplicity of Language* is measured through two widely-used readability formulae: **Flesch Reading Ease** where 0=very difficult and 100=very easy (Flesch, 1948) and **Flesch-Kincaid Grade Level**, where 0 = very easy and 12=very difficult (Kincaid et al., 1975).

Other rules in the guidelines, which were not evaluated in this experiment due to lack of relevant linguistic indices, 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.*

3.3 Results

The results indicate that, indeed, the documents in the sample used fairly short sentences of 6.3 words on average, as well as short words of 1.44 syllables on average. Most of the words have also been acquired early in childhood (**AOA** = 317.4) and are highly familiar (**familiarity index** = 580.82). **Imagability** (419.78) and **concreteness** (388.87) are also high, meaning that most of the words were not abstract. Overall, we can conclude that the lexical component of the sample complies with the requirements of the guidelines. We can also observe very few uses of **passive voice** (7.53) or **negation** (9.16) and a very high **second person pronoun incidence** (34.24), showing that the reader has often been addressed directly.

The **Flesch** and **Flesch-Kincaid** formulae also demonstrate a good level of readability of the texts. The Flesch formula has an average score of

78.84 for the sample, where a score of 0 stands for "very difficult" and a score of 100 stands for "very easy"; the Flesch-Kincaid Grade Level goes in the opposite direction (the lower the score, the easier the text) and gave an average value of 3.83 for our sample.

Even though all measures indicate a very good level of accessibility of the documents, the SD measures vary greatly, which means that some of the documents score very highly in some of the measures, while others had very low scores. To investigate this further, we ranked all texts based on the scores of the Flesch Reading Ease formula. Focusing on the lower quartile, we identified 11 texts from miscellaneous sources, the Flesch readability of which was under the recommended threshold of 65 for documents written in plain English (Flesch, 1948), with some of them going as low as 43.1 or 48.77. The Flesch Reading Ease measure was consistent with the rest of the measures in identifying these 11 texts as deviant from the other 139 ones and thus we regard these as easy-read documents with lower compliance to the guidelines and thus with potentially lower accessibility.

As a whole, the results indicate that the selected sample of accessible texts complies with the standard set out in the easy-read guidelines. Only 7.33% of the texts (11 documents) showed readability under the threshold for what could be considered an accessible document. However, it is known that readability indices are an approximation only and do not account for all aspects of the text and reader interaction (DuBay, 2004). Thus, we can conclude that easy-read documents randomly selected from various domains on the Web overall comply with the rules in the easy-read guidelines.

In the next section we compare the sample of 150 documents to other corpora, which have previously been used as a gold standard in text simplification for people with disabilities.

4 Assessing the Suitability of Easy-Read Documents for Different Target Populations

Knowing that the majority of the easy-read documents available on the Web comply with the rules of easy-read production guidelines is not enough to accept them as apt for all types of readers with disabilities without evaluating their suitability: a text which is too simplistic or too challenging for a particular group of readers may cause them to lose interest in the text and may diminish their motivation. We evaluate the suitability of easy-read documents with respect to readers with autism (Section 4.1) and readers with mild ID (Section 4.2) by comparing them with corpora evaluated by these readers (LocalNews corpus in the case of ID) or developed by experts (FIRST corpus in the case of autism). If the easy-read sample is significantly more complex or simplistic than the texts in the FIRST and LocalNews corpora, its suitability as a gold standard for accessible texts for readers with autism might be disputed based on the level of simplification the users (LocalNews) and the experts (FIRST) have perceived as suitable for the relevant populations.

4.1 Autism

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised with impairment in social interaction and communication, which influences the language comprehension abilities of the affected individuals (APA, 2013). In previous research we evaluated the suitability of 7 randomly selected easy-read documents from the same 150 document sample used in this study on 20 adults with autism (without intellectual disability) and 20 non-autistic adults matched for age and level of education (Yaneva, Temnikova and Mitkov, 2015). The level of comprehension was assessed through: 1) multiple choice questions, 2) reading times and 3) self-reported difficulty measures for each document. The results indicate that all documents were well understood by all participants, with the autistic participants requiring significantly more time to read them compared with the non-autistic ones. In addition, the autistic participants rank the texts from predominantly "very easy" and "easy" to "moderate" and in a few cases "difficult", while the vast majority of the non-autistic participants rank them as "very easy". The study concluded that easy-read documents are understandable enough for adult readers with autism (without intellectual disability), while their perceived level of difficulty is not so trivial as to bore the readers.

As mentioned in Section 2, the FIRST corpus consists of 25 original texts and their parallel simplified versions from the genres of news, education and popular culture. It has been produced by experts specifically working with autistic adults but has never been evaluated by its target population. Our easy-read sample, on the other hand, has been partially evaluated with participants (7 documents only (Yaneva et al., 2015))

	Sent length	Word length	Freq	AOA	Famil-iarity	Concr	Imaga bility	Pas-sive	Nega-tion	1st pers.	2nd pers.	Flesch	Flesch-Kinc
FIRST vs.ER	**-4.4**	-0.1	-1.4	-1.5	**-3**	-0.01	-0.01	-1.8	-0.2	-0.6	**-3.6**	-1.7	**-3.5**
LN vs. ER	**-2.9**	-0.5	**-2.7**	-0.5	-1.7	-1.3	-1.3	-1.8	-0.2	-1.6	-1	-0.6	-1.8

Table 2: Wilcoxon test Z scores for the FIRST and LocalNews (LN) corpora compared to the Easy-Read (ER) sample

and has been produced for people with disabilities as a primary purpose. The comparison of the two corpora is based on the same 13 features as described in Section 3.2.

A Shapiro-Wilk test showed that the data was non-normally distributed, so a Wilcoxon paired signed rank test was applied to compare the differences between the two corpora. Table 2 shows the results of the Wilcoxon test, where the z scores marked in bold indicate 0.001 level of significance.

The main difference between the FIRST corpus and the easy-read texts is that the sentences in FIRST are significantly longer than the ones in the easy-read sample. This difference in sentence length is also the reason why the Flesch Reading Ease formula does not find a significant difference between the levels of difficulty of the two corpora, while Flesch-Kincaid distinguishes between their levels of difficulty, due to its subtler sensitivity to sentence length. The lexical component in both corpora is equally simple, except the fact that the words in the easy-read documents have a higher familiarity level. Finally, the FIRST corpus does not contain many instances where the readers are addressed by second person pronouns, but this could be attributed to the lack of instructional texts in the FIRST corpus compared with the easy-read sample.

4.2 Mild Intellectual Disability

Intellectual Disability (ID) is a condition involving impairment in the general mental abilities of the affected individuals (APA, 2013). The LocalNews parallel corpus (Feng et al., 2009) contains 11 newspaper articles simplified by experts working with adults with mild intellectual disability (MID). Unlike FIRST, LocalNews has been evaluated by 20 adults with MID (Feng et al., 2009). In order to avoid genre bias we only compare the LocalNews corpus against 50 easy-read newspaper articles from our sample. A Shapiro-Wilk test identified the data as non-normally distributed, which is why, similar to the experiment with FIRST, a Wilcoxon signed-rank

pair test was applied. The z scores for all 13 features are summarised in Table 2.

Similar to the results from the comparison with the FIRST corpus, again the average sentence length for each document from the easy-read sample is shorter than the average sentence length in the LocalNews corpus, though not to the extent to influence the Flesch-Kincaid formula, which in this case did not differentiate significantly between the two groups of texts. The only other difference is the fact that the words in the easy-read sample had higher frequency scores than those in the LocalNews corpus.

5 Conclusions and Future Work

The results of the presented studies showed that easy-read documents, which were randomly accessed from various domains on the Web, such as charity organisations and government or healthcare websites, comply with the accessibility standard set in the easy-read guidelines. Second, these texts did not exceed the level of difficulty of corpora previously used as a gold standard for accessible writing for autism and mild intellectual disability. Quite the opposite, a presence of shorter sentences and more familiar words was shown, but these did not influence the indices to an extent that would put the easy-read documents in a whole new class of documents, which might be deemed as too simplistic. By satisfying the prerequisites of having good compliance and suitability for autism and MID, easy-read documents show the potential of being a valid gold standard for accessible documents.

Future challenges include exploring the possibility of creating a monolingual comparable corpus of easy-read documents and documents developed for the general audience (e.g. the Conservative Party manifesto versions in Section 1). The creation of such a corpus would allow investigation of ways of aligning parts of these documents where possible, for the purposes of improving automatic text simplification for people with disabilities.

References

David Allen, 2009. A study of the role of relative clauses in the simplification of news texts for learners of English. *System*, 37(4):585–599

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association.

Regina Barzilay and Noemie Elhadad. 2003. Sentence alignment for monolingual comparable corpora. In *Proceedings of the Conference on Empirical methods in Natural Language Processing (EMNLP)*, pages 25–32. ACL.

Conservative Party Manifesto Easy-Read Version. 2015[online] Available at: https://www.mencap.org.uk/sites/default/files/docu ments/Easy%20Read%20Manifesto_Conservative_ V2.pdf. Last accessed: 23.06.2015

Conservative Party Manifesto.2015[online] Available at: https://www.conservatives.com/Manifesto. Last accessed: 23.06.2015

William Coster and David Kauchak. 2011. Simple EnglishWikipedia: A New Text Simplification Task. Proceedings of the *49th Annual Meeting of the Association for Computational Linguistics*: shortpapers, pages 665–669, Portland, Oregon, June 19-24, 2011

William DuBay. 2004. The principles of readability. Costa Mesa, CA: Impact Information.

Equality Act 2010. UK. [online] Avaliable at: http://www.legislation.gov.uk/ukpga/2010/15/secti on/6. [Last accessed 26.06.2015]

Lijun Feng. 2009. Automatic readability assessment for people with intellectual disabilities. *ACM SIGACCESS accessibility and computing*, 93, pp. 84-91

Rudolf Flesch. 1948. A New Readability Yardstick. *Journal of Applied Psychology*, 32(3), 221-233.

Geert Freyhoff, Gerhard Hess, Linda Kerr, Bror Tronbacke and Kathy Van Der Veken (1998) Make it Simple. European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability. ILSMH European Association

Ken J. Gilhooly and Robert H. Logie. 1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*. Volume 12 (4), pp: 395-427

Vesna Jordanova, Richard Evans and Arlinda Cerga-Pashoja. 2013. FIRST Deliverable – Benchmark report (result of piloting task). Technical Report D7.2, Central and Northwest London NHS FoundationTrust, London, UK.

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. CNTECHTRA Research Branch Report 8-75.

Juan Martos, Sandra Freire, Ana Gonz´alez, David Gil, Richard Evans, Vesna Jordanova, Arlinda Cerga, Antoneta Shishkova and Constantin Orasan. 2013. FIRST Deliverable - User preferences: Updated. Technical Report D2.2, Deletrea, Madrid, Spain.

Aurelien Max. 2006. Writing for Language-impaired Readers. CICLing'06 Proceedings of the *7th international conference on Computational Linguistics and Intelligent Text Processing.* pp. 567 – 570.

Danielle S. McNamara, Max M. Louwerse, Z. Cai and Art Graesser. 2013. Coh-Metrix version 3.0. Retrieved [17.05.2015], from http://cohmetrix.com

Misako Nomura, Gyda Skat Nielsen, Bror Tronbacke and International Federation of Library Associations and Institutions (2010) Guidelines for Easy-to-read Materials/rev. IFLA Headquarters, The Hague.

Sarah E. Petersen, and Mari Ostendorf. 2007. Text simplification for language learners: a corpus analysis. In Proc. of *Workshop on Speech and Language Technology for Education*, pages 69–72

Sanja Štajner, Richard Evans, Constantin Orasan and Ruslan Mitkov. 2012. What can readability measures really tell us about text complexity? In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, held in conjunction with LREC 2012. Istanbul, Turkey, May 27, pp. 14-21

World Health Organisation. 2011. World Report on Disability.[online]Available at: http://www.who.int [Last accessed 26.06.2015]

Wei Xu, Chris Callison-Burch and Courtney Napoles. 2015. Problems in Current Text Simplification Research: New Data Can Help. *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 283–297.

Victoria Yaneva, Irina Temnikova and Ruslan Mitkov. 2015. Accessible Texts for Autism: An Eye-Tracking Study. To appear in *ASSETS 2015. The 17th International ACM SIGACCESS Conference of Computers and Accessibility*, Lisbon, Portugal, 26-28 October.