Six Good Predictors of Autistic Text Comprehension

Victoria Yaneva
Research Group in Computational Linguistics
University of Wolverhampton
V.Yaneva@wlv.ac.uk

Richard Evans
Research Group in Computational Linguistics
University of Wolverhampton
R.J.Evans@wlv.ac.uk

Abstract

This paper presents our investigation of the ability of 33 readability indices to account for the reading comprehension difficulty posed by texts for people with autism. The evaluation by autistic readers of 16 text passages is described, a process which led to the production of the first text collection for which readability has been evaluated by people with autism. We present the findings of a study to determine which of the 33 indices can successfully discriminate between the difficulty levels of the text passages, as determined by our reading experiment involving autistic participants. The discriminatory power of the indices is further assessed through their application to the FIRST corpus which consists of 25 texts presented in their original form and in a manually simplified form (50 texts in total), produced specifically for readers with autism.

1 Introduction

Autism Spectrum Disorder (ASD) is a developmental disorder of neural origin, characterised by impairment in communication and social interaction and stereotyped repetitive behaviour (American Psychiatric Association, 2013). Currently about 1 in 100 people in the UK are diagnosed with this condition (Brugha et al., 2012), and there are assumed to be two undiagnosed cases for every three diagnosed (Baron-Cohen et al., 2009). In many countries there are no official statistics about the number of affected individuals, but with rising awareness of the condition, this number has been continually increasing to the extent that it is now referred to as an autism epidemic (Wazana, 2007).

One of the central characteristics of autism is impairment in communication both in terms of language comprehension and social interaction. Depending on the severity with which the condition affects individuals, they may be low-functioning and often non-verbal or medium and high-functioning, requiring help with only the social aspects of language use. While most medium- and high-functioning autistic people have a high level of word decoding skills when reading, they struggle at semantic, syntactic and most of all, pragmatic levels of understanding. For example it may be challenging for autistic readers to access the meaning of some words if they are very abstract or are too long; they may have difficulty in processing long and complex sentences due to the cognitive load that these impose on the reader and the comparatively short working memory span that people with autism may have (Bennetto et al., 1996). However, the area of utmost difficulty for autistic individuals, which differentiates them from non-autistic readers in the way that they read, is their inability to "refer to the whole", to struggle to infer meaning from both the semantic and the social context of a text (Frith and Snowling, 1983; Happé, 1997). These characteristics of their reading can be illustrated by the ability of autistic readers to use syntactic context but not semantic context to disambiguate homophones (Happé, 1997) and by their reduced ability to understand non-literal language, sarcasm, irony and authors' intentions (O'Connor and Klein, 2004; MacKay and Shaw, 2004).

There are a number of software tools designed to assist people with autism in their use of language, including automatic text simplification tools (Section 2.1). The emergence of such software entails a need, at the very least, to assess the accessibility of instruction manuals provided for users with autism. In the case of text simplification software, it is necessary to assess (1) the extent to which texts require conversion to a more accessible form, (2) the types of conversion oper-

Proceedings of Recent Advances in Natural Language Processing, pages 697–706,
Hissar, Bulgaria, Sep 7–9 2015.

ations that are required, and (3) the suitability of the converted output for readers with autism. It is expected that people working to improve the accessibility of a given text, both in automatic and manual text conversion, will benefit from relevant feedback concerning the effects of different conversion operations and the extent to which different versions of a text meet the particular requirements of intended readers. So far the only way to perform such evaluation has been to conduct time-consuming and expensive user-focused evaluation studies. Automatic methods to assess the readability of texts for people with autism have proven useful in the process of automatic text simplification but these have not been applied to user-evaluated texts and thus their merit is unknown. In this paper, the term *user-evaluated texts* is used to denote texts whose readability has been evaluated via reading comprehension testing and recording of the reading times of people with autism. So far, their scarcity has meant that user-evaluated texts have not been exploited in the development of language technology intended to provide reading support. (Section 2.2). There has also been no user-focused research on readability in autism.

The research described in the current paper includes:

- the production of reading passages at different readability levels evaluated by 20 participants with autism with no developmental delay.
- evaluation of the effectiveness of 33 automatically computed readability indices to discriminate between texts classified by the users as easy or difficult. Some of these indices have been used in the past to account for reading difficulties in autism but this is the first time that their effectiveness has been tested on text passages evaluated by users.
- evaluation of the indices on the FIRST corpus which consists of 25 texts presented in their orginal form and in a more accessible form, converted by experts working with people with autism and following ASD-specific text simplification guidelines (Jordanova et al., 2013).

These are contributions toward a better understanding of text readability from the perspective of people with autism.

2 Related Work

2.1 Assistive Language Technology for People with Autism

Assistive software and technologies have repeatedly been reported to be well-received among autistic individuals for various reasons, including their demand for structure and uniformity, the ability of automatic tools to repeat the same action or instruction multiple times and the ability of these tools to reduce the complexity of social situations (Bosseler and Massaro, 2003; Putnam and Chong, 2008). As the need of autistic individuals for assistance with language-related tasks is well-known, a number of software tools have been developed to assist the language development of autistic individuals of various age groups and at various levels of ability.

A suitable tool for people with ASD who are severely impaired and who may remain completely or partially non-verbal, are the various types of picture exchange communication systems (PECS), which allow them to produce sentences by combining images and words on a tablet screen or PDA (Charlop-Christy et al., 2002). For those who are not so severely impaired as to remain non-verbal but are still in the process of acquiring verbal skills, the *VAST-Autism* app[1] combines videos with written words and auditory cues to help autistic and apraxic individuals acquire certain words, phrases or sentences. *Stories About Me*,[2] is another iPad application, which helps autistic users produce stories by combining photographs with text and voice recordings.

For autistic individuals who are fairly able, the *OpenBook* tool[3] provides semi-automatic conversion of text documents by reducing syntactic complexity and disambiguating meaning by resolving pronominal reference, performing word sense disambiguation and detecting conventional metaphors. The output is an accessible version of the original document supplemented with additional elements such as glossaries, illustrative images, and document summaries. The system is deployed as an editing tool for healthcare and educational service providers.

Systems such as *OpenBook* can benefit from ad-

[1] https://itunes.apple.com/us/app/vast-autism-1-core/id426041133?mt=8, last accessed May 2015.

[2] https://itunes.apple.com/us/app/stories-about-me/id531603747?mt=8, last accessed May 2015.

[3] http://openbooktool.net, last accessed May 2015.

vances in autism-specific automatic readability assessment, as this process can be used to evaluate each conversion operation applied.

2.2 Readability Assessment

Readability assessment has been used to match intended readers to texts with a view to the specific purpose of reading (Chall and Dale, 1995). Classic readability formulae typically exploit textual features such as sentence length, word length, and the average number of syllables per word, or make use of word lists such as Dale and Chall's list of 3 000 EasyWords (Dale and Chall, 1948). Dubay (2004) provides information on a large number of readability formulae. More sophisticated systems, such as the *Coh-Metrix* system (Graesser et al., 2004) and the Lexile Framework (Smith et al., 1989), are based on surface features, cognitively-motivated features and features of cohesion and syntactic complexity, exploiting human-evaluated databases such as the Colorado Norms for word familiarity, and age of acquisition and concreteness indices, among others (Smith et al., 1989; McNamara et al., 2010).

Readability formulae are developed with particular target populations and text types in mind (Siddharthan, 2004; Benjamin, 2012; Bruce et al., 1981), which is why readability features relevant specifically to people with special needs have also been explored. For example, people with intellectual disability have been found to have decreased working memory capacity (a characteristic they share with some people with autism), which results in their difficulty in remembering relations within and between sentences (Jansche et al., 2010). Thus, features developed for and evaluated on this reader population include entity density (counts of entities such as person, location and organisation per sentence) and lexical chains (synonymy or hyponymy relations between nouns) (Jansche et al., 2010; Feng et al., 2010; Huenerfauth et al., 2009). Word frequency and word length have been found to affect readability for Spanish readers with dyslexia based on data from eye tracking techniques and comprehension questions (Rello et al., 2012a; Rello et al., 2012b).

Previous assessments of the readabiliy of texts to be read by people with autism have explored features hypothesised to be related to those aspects of language known to pose reading comprehension difficulties for this population (Martos et al., 2013; Štajner et al., 2012; Štajner et al., 2014).[4] In previous research, a set consisting of three groups of readability indices, used to estimate syntactic complexity and ambiguity in meaning, together with several other exisiting readability formulae were used to assess the readability of texts of the registers of news, health, and fiction. The scores obtained were compared with those obtained when estimating the readability of texts from Simple Wikipedia, which were assumed to be a gold standard of readability. This assumption is disputed (Štajner et al., 2012) but at the time of their experiments, no user-evaluated text resources were available. Readability indices such as the number of metaphors or passive verb constructions per text have been considered (Jordanova et al., 2013) but their discriminative power has not previously been evaluated on texts whose difficulty for autistic readers is known. The research presented in this paper builds upon these previous studies by evaluating text passages with respect to 20 participants with autism and testing the effectiveness of various readability indices, including those developed by Jordanova et al. (2013), to discriminate between the levels of difficulty of the passages.

3 Production of User-Evaluated Text Passages

This section presents the experimental design and procedure for evaluating the difficulty of 16 text passages by 20 participants diagnosed with autism spectrum disorder.

3.1 Design and Materials

The participants were asked to read text passages and answer three multiple choice questions (MCQs) per passage. Evaluation of the difficulty of the texts is then based on their answers to the questions and their reading time scores, produced by dividing the amount of time a participant spends reading the text (measured in seconds) by the number of words in the text to control for the differences in length between the texts.

3.1.1 Text Passages

To avoid bias, the study included a total of 16 text passages from miscellaneous domains and registers (3 newspaper articles, 3 educational articles, 3 general informational texts obtained from the web,

[4] Hypotheses that were not formally tested.

and 7 easy-read documents, which are simple documents developed specifically for people with disabilities) (Table 1).The presented texts vary in difficulty and avoid potentially sensitive topics such as religion, sexuality, and disabilities.

One of the biggest challenges in the design of this study and the selection of materials was the fact that people with autism are prone to experience difficulties with concentration and attention, resulting in fatigue (Happé and Frith, 2006; Lai et al., 2014). For this reason, the evaluation by a single participant of a large set of long text passages is not feasible. The length of each text and the number of texts presented to each participant were selected with a view to avoid fatigue and to comply with ethical considerations. Table 1 summarises some of the characteristics of the texts included in this study.

Text Number	Register	#Words	Flesch-Kincaid Grade Level[5]	Flesh Reading Ease Score[6]
1	Informational	163	4.93	79.548
2	Educational	178	4.671	80.22
3	Educational	206	7.577	65.437
4	Educational	189	9.276	56.758
5	Newspaper	226	11.983	40.658
6	Newspaper	160	8.866	59.82
7	Informational	163	8.765	66.657
8	Informational	185	14.678	45.34
9	Newspaper	188	9.823	58.298
10	Easy-Read	77	8.16	60.11
11	Easy-Read	96	6.73	67.33
12	Easy-Read	74	2.71	92.54
13	Easy-Read	178	5.52	75.33
14	Easy-Read	77	5.79	70.67
15	Easy-Read	121	1.75	95
16	Easy-Read	58	6.63	68.16

Table 1: Characteristics of the 16 texts included in the study.

3.1.2 Questions

Three multiple choice questions (MCQs) with four possible answers were developed for each text. Three different types of MCQs were presented to assess different types of reading comprehension:

1. Literal MCQs, examining literal understanding of the texts;
2. Reorganisation MCQs, examining the ability of participants to combine information from different parts of the text. One characteristic of autistic readers is that they make little use of context (Oliver, 1998; O'Connor and Klein, 2004), which is crucial for performing the task of reorganisation;
3. Gap Inference MCQs, examining participants' abilities to use two or more pieces of information from a text in order to arrive at a third piece of information that is implicit (Kispal, 2008). Since this type of question is based on literal understanding, they evaluate the role of context and structure of the text. Inferences involve both literal understanding and general knowledge, intuition, and pragmatic understanding of the text (Day and Park, 2005), which is a central area of impairment in ASD.

In the case of the easy-read documents, only literal questions were presented due to the simplicity of the information contained in the text. All MCQs developed for the 16 texts used simple language with highly frequent words combined in sentences containing a maximum of three clauses.

3.2 Participants

Participants in the study were 20 adults (7 female, 13 male) with a confirmed diagnosis of autism recruited through 4 local charity organisations. None of the 20 participants had comorbid conditions affecting reading (e.g. dyslexia, learning difficulties, aphasia etc.). Mean age (m) for the group in years was m=30.75, with standard deviation SD=8.23, while years spent in education, as a factor influencing reading skills, were m=15.31, with SD=2.9. None of the participants had been diagnosed with a learning disability or developmental delay. All participants were native speakers of English.

3.3 Apparatus and Procedure

The texts were displayed on a 19" LCD monitor via software specifically designed following analysis of the requirements of people with ASD (Martos et al., 2013): there were no bright colours, complex navigation systems or distracting logos or images. Reading time was measured in seconds using the software, which also randomised both the order of presentation of the texts and the

[5] *Flesch-Kincaid Grade Level* is inversely proportional to text readability. For text passages of less than 100 words, the Flesch-Kincaid Grade Level and the Flesch Score have been computed for whole documents rather than selected text snippets, due to the fact that these formulae are not recommended for texts shorter than 100 words (Dubay, 2004).

[6] *Flesch Reading Ease Score* is proportional to text readability.

questions pertaining to texts for each participant, to avoid bias. Each session lasted between 40 and 70 minutes. Informed consent was first obtained and demographic information about diagnoses, age and level of education collected. Participants then read all texts and answered all questions, taking as many breaks as they requested. At the end of the experiment, participants were debriefed.

3.4 Results from the Reading Comprehension Experiment

A Shapiro-Wilk test showed that the answers to reading comprehension questions based on the texts are non-normally distributed. A Friedman test was performed, confirming that there are significant differences between scores obtained for different texts ($\chi^2(12) = 39.698$, $p < 0.001$). A post-hoc Wilcoxon Signed Rank test with Bonferroni adjustment of the significance level identified the differences between the particular texts and on this basis, they were divided into two groups: *easy* and *difficult*. All easy-read texts (10 to 16) and texts 1 to 4 were classified as *easy*, with only text 1 being significantly easier than the other texts in this group (text 2 and text 1: $p = 0.008$). Texts 5 to 9 varied in difficulty but were classified as significantly more difficult than the first 4 texts and the easy-read texts. Therefore they were assigned to a separate class: *difficult* (text 5 and text 4: $p = 0.012$; text 6 and text 5: $p = 0.083$; text 7 and text 6: $p = 0.034$; text 8 and text 7: $p = 0.037$; text 9 and text 8: $p = 0.021$).

These differences in the level of difficulty of the texts were also confirmed by the reading time score. The data from the reading time scores was also non-normally distributed according to the Shapiro-Wilk test and a Friedman test identified significant differences between the 9 texts ($\chi^2(12) = 45.060$, $p < 0.001$). A post-hoc Wilcoxon Signed Rank test with Bonferroni adjustment confirmed that texts 5 to 9 were to be considered more difficult than texts 1 to 4 (text 5 and text 4: $p = 0.001$), with no significant differences in the reading time scores between texts 5 and 6 (text 6 and text 5: $p = 0.409$; text 7 and text 6: $p = 0.683$; text 8 and text 7: $p = 0.331$; text 9 and text 8: $p = 0.601$).

Both measures, *question answers* and *reading time scores*, classified texts 1 to 4 and texts 10 to 16 as *easy*, while texts 5 to 9 were significantly more complex and were thus classified as *difficult*. The next section describes the readability indices applied to these two classes of text in order to find the most suitable indices for predicting reading difficulty for people with autism.

4 Readability Indices

Four groups of readability metrics, comprising 33 indices overall, were selected on the basis of their relationship to the types of difficulties that readers with autism face. All of the selected metrics were automatically computed with the exception of the *metaphor index*, which required manual counting of metaphors, due to the scarcity of accurate systems for automatic figurative language detection. The sets of syntactic and cognitively-motivated lexical features were computed using the *Coh_Metrix 3.0* tool (McNamara et al., 2010), which exploits the Charniak parser (Charniak, 2000).

4.1 Indices Previously Used to Assess Text Difficulty for Readers with ASD

The indices described in this section were proposed during the development of the OpenBook tool and are based on a user study identifying 43 user requirements (Jordanova et al., 2013; Martos et al., 2013). Indices (1), (2), (3) and (7) relate to features of syntactic and lexical complexity, while (4), (5) and (8) are intended to measure ambiguity in meaning. Index (6), the only index whose evaluation requires human input, estimates the difficulty posed by texts to autistic readers due to their difficulties in understanding metaphor and figurative language.

Definitions:

(1) **Comma index (C)** is proportional to the ratio of commas to words in the text. It indicates the average syntactic complexity of the sentences occurring in the text.

(2) **Index of words with three or more syllables (MSW)** is proprotional to the ratio of the number of words in the text with three or more syllables to the number of words in the text.

(3) **Index of words per sentence (WPS)** is the ratio of words to sentences in the text.

(4) The **Index of word diversity (WD)** is the type-token ratio of the text. The greater the number of word types occurring, the greater the likelihood that one or more of them will be semantically ambiguous.

(5) **Pronoun index (P)** is proportional to the ratio of the number of pronouns in the text to the number of words in the text. This index is relevant to the difficulty some autistic readers have in resolving anaphors (Martos et al., 2013).
(6) **Metaphor index (M)** is proportional to the ratio of the number of phraseological units and non-lexicalised metaphors in the text to the number of sentences in the text.
(7) **Passive verb index (PV)** is proportional to the ratio of passive verbs in the text to the number of sentences in the text. LT was developed to detect the occurrence of passive verbs in English on the basis of part-of-speech patterns, exploiting the LT TTT package (Grover et al.,).
(8) **Polysemic word index (PW)** is proportional to the ratio of the number of words in the text that belong to more that one synset in a language-specific ontology to the number of words in the text.

4.2 Syntactic Complexity Features

Syntactic complexity features account for the difficulties readers with ASD may experience in processing long and complex sentences. For example, the metric *Words Before Main Verb* estimates the working memory load imposed by a sentence (McNamara et al., 2010), and is particularly relevant to autism, as some autistic individuals have been shown to have decreased working memory capacity (Bennetto et al., 1996).

The set of 12 syntactic complexity features includes *Words before Main Verb* (the mean number of words occurring before the main verb of the main clause in each sentence), *Mean Number of Modifiers per Noun-Phrase*; *Syntactic Structure Similarity (Adjacent)* (proportional to the number of nodes in syntactic sub-trees shared by adjacent sentences, averaged over all pairs of adjacent sentences), *Syntactic Structure Similarity (All)* (computed in a similar way, but between all pairs of sentences in the text, not just adjacent ones), and *incidence scores* of *nouns*, *verbs*, *adverbial and prepositional phrases*, *passive voice forms*, *negation expressions*, *gerunds* and *infinitives*.

4.3 Cognitively Motivated Lexical Features

Cognitively-motivated readability features evaluate various phenomena relevant to autistic readers such as references to highly abstract concepts, which some readers with ASD may be unable to understand, and unfamiliar words that may pose difficulties because some readers are unable to exploit context to comprehend them. A set of 5 cognitively-motivated indices, based on word norms from the MRC psycholinguistic database (Gilhooly and Logie, 1980) and obtained using the *Coh_Metrix 3.0* system, were included in the study: *Frequency of Words*, *Age of Acquisition*, *Familiarity*, *Concreteness*, and *Imagability*.

4.4 Readability Formulae

Readability formulae included in the study were *Flesch Reading Ease* (Flesch, 1948), *Flesch-Kincaid Grade Level* (Kincaid et al., 1975; Kincaid et al., 1981), *Army's Readability Index* (*ARI*) (Senter and Smith, 1967), *Fog Index* (Gunning, 1952), *Lix* (Björnsson, 1968); and *SMOG* (McLaughlin, 1969).

5 Data Analysis and Results

A Shapiro-Wilk test showed that some of the datasets were normally distributed, while others were not. A paired samples t-test with corrections for outliers and a Wilcoxon signed rank test were both applied, showing consistent results.[7]

A paired samples t-test was used to evaluate whether each of the readability indices described in Section 4 could discriminate significantly between the two classes of easy and difficult texts. After that, a bootstrap for the paired samples test was used to calculate 95% confidence intervals (CI) based on $1\,000$ bootstrap samples of each measure. Table 2 presents values of p, t-test results, and the 95% CI endpoints of each of the three discriminative sets of readability features. Of the set of readability indices developed to evaluate texts for readers with ASD, statistical analysis indicates that a two-tailed significant difference was yielded by two indices: *words in sentences* and *metaphor index*. Of the syntactic set, significant results were yielded by the *Words Before Main Verb* measure of cognitive load and the *Syntactic Structure Similarity (Adjacent)* measure. *Syntactic Structure Similarity (All)* did show significance at the t-test ($t = 2.932$, $p < 0.05$ with 95% CI ($0.01\,800$, $0.08\,540$)) but the p value after bootstrapping increased to $p = 0.086$, indicating that it is not a significant discriminator. The third set of cognitively motivated features failed to discriminate between the two classes, while the only readability formula of the fourth set which

[7] For brevity, only the t-test results are reported in this paper.

Index	t	p	*95% CI Endpoints*	
			Lower	*Upper*
ASD-Specific				
Words in Sentences	-6.514	< 0.05	$-8.75\,421$	$-511\,480$
Metaphor Index	-3.723	< 0.05	$-0.66\,997$	$-0.26\,537$
Syntactic				
Words Before Main Verb	-3.264	< 0.05	$-3.21\,221$	$-1.05\,580$
Syntactic Structure Similarity (Adjacent)	3.510	< 0.05	$0.03\,080$	$0.09\,540$
Readability				
Flesch-Kincaid Reading Ease	-3.362	0.028	$-7.02\,138$	$-0.66\,982$
ARI	-3.706	< 0.05	$-5.46\,000$	$-2.12\,000$

Table 2: Six features discriminative between *easy* and *difficult* texts.

managed to do so was *Flesch-Kincaid Reading Ease*. The t-test indicated significance of the *Lix* measure in discriminating between *easy* and *difficult* texts ($t = -2.824$, $p < 0.05$, with 95% CI ($-16.5\,800$, $-3.78\,000$)), but bootstrapping contradicted this ($p = 0.090$).

6 Application to Manual Text Simplification Evaluation

6.1 Materials

The effectiveness of the readability indices described in Section 4 was assessed over a larger set of texts specifically designed for people with autism. They were applied to the FIRST corpus, which consists of 25 documents of the registers of popular science and literature (13 texts) and newspaper articles (12 texts) (Jordanova et al., 2013). These texts were presented in both their original form and in a form intended to facilitate reading comprehension, so that the corpus contains 25 paired original and simplified documents (50 documents in total). The simplification was performed by 5 experts working with autistic people, who were given ASD-specific text simplification guidelines, specified by Jordanova et al. (2013), which contains full details of the simplification procedure and the characteristics of the corpus. It is important to note that no user-based evaluation of those texts has been conducted. Evaluating the readability indices on the FIRST corpus would test their efficacy in discriminating between original and manually simplified versions of texts.

6.2 Results

All readability indices that successfully discriminated between *easy* and *difficult* user-evaluated texts and all 7 readability formulae discriminated successfully between the original and simplified versions of texts with $p < 0.0001$. Other indices from the first set of ASD-specific features that performed well were the *Comma Index*, *Syllables in Long Words*, *Word Diversity*, and *Pronoun Index*. Successful discriminators from the cognitive set were the features *Average Word Length in Syllables*, *word frequency*, *Age of Acquisition*, *Familiarity*, and *Polysemy*. Finally, of the syntactic set, *Mean Number of Modifiers per NP*, *incidence score of preposition phrases*, and *gerunds* were significant discriminators. Table 3 displays p-values and t-test results of each of these features.

Index	t	p
ASD-Specific		
Comma index	-8.077	0.0001
Syllables in long words	-3.006	0.0001
Word Diversity	-5.840	0.0001
Pronoun Index	4.211	0.0001
Cognitive		
Average word length (syllables)	-2.500	0.016
Word frequency	4.727	0.0001
Age of Acquisition	-3.438	0.002
Familiarity	4.426	0.001
Polysemy	3.048	0.006
Syntactic		
Mean number of modifiers per NP	-3.934	0.001
Incidence Score of Prepositional Phrases	-2.446	0.022
Incidence Score of Gerunds	-3.544	0.002

Table 3: Features discriminative between original and manually simplified versions of texts.

7 Discussion

The study shows that the main differences between the *easy* and *difficult* texts evaluated by autistic users were that, unsurprisingly, the easy texts contain shorter words and sentences. However, an even more marked characteristic of the easier texts is the fact that they contain fewer metaphors. The *metaphor index* was far more predictive than commonly used readability features such as modifiers per noun phrase, type-token ratio, or instances of

passive voice. This feature is directly related to the inability of even some of the highly skilled readers with autism to comprehend figurative constructions. One limitation is that the *metaphor index* needs to be derived manually and that manual annotation of metaphors can be an onerous and unreliable process. However, we argue that, in the case of readability assessment for autism, a very detailed annotation scheme encoding fine-grained distinctions is unnecessary and that a less detailed approach would be sufficient. In due course, advances in NLP may make the automatic tagging of metaphors a feasible option.

One feature, whose use is relatively uncommon in the metrics used to assess readability for other populations, is the occurrence of fewer words before the main verb in a sentence, which has proven effective due to the decreased working memory capacity of people with autism. That is, the closer that main verbs are to the starts of sentences, the more comprehensible the text is for readers with autism. Consistency of syntactic structure was also found to be a highly-discriminative measure, meaning that sentences in *easy* texts have greater uniformity of syntactic structure. Furthermore, the results indicate that it is more important for autistic readers that syntactic structure is similar in adjacent sentences rather than over whole documents, as the latter index was insignificant after bootstrapping. One possible explanation for the significance of this index is that the syntactic structure of texts of the register of news is quite diverse, possibly due to the variety of sources, including reported speech and reportage, included in news articles. It would be interesting to investigate whether this index is as discriminative when applied only to educational texts. Finally, *Flesch-Kincaid Grade Level* and *ARI* were found to be suitable formulae for assessing the readability of texts for autistic readers. This may be due to the sensitivity of autistic readers to sentence length, a feature which is weighted more heavily in the *Flesch-Kincaid* formula than in others, such as the original *Flesch* formula (Dubay, 2004). The occurrence of passive verbs and the frequent use of pronouns, which were previously thought to increase reading difficulty for people with autism, did not prove to be significant in our experiments.

All indices which successfully discriminated between the user-evaluated texts retained their significance when applied to the FIRST corpus with $p < 0.0001$, showing that they are suitable for use in text simplification tasks. Due to the considerable number of simplification operations applied in the FIRST corpus, which resulted in larger differences between the two classes of texts than between texts included in the user-evaluated materials, many other indices were also discriminative.

8 Conclusions and Future Work

The study identified six readability indices as being highly-discriminative of text complexity for readers with autism: the *number of words per sentence*, the *number of metaphors per text*, the *average number of words occurring before the main verb in a sentence*, *syntactic structure similarity for adjacent sentences*, *Flesch-Kincaid Grade Level*, and the *Automated Readability Index*. These indices discriminated successfully both between texts evaluated as *easy* or *difficult* by reference to comprehension testing and reading times of participants with ASD and between texts in the FIRST corpus in original and simplified forms.

An additional set of autism-specific, syntactic and cognitively-based readability indices and readability formulae discriminated successfully between original and simplified texts of the FIRST corpus, but this is most likely explained by the considerable number of simplification operations applied to it. On the assumption that this corpus of simplified texts is more accessible for readers with autism, this extended set of indices could be considered suitable for this target population.

This study shares the limitations of all research involving participants with autism: small sample sizes and strict limits on the demands that can be placed on participants, due to their condition. The results should therefore be applied with caution and not necessarily generalised to children, people at the lower ends of the autism spectrum, or people with other types of disabilities. Future work would include evaluation of a larger set of texts by a larger group of particpants and the exploration of new readability indices tailored to the specific reading difficulties of autistic individuals.

9 Acknowledgments

The authors gratefully acknowledge the volunteers who took part in this study, the charity organisations which facilitated their recruitment, and the three anonymous reviewers for their valuable comments.

References

American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (5th ed.).

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

Loisa Bennetto, Bruce F. Pennington, and Sally J. Rogers. 1996. Intact and Impaired Memory Functions in Autism. *Child Development*, 67(4):1816–1835.

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

Alexis Bosseler and Dominic W. Massaro. 2003. Development and evaluation of computer-animated tutor for vocabulary and language learning in children with autism. *Journal of autism and developmental disorders*, 33(6):553–567.

Bertram C. Bruce, Ann D. Rubin, and Kathleen S. Starr. 1981. Why readability formulas fail. *IEEE Transactions on Professional Communication*, PC-24:50–52.

Terry S. Brugha, Sally Anne Cooper, and Sally McManus. 2012. Estimating the Prevalence of Autism Spectrum Conditions in Adults: Extending the 2007 Adult Psychiatric Morbidity Survey. Technical report, NHS, The Health and Social Care Information Centre., London.

Jeanne S. Chall and Edgar Dale. 1995. *Readability Revisited: the new Dale-Chall readability formula*. Brookline Books, Cambridge, Massachusetts.

Marjorie H. Charlop-Christy, Michael Carpenter, Loc Le, Linda A. LeBlanc, and Kristen Kellet. 2002. Using the picture exchange communication system (pecs) with children with autism: assessment of pecs acquisition, speech, social-communicative behavior, and problem behaviour. *JOURNAL OF APPLIED BEHAVIOR ANALYSIS*, 3(3):213–231.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the First Conference on North American Chapter of the Association for Computational Linguistics*, pages 132–139, San Francisco.

Edgar Dale and Jeanne S. Chall. 1948. A formula for predicting readability: Instructions. *Educational Research Bulletin*, 27(2):37–54.

Richard R. Day and Jeong-Suk Park. 2005. Developing Reading Comprehension Questions. *Reading in a Foreign Language*, 17(1).

William H. Dubay. 2004. *The Principles of Readability*. Impact Information.

Lijun Feng, Martin Jansche, Matt Huenerfauth, and Noémie Elhadad. 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.

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

Uta Frith and Maggie Snowling. 1983. Reading for meaning and reading for sound in autistic and dyslexic children. *Journal of Developmental Psychology*, 1:329–342.

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*, 12(4):395–427.

Arthur C. Graesser, Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai. 2004. Coh-metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36:193–202.

Claire Grover, Colin Matheson, Andrei Mikheev, and Marc Moens. LT TTT - a flexible tokenisation tool.

Robert Gunning. 1952. *The technique of clear writing*. McGraw-Hill, New York.

Francesca Happé and Uta Frith. 2006. The weak coherence account: Detail focused cognitive style in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36:5–25.

Francesca Happé. 1997. Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, 15:1–12.

Matt Huenerfauth, Lijun Feng, and Noémie Elhadad. 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.

Martin Jansche, Lijun Feng, and Matt Huenerfauth. 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.

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 Foundation Trust, London, UK.

J. Peter Kincaid, Robert P. Fishburne, Richard L. Rogers, and Brad S. Chissom. 1975. *Derivation of new readability formulas (Automatic Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. CNTECHTRA, 8-75 edition.

J. Peter Kincaid, James A. Aagard, John W. O'Hara, and Larry K. Cottrell. 1981. Computer readability editing system. *IEEE transactions on professional communications*.

Anne Kispal. 2008. Effective Teaching of Inference Skills for Reading. Literature Review.

Meng-Chuan Lai, Michael V. Lombardo, and Simon Baron-Cohen. 2014. Autism. *Lancet*, 383(9920):896–910.

Gilbert MacKay and Adrienne Shaw. 2004. A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy*, 20(13).

Juan Martos, Sandra Freire, Ana González, 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.

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

Danielle S. McNamara, Max M. Louwerse, Philip M. McCarthy, and Arthur C. Graesser. 2010. Coh-Metrix: Capturing Linguistic Features of Cohesion, May.

Irene M. O'Connor and Perry D. Klein. 2004. Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 34:2:115–127.

Stephen Oliver. 1998. *Understanding Autism*. Oxford Brookes University, UK.

Cynthia Putnam and Lorna Chong. 2008. Software and technologies designed for people with autism: What do users want? In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, pages 3–10, New York, NY, USA. ACM.

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

Luz Rello, Clara Bayarri, and Azuki Gorriz. 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.

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

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

Dean R. Smith, A. Jackson Stenner, Ivan Horabin, and III Malbert Smith. 1989. The lexile scale in theory and practice: Final report. Technical report, MetaMetrics (ERIC Document Reproduction Service No. ED307577)., Washington, DC:.

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

Sanja Štajner, Ruslan Mitkov, and Gloria Corpas Pastor, 2014. *Simple or not simple? A readability question*. Springer-Verlag, Berlin.

Ashley Wazana. 2007. The Autism Epidemic: Fact or Artifact? *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(6):721 – 730.