Readability Indices for Automatic Evaluation of Text Simplification Systems: A Feasibility Study for Spanish

Sanja Štajner
Research Group in Computational Linguistics
University of Wolverhampton, UK
sanjastajner@wlv.ac.uk

Horacio Saggion
TALN Research Group
Universitat Pompeu Fabra, Spain
horacio.saggion@upf.edu

Abstract

This paper addresses the problem of automatic evaluation of text simplification systems for Spanish. We test whether already-existing readability formulae would be suitable for this task. We adapt three existing readability indices (two measuring lexical complexity and one measuring syntactic complexity) to be computed automatically, which are then applied to a corpus of original news texts and their manual simplifications aimed at people with cognitive disabilities. We show that there is a significant correlation between each of the three readability indices and several linguistically motivated features which might be seen as reading obstacles for various target populations. Furthermore, we show that there is a significant correlation between the two readability indices which measure lexical complexity.

1 Introduction

In recent years, there has been growing effort to simplify written material and make it equally accessible to everyone. Various studies indicate that lexically and syntactically complex texts can be very difficult for non-native speakers and people with various reading impairments (e.g. autistic, aphasic, dyslexic or deaf people). Aphasic people, for instance, may encounter problems with less frequent words and some particular sentence constructions (Devlin, 1999). They also have problems in understanding syntactic constructions which do not follow the canonical subject-verb-object structure (e.g. passive constructions), and especially those sentences which are semantically reversible, e.g. *"The boy was kissed by the girl"* (Carroll et al., 1999). Additionally, aphasic readers may have additional problems with

comprehending newswire texts which have some genre-specific characteristics. These types of texts tend to use long sentences, noun compounds and long sequences of adjectives, e.g. *"Twenty-five-year-old blonde-haired mother-of-two Jane Smith"* (Carroll et al., 1999). People with intellectual disabilities have problem with both lexically and syntactically complex texts, as well as with processing and loading large amounts of information (Feng, 2009).

Since the late nineties, several initiatives which proposed guidelines for producing plain, easy-to-read and more accessible documents have emerged, e.g. "The Plain Language Action and Information Network (PLAIN)"[1], "Make it Simple, European Guidelines for the Production of Easy-to-Read Information for people with Learning Disability" (Freyhoff et al., 1998), "Am I making myself clear? Mencap's guidelines for accessible writing"[2], and "Web content accessibility guidelines"[3]. All these guidelines share similar instructions for accessible writing, some of them more detailed than others. They all advise the writer to use the active voice instead of passive, use the simplest form of a verb (present and not conditional or future), avoid hidden verbs (i.e. verbs converted into a noun), use short, simple words and omit unnecessary words, write short sentences and cover only one main idea per sentence, etc.

Armed with these guidelines and the aim of making written documents equally accessible to everyone, many attempts have been made to completely or at least partially automate the process of text simplification, which is very expensive and time-consuming when performed manually. So far, text simplification systems have been devel-

---

[1] http://www.plainlanguage.gov/
[2] http://november5th.net/resources/Mencap/Making-Myself-Clear.pdf
[3] http://www.w3.org/TR/WCAG20/

oped for English (Zhu et al., 2010; Coster and Kauchak, 2011; Woodsend and Lapata, 2011; Wubben et al., 2012), Spanish (Saggion et al., 2011), and Portuguese (Aluísio et al., 2008), with recent attempts at Basque (Aranzabe et al., 2012), Swedish (Rybing et al., 2010), and Dutch (Ruiter et al., 2010). With the emergence of these systems, the question we are faced with is how to automatically evaluate their performance given that the access to the target users might be difficult.

This study is an attempt to address this issue. We focus on text simplification systems for Spanish and investigate whether some of the already existing readability indices could be used for the automatic evaluation of these systems. Using a corpus of original news texts and their manual simplifications which followed specific guidelines for writing for people with cognitive disabilities, we show that two lexical complexity indices – one suggested by Anula (2007), and other by Spaulding (1956) – are highly correlated in both these text sets. Furthermore, we show that both these indices and the third readability index concerned with syntactic complexity (Anula, 2007) could be used for automatic evaluation of text simplification systems, as each index is correlated with some subset of the linguistically motivated complexity features considered as obstacles for people with different reading impairments.

The remainder of the article is structured as follows: Section 2 presents the most important previous work on readability prediction and linguistically motivated complexity features considered as obstacles for people with different reading difficulties; Section 3 describes the corpora, features, and readability indices used in this study; Section 4 presents and discusses the results of analysis of three chosen readability indices, twelve linguistically motivated complexity features, and their mutual correlation; while Section 5 concludes the article by summarising the main contributions and proposing possible directions for future work.

2 Related Work

Since the 1950s, over 200 readability formulae have been developed (for the English language), with over 1000 studies of their application (DuBay, 2004). Initially, they were used to assess the grade level of textbooks. Later, they were adapted to different domains and purposes, e.g. to measure readability of technical manuals

(Automated Readability Index (Smith and Senter, 1967)), or US healthcare documents intended for the general public (the SMOG grading (McLaughlin, 1969)). Some of these first readability formulae are still widely in use, given their simplicity (they require only the average sentence and word length) and good correlation with the reading tests. One of the most used readability formulae – the Flesch Reading Ease score (Flesch, 1949) – for example, "correlates .70 with the 1925 McCall-Crabbs reading test and .64 with the 1950 version of the same test" (DuBay, 2004). Another set of readability formulae are those which depend on average sentence length and the percentage of words which cannot be found on a list of the "easiest" words, e.g. the Dale-Chall readability formulae (Dale and Chall, 1948). These formulae have been adapted to other languages by changing the coefficient before the factors (e.g. the Flesch-Douma (Douma, 1960) and Leesindex Brouwer (Brouwer, 1963) formulae for Dutch represent the adaptations of the Flesch Reading Ease score, while Spaulding's Spanish readability formula (Spaulding, 1956) could be seen as an adaptation of the Dale-Chall formula (Dale and Chall, 1948)). Oosten et al. (2010) showed that readability formulae which are solely based on superficial text characteristics (average sentence and word length) seem to be strongly correlated even across different languages (English, Dutch, and Swedish).

With the recent advances of natural language processing (NLP) tools and techniques, new approaches to readability assessment have emerged. Schwarm and Ostendorf (2005), and Petersen and Ostendorf (2009), used statistical language modeling and support vector machines to show that more complex features (e.g. average height of the parse tree, average number of noun and verb phrases, etc.) give better readability prediction than the traditional Flesch-Kincaid readability formula. They based their approach on the texts from Weekly Reader[4], and two smaller corpora: Encyclopedia Britannica and Britannica Elementary (Barzilay and Elhadad, 2003), and CNN news stories and their abridged vesions[5]. Feng et al. (2009) introduced some new cognitively motivated features which should improve automatic readability assessment of texts for people with cognitive dis-

---

abilities. In addition to three previously used corpora (Weekly Reader, Britannica, and CNN news stories) aimed at second language learners or children, Feng et al. (2009) used a corpus of local news articles which were simplified by human editors in order to make them more accessible for people with mild intellectual disabilities (MID). The texts were further rated for readability by people with MID. The study (Feng et al., 2009) showed that their newly introduced cognitively motivated features (e.g. entity mentions, lexical chains, etc.) are better correlated with the user-study comprehension than the Flesch-Kincaid Grade Level index (Kincaid et al., 1975).

Štajner et al. (2012) stated that many features which could be automatically extracted from a parser's output can indicate the occurrence of the obstacles to reading comprehension faced by people with autism. The authors referred to the syntactic concept of the projection principle (Chomsky, 1986) that "lexical structure must be represented categorically at every syntactic level" which implies "that the number of noun phrases in a sentence is proportional to the number of nouns in that sentence, the number of verbs in a sentence is related to the number of clauses and verb phrases, etc." (Štajner et al., 2012). Therefore, they automatically extracted nine features which account for indicators of structural complexity (nouns, adjectives, determiners, adverbs, verbs, infinitive markers, coordinating conjunctions, subordinating conjunctions, and prepositions), and three which account for indicators of ambiguity in meaning (pronouns, definite descriptions, and word senses). Štajner et al. (2012) showed that many of these features are significantly correlated with the Flesch Reading Ease score (Flesch, 1949). Given that all of the reading obstacles for people with autism (Štajner et al., 2012) would also be difficult to understand for people with cognitive disabilities (Freyhoff et al., 1998; Feng, 2009), we believe that these features (Section 3.3) could also be a good measure of complexity reduction achieved in a text simplification system.

Motivated by the study of Štajner et al. (2012), we wanted to explore how these features are correlated with the existing readability formulae (this time for Spanish instead of English). These formulae were not initially intended to be used for the evaluation of text simplification systems but rather to measure the grade level necessary to understand a given text. Therefore, we wanted to establish whether those readability indices could be used in an automatic evaluation of text simplification systems. To the best of our knowledge, this is the first study of this type for Spanish. Unlike the study of Štajner et al. (2012) which uses the Simple Wikipedia[6] as an example of simplified texts (which do not comply totally with easy-to-read guidelines for people with cognitive disabilities, but are rather intended for a much wider audience), our study uses the original news texts and their manual simplifications aimed at people with cognitive disabilities, following specifically tailored easy-to-read guidelines for this target population (Section 3).

3 Methodology

The corpora, readability indices and linguistically motivated complexity features used in this study are presented in the next three subsections.

3.1 Corpora

We first compared all features and readability measures on a parallel corpus of original and manually simplified texts (Table 1) in order to investigate whether these complexity measures differ significantly on these two types of texts, thus justifying the idea to use them to measure the degree of the performed simplification. The corpus contains 200 original news articles in Spanish (provided by the Spanish news agency Servimedia[7]) and their manually simplified versions. Simplification was done by trained human editors, familiar with the particular needs of a person with cognitive disabilities and following a series of easy-to-read guidelines suggested by Anula (2007), as a part of the Simplext project[8] (Saggion et al., 2011).

Table 1: Corpora

Corpus	Texts	Sentences	Words
Original	200	1150	37121
Simplified	200	1804	24332

The simplification operations applied by human editors could be classified in the following four categories (Drndarevic et al., 2013):

---

[6]http://simple.wikipedia.org
[7]http://www.servimedia.es
[8]http://www.simplext.es/

1. **Syntactic operations**: changes applied at the sentence level, such as sentence splitting or quotation inversion.

2. **Lexical operations**: infrequent, long or technical terms are substituted with their simpler synonyms, and certain expressions are paraphrased or otherwise modified.

3. **Content reduction**: a significant portion of original content is eliminated through summarisation and paraphrases, in accordance with the guidelines that indicate that only the most essential piece of information should be preserved.

4. **Clarification**: certain complex terms and concepts, for which no synonym can be found, are explained by means of a definition.

3.2 Readability Indices

In this study, we focused on three readability formulae for Spanish: two concerned with the lexical complexity of the text – LC (Anula, 2007) and SSR (Spaulding, 1956); and the third one concerned with the syntactic complexity of the given text – SCI (Anula, 2007).

The Spaulding's Spanish Readability index (SSR) has been used for assessing the reading difficulty of fundamental education materials for Latin American adults of limited reading ability and for the evaluation of text passages of the foreign language tests (Spaulding, 1956). It predicts the relative difficulty of reading material based on the vocabulary and sentence structure, using the following formula:

$$SSR = 1.609 \times \frac{|w|}{|s|} + 331.8 \times \frac{|rw|}{|w|} + 22.0$$

Here, $|w|$ and $|s|$ denote the number of words and sentences in the text, while $|rw|$ denotes the number of rare words in the text. According to Spaulding (1956), *rare words* are those words which cannot be found on the list of 1500 most common Spanish words provided in the same study[9]. Given that the SSR index was used for assessing the reading difficulty of the materials aimed at adults of limited reading ability, it is reasonable to expect that this formula could be used for estimating the level of simplification performed by text simplification systems aimed at making texts more accessible for this target population.

The Lexical Complexity index (LC) was suggested by Anula (2007) as a measure of lexical complexity of literary texts aimed at the second language learners. It is calculated using the following formula:

$$LC = \frac{LDI + ILFW}{2}$$

where *LDI* and *ILFW* represent the *Lexical Density Index* and *Index of Low-Frequency Words*, respectively:

$$LDI = \frac{|dcw|}{|s|},$$

$$ILFW = \frac{|lfw|}{|cw|} \times 100$$

Here, $|dcw|$, $|s|$, $|lfw|$, and $|cw|$ denote the number of distinct content words, sentences, low-frequency words, and content words (nouns, adjectives, verbs, and adverbs), respectively. Anula (2007) considers as *low frequency words* those words whose frequency rank in the Referential Corpus of Contemporary Spanish[10] is lower than 1,000.[11]

The Sentence Complexity Index (SCI) was proposed by Anula (2007) as a measure of sentence complexity in a literary text aimed at second language learners. It is calculated by the following formula:

$$SCI = \frac{ASL + ICS}{2}$$

where *ASL* denotes the average sentence length, and *ICS* denotes the index of complex sentences. They are calculated as follows:

$$ASL = \frac{|w|}{|s|},$$

$$ICS = \frac{|cs|}{|s|} \times 100$$

---

[9] Detailed instructions on what should be considered as a *rare word* (e.g. special cases of numbers, names of months and days, proper and geographic names, initials, diminutives and augmentatives, etc.) can be found in (Spaulding, 1956). Here we do not apply rules (a)–(g) specified in (Spaulding, 1956).

[10] http://corpus.rae.es/lfrecuencias.html

[11] Both lists (from Referential Corpus of Contemporary Spanish and the Spaulding's list of 1500 most common Spanish words) were lemmatised using Connexor's parser in order to retrieve the frequency of the lemma and not a word form (action carried out manually in the two cited works).

Here, $|w|$, $|s|$, and $|cs|$ denote the number of words, sentences and complex sentences in the text, respectively.[12]

3.3 Linguistically Motivated Complexity Features

Inspired by the work of Štajner et al. (2012), and easy-to-read guidelines for writing for people with cognitive disabilities (Freyhoff et al., 1998), this study employs twelve linguistically motivated complexity features (Table 2). The first nine features (1–9) are indicators of structural complexity and the final three features (10–12) are indicators of ambiguity in meaning.

Table 2: Linguistically motivated features

#	Code	Feature
1	*N*	Noun
2	*Det*	Determiner
3	*Adj*	Adjective
4	*V*	Verb
5	*Inf*	Infinitive
6	*Adv*	Adverb
7	*Prep*	Preposition
8	*CC*	Coordinating conjunction
9	*CS*	Subordinating conjunction
10	*Pron*	Pronoun
11	*Sens*	Number of senses per word
12	*Amb*	Percentage of ambiguous words

The corpora were parsed with the Connexor's Machinese parser[13] and the features 1–10 (Table 2) were automatically extracted using the parser's output. Features 11 and 12 were extracted using two lexical resources – the Spanish Open Thesaurus (version 2)[14] and the Spanish EuroWord-Net (Vossen, 1998). The Spanish Open Thesaurus lists 21,831 target words (lemmas) and provides a list of word senses for each word. Each word sense is, in turn, a list of substitute words. There is a total of 44,353 such word senses. The Spanish part of EuroWordNet is far more exhaustive containing 50,526 word meanings and 23,370 synsets. For computation of measures related to word sentences we only considered the lemmas present in the lexical resources used. For each text we com-

pute the average number of senses per word (code *Sens*, Table 2) as well as the percentage of ambiguous words in the text (code *Amb*, Table 2) producing two measures for each lexical resource used (*SensWN*, *SensOT*, *AmbWN*, *AmbOT*, Section 4). In the computation we consider all occurrences of lemmas including repeated lemmas.

4 Results and Discussion

The results of the analysis of readability indices on the corpora and their mutual correlation are presented in Section 4.1, and the results of the analysis of linguistically motivated complexity features are presented in Section 4.2, while their correlation with the readability indices is presented and discussed in Section 4.3.

4.1 Analysis of Readability Indices

The results of the comparison of readability indices across the corpora are given in Table 3. Columns 'Original' and 'Simple' contain the mean value of the corresponding readability indices in each of the two corpora, while the column 'Rel.diff.' contain the mean value of the relative differences between the text pairs (original and simplified). Column 'Sign.' presents the level of significance at which the differences between the two corpora are statistically significant. For the indices which follow approximately normal distribution, this column contains the result of the paired t-test. For those which do not follow normal distribution, it contains the result of the alternative non-parametric test – the Wilcoxon signed-rank test. All tests of normality and statistical significance were performed in SPSS.

Table 3: Readability indices

Index	Original	Simple	Rel.diff.	Sign.
LC	21.05	12.76	-39.06%	0.001
SSR	184.20	123.82	-32.60%	0.001
SCI	41.36	29.99	-27.43%	0.001

The results presented in Table 3 clearly demonstrate that there is a significant difference between the original and manually simplified texts for all three readability indices. The text pairs show an average relative difference of almost 40% for LC and about 30% for SSR and SCI, thus justifying the idea that those readability indices might be used in an automatic evaluation of text simplifica-

---

[12]We consider a complex sentence one that contains multiple finite predicates according to the output of Connexor's Machinese parser.

[13]www.connexor.eu
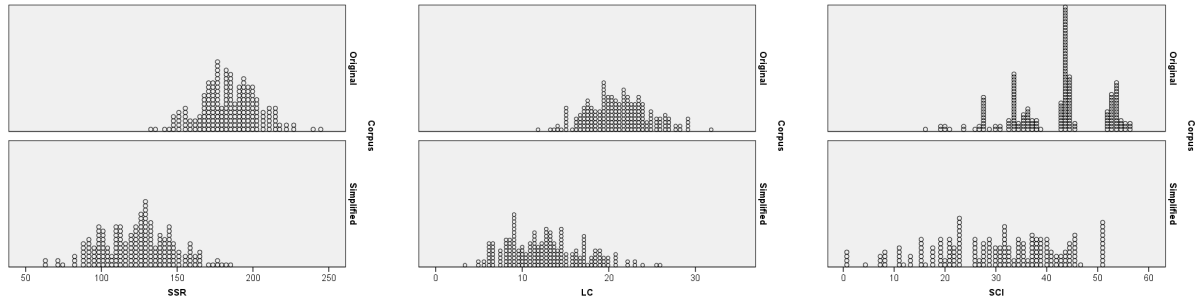
[14]http://openthes-es.berlios.de/

Figure 1: Readability indices across the corpora

tion systems as a measure of the degree of simplification. The distribution of the three readability indices (LC, SSR, and SCI) is presented in Figure 1, which shows that the distribution of all three indices is shifted left in the case of the simplified texts, thus indicating lower level of complexity.

The correlations between each pair of readability indices (LC–SSR, LC–SCI, and SSR–SCI), calculated using both corpora, are given in Table 4. All correlations which were reported as statistically significant at a 0.001 level of significance are presented in bold. As expected, the two readability indices concerned with the lexical complexity (LC and SSR) are significantly correlated, while the third one concerned with the syntactic complexity (SCI) is not significantly correlated with any of the other two (LC and SSR). The linear correlation between LC and SSR (measured by the Pearson's coefficient) is, however, much less strong than the one among the four readability indices for English: Flesch Reading Ease, Flesch-Kincaid, Fog and SMOG, reported by Štajner et al. (2012).

Table 4: Correlation among readability indices

Corpus	Indices	Pearson	Spearman
	LC–SSR	**0.445**	**0.440**
Original	LC–SCI	-0.075	-0.085
	SSR–SCI	0.045	0.043
	LC–SSR	**0.353**	**0.378**
Simplified	LC–SCI	0.093	-0.116
	SSR–SCI	-0.159	-0.136

4.2 Analysis of Linguistically Motivated Complexity Features

Occurrences of each feature which is an indicator of structural complexity, and prepositions (*Prep*)

were calculated as number of occurrences per 100 words. Average number of senses per word and percentage of ambiguous words in text were calculated in two different ways – using the Spanish EuroWordNet (SenseWN and AmbWN) and using the Spanish Open Thesaurus (SenseOT and AmbOT). The results of the analysis are presented in Table 5, using the same notation as in the case of readability indices in Table 3.

Table 5: Complexity Features

Feature	Original	Simple	Rel.diff.	Sign.
N	33.12	33.32	+1.13%	no
Det	14.82	17.13	+17.65%	0.001
Adj	7.24	4.89	-31.10%	0.001
V	10.39	14.56	+45.70%	0.001
Inf	1.65	2.22	+38.14%	0.001
Adv	2.27	3.35	+83.45%	0.001
Prep	19.75	17.12	-12.42%	0.001
CC	2.97	1.63	-41.79%	0.001
CS	1.82	2.55	+53.96%	0.001
Pron	19.75	17.12	+11.82%	0.001
SenseWN	3.78	4.01	+6.99%	0.001
AmbWN	66.02	72.19	+9.62%	0.001
SenseOT	3.52	3.65	+4.47%	0.001
AmbOT	78.89	82.71	+5.13%	0.001

The results in Table 5 show that the number of occurrences (per 100 words) of nouns does not differ significantly between the two corpora. Simplified texts have significantly lower number of occurrences (per 100 words) of adjectives, prepositions and coordinating conjunctions. This could be interpreted as an indication of omitting unnecessary information (adjectives), removing/resolving syntactic ambiguity and complexity (prepositions) and sentence splitting (coordinating conjunctions) in the process of simplification. The increased percentage of verbs might be a reflection of omitting

379

Table 6: Spearman's correlation between readability indices and complexity features (Original)

Feature	LC	SSR	SCI
V	*-0.178	-0.192	0.423
Inf	*-0.154	*-0.151	0.303
Adj	*-0.159	0.137	-0.100
Adv	-0.024	-0.047	0.123
Det	-0.022	-0.243	-0.076
N	*0.177	0.193	-0.433
Prep	0.088	0.049	-0.122
CC	0.065	0.116	-0.086
CS	-0.092	*-0.150	0.459
Pron	0.072	-0.248	0.097
SensWN	-0.285	-0.231	0.236
AmbWN	-0.243	-0.080	*0.154
SensOT	-0.077	-0.093	0.088
AmbOT	-0.208	-0.083	0.099

Table 7: Spearman's correlation between readability indices and complexity features (Simplified)

Feature	LC	SSR	SCI
V	0.000	-0.059	0.672
Inf	-0.025	-0.074	0.573
Adj	-0.241	0.086	*-0.145
Adv	-0.113	-0.118	0.246
Det	-0.086	-0.438	0.034
N	*0.161	0.375	-0.606
Prep	*0.156	0.088	*-0.153
CC	0.027	0.108	*-0.150
CS	-0.030	*-0.159	* 0.595
Pron	0.002	-0.074	-0.186
SensWN	-0.064	-0.070	0.225
AmbWN	-0.110	-0.075	0.115
SensOT	0.053	0.025	0.113
AmbOT	0.110	0.113	0.045

the unnecessary words (e.g. adjectives, coordinating conjunctions, prepositions) and leaving only the main ideas expressed by verbs.

It is interesting to note that both the average number of senses per word and the percentage of ambiguous words are higher in simplified than in original texts, using both sources (EuroWordNet and Open Thesaurus). One possible explanation (which would have to be explored further) is that the shorter and more commonly used words are more ambiguous than the original words which they substituted in the process of simplification.

4.3 Correlation between Readability Indices and Complexity Features

The Spearman's rho coefficient of correlation between readability indices and the twelve linguistically motivated complexity features is given in Table 6 (for original texts) and in Table 7 (for simplified texts). Correlations which are significant at a 0.001 level of significance (2-tailed) are presented in bold, while those which are significant at a 0.05 but not at a 0.001 level of significance are presented in bold with an '*' preceding. Other correlations are not statistically significant.

From the results presented in Table 6 and Table 7 it can be noted that each of the readability indices is significantly correlated with several linguistically motivated complexity features. LC is, for example, positively correlated with occurrences of nouns (*N*) and negatively correlated with occurrences of adjectives (*Adj*) in both corpora. SSR

is positively correlated with occurrences of nouns (*N*) and negatively correlated with occurrences of determiners (*Det*) and subordinating conjunctions (*CS*). SCI is, on the other hand, negatively correlated with the number of occurrences of nouns (*N*), and positively correlated with number of occurrences of verbs (*V*), infinitive forms (*Inf*), subordinating conjunctions (*CS*), and average number of senses per word according to Spanish EuroWord-Net (*SensWN*).

These results indicate that there is no one readability index which correlates significantly with all of the linguistically motivated complexity features. However, it seems that they complement each other well as each one of them is significantly correlated with a different subset of features. Each of these three readability indices could, therefore, be seen as a measure of a different kind of complexity reduction performed by a text simplification system and thus be used in an automatic evaluation of a text simplification system. That automatic evaluation would, of course, account only for measuring the complexity reduction performed by the system, while a human-oriented evaluation would be needed for assessing the preservation of meaning and grammaticality of the simplified text generated by the system (Drndarevic et al., 2013).

5 Conclusions and Future Work

The results presented in this study revealed that there are significant differences between the val-

ues of the three readability indices (LC, SSR, and SCI) applied to the corpus of original news texts and the same applied to manually simplified versions of those texts (aimed at people with cognitive disabilities). Another set of experiments indicated that the two corpora also significantly differ in all but one of the twelve linguistically motivated complexity features.

The study also revealed that the two readability indices which measure lexical complexity of a given text are highly correlated. It also showed that each of the three readability indices (LC, SSR and SCI) significantly correlates with several linguistically motivated complexity features in both corpora. Each of them could thus be used in an automatic evaluation of a text simplification system, each measuring a different kind of complexity reduction performed. Furthermore, it seems that those three readability indices complement each other very well in terms of their correlation with different complexity features. Therefore, it might be possible to find some combination of all three of them which could be used as a single measure in an automatic evaluation of text simplification systems.

The search for this ideal combination will be one of the directions of our future work. We also plan to repeat all these experiments on a different set of texts, this time aimed at a different target population, in order to see whether these readability indices show the same properties for texts simplified in a different manner, i.e. whether they could be used in automatic evaluation of any text simplification system. Furthermore, we wish to apply these indices on texts which were automatically simplified. We would like to explore how well the conclusions drawn based on differences of readability indices between original and automatically simplified texts correlate with human judgments of the level of simplification performed.

Acknowledgements

References

S. M. Aluísio, L. Specia, T. A. S. Pardo, E. G. Maziero, H. M. Caseli, and R. P. M. Fortes. 2008. A corpus analysis of simple account texts and the proposal of simplification strategies: first steps towards text simplification systems. In *Proceedings of the 26th annual ACM international conference on Design of communication*, SIGDOC '08, pages 15–22, New York, NY, USA. ACM.

A. Anula. 2007. Tipos de textos, complejidad lingüística y facilicitación lectora. In *Actas del Sexto Congreso de Hispanistas de Asia*, pages 45–61.

M. J. Aranzabe, A. Díaz De Ilarraza, and I. González. 2012. First Approach to Automatic Text Simplification in Basque. In *Proceedings of the first Natural Language Processing for Improving Textual Accessibility Workshop (NLP4ITA)*.

R. Barzilay and N. Elhadad. 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.

R. H. M. Brouwer. 1963. Onderzoek naar de leesmoeilijkheden van nederlands proza. *Pedagogische studiën*, 40:454–464.

J. Carroll, G. Minnen, D. Pearce, Y. Canning, S. Devlin, and J. Tait. 1999. Simplifying text for language-impaired readers. In *Proceedings of the 9th Conference of the European Chapter of the ACL (EACL'99)*, pages 269–270.

N. Chomsky. 1986. *Knowledge of language: its nature, origin, and use.* Greenwood Publishing Group, Santa Barbara, California.

W. Coster and D. Kauchak. 2011. Learning to Simplify Sentences Using Wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 1–9.

E. Dale and J. S. Chall. 1948. A formula for predicting readability. *Educational research bulletin*, 27.

S. Devlin. 1999. *Simplifying natural language text for aphasic readers.* Ph.D. thesis, University of Sunderland, UK.

W.H. Douma. 1960. De leesbaarheid van landbouwbladen: een onderzoek naar en een toepassing van leesbaarheidsformules. *Bulletin*, 17.

B. Drndarevic, S. Štajner, S. Bott, S. Bautista, and H. Saggion. 2013. Automatic Text Simplication in Spanish: A Comparative Evaluation of Complementing Components. In *Proceedings of the 12th International Conference on Intelligent Text Processing and Computational Linguistics. Lecture Notes in Computer Science. Samos, Greece, 24-30 March, 2013.*, pages 488–500.

W. H. DuBay. 2004. The Principles of Readability. *Impact Information*.

L. Feng, N. Elhadad, and M. Huenerfauth. 2009. Cognitively motivated features for readability assessment. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 229–237, Stroudsburg, PA, USA. Association for Computational Linguistics.

L. Feng. 2009. Automatic readability assessment for people with intellectual disabilities. In *SIGACCESS Access. Comput.*, number 93, pages 84–91. ACM, New York, NY, USA, jan.

R. Flesch. 1949. *The art of readable writing*. Harper, New York.

G. Freyhoff, G. Hess, L. Kerr, B. Tronbacke, and K. Van Der Veken, 1998. *Make it Simple, European Guidelines for the Production of Easy-toRead Information for People with Learning Disability*. ILSMH European Association, Brussels.

J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom. 1975. Derivation of new readability formulas for navy enlisted personnel. *Research Branch Report 8-75*.

G. H. McLaughlin. 1969. SMOG grading – a new readability formula. *Journal of Reading*, 22:639–646.

S. Petersen and M. Ostendorf. 2009. A machine learning approach to reading level assessment. *Computer speech & language*, 23(1):89–106.

M. B. Ruiter, T. C. M. Rietveld, C. Cucchiarini, E. J. Krahmer, and H. Strik. 2010. Human Language Technology and communicative disabilities: Requirements and possibilities for the future. In *Proceedings of the the seventh international conference on Language Resources and Evaluation (LREC)*.

J. Rybing, C. Smithr, and A. Silvervarg. 2010. Towards a Rule Based System for Automatic Simplification of Texts. In *The Third Swedish Language Technology Conference*.

H. Saggion, E. Gómez Martínez, E. Etayo, A. Anula, and L. Bourg. 2011. Text Simplification in Simplext: Making Text More Accessible. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*.

S. E. Schwarm and M. Ostendorf. 2005. Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In *Proceedings of the 43rd annual meeting of the Association of Computational Linguistics (ACL)*, pages 523–530.

E. A. Smith and R. J. Senter. 1967. Automated Readability Index. Technical report, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

S. Spaulding. 1956. A Spanish Readability Formula. *Modern Language Journal*.

P. van Oosten, D. Tanghe, and V. Hoste. 2010. Towards an Improved Methodology for Automated Readability Prediction. In *Proceedings of the seventh international conference on language resources and evaluation (LREC10). Valletta, Malta: European Language Resources Association (ELRA)*.

P. Vossen, editor. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.

S. Štajner, R. Evans, C. Orasan, and R. Mitkov. 2012. What Can Readability Measures Really Tell Us About Text Complexity? In *Proceedings of the LREC'12 Workshop: Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, Istanbul, Turkey.

K. Woodsend and M. Lapata. 2011. Learning to Simplify Sentences with Quasi-Synchronous Grammar and Integer Programming. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

S. Wubben, A. van den Bosch, and E. Krahmer. 2012. Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 1015–1024, Stroudsburg, PA, USA. Association for Computational Linguistics.

Z. Zhu, D. Berndard, and I. Gurevych. 2010. A Monolingual Tree-based Translation Model for Sentence Simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361.