Automatic Text Simplification in Spanish: A Comparative Evaluation of Complementing Modules

Biljana Drndarević[1], Sanja Štajner[2], Stefan Bott[1], Susana Bautista[3], and Horacio Saggion[1]

[1] Universitat Pompeu Fabra, Barcelona, Spain
{biljana.drndarevic,stefan.bott,horacio.saggion}@upf.edu
[2] University of Wolverhampton, Wolverhampton, UK
sanjastajner@wlv.ac.uk
[3] Universidad Complutense de Madrid, Madrid, Spain
subautis@fdi.ucm.es

Abstract. In this paper we present two components of an automatic text simplification system for Spanish, aimed at making news articles more accessible to readers with cognitive disabilities. Our system in its current state consists of a rule-based lexical transformation component and a module for syntactic simplification. We evaluate the two components separately and as a whole, with a view to determining the level of simplification and the preservation of meaning and grammaticality. In order to test the readability level pre- and post-simplification, we apply seven readability measures for Spanish to three sets of randomly chosen news articles: the original texts, the output obtained after lexical transformations, the syntactic simplification output, and the output of both system components. To test whether the simplification output is grammatically correct and semantically adequate, we ask human annotators to grade pairs of original and simplified sentences according to these two criteria. Our results suggest that both components of our system produce simpler output when compared to the original, and that grammaticality and meaning preservation are positively rated by the annotators.

1 Introduction

Automatic text simplification as an NLP task arose from the necessity to make electronic textual content equally accessible to everyone. Organisations such as Inclusion Europe[1] point out to the essential right for every person to take active part in the life of their society through access to information. Nevertheless, numerous people experience difficulties reading government reports, laws, news articles and other written material that enables their inclusion in the community. Some Internet portals have created simplified variants of their content, as is the case with Simple English Wikipedia[2]. However, simplifying text manually

[1] http://inclusion-europe.org/en
[2] http://simple.wikipedia.org/wiki/Main_Page

A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 488–500, 2013.

is time-consuming and not cost-effective, especially in the case of news articles, which are constantly being generated and updated. That is why attempts have been made to automate the laborious process of text simplification. So far, systems have been developed for English [1], Portuguese [2] and Japanese [3], with recent attempts at Basque [4] and Swedish [5] text simplification.

Automatic text simplification is a complex task which encompasses a number of operations applied at different linguistic levels. The aim is to turn a complex text into its simplified variant, taking into consideration the specific needs of a particular target user. Our Simplext project is one such aspiration [6]. We have been developing a system for automatic text simplification in Spanish, aimed at producing more readable news articles for people with cognitive disabilities. We conduct simplification at the syntactic and the lexical levels of the input text. Easy-to-read guidelines indicate that a single idea should be expressed per sentence [7], so we divide a complex sentence into as many simple sentences as possible, as part of our syntactic simplification strategy. The guidelines also suggest that common and simple words should be used to express the desired idea, and that the use of technical and complex vocabulary should be avoided [8]. This entails treatment of the lexical items of the input text. We here describe one component of our lexical simplification module, which applies rule-based transformations to phrases and expressions that cannot be simplified through a more traditional synonym substitution approach. However, we do employ the latter approach for the second component of our lexical module [9], currently under development and not presented on this occasion. Our main goal here is to concentrate on the evaluation of the two existing components of our system, and test their performance in terms of the grade of simplification, the grammaticality of the output, and the preservation of original meaning.

The remainder of the article is organised as follows: in Section 2 we present an overview of the most relevant work in the field of automatic text simplification; in Section 3 we outline our approach to the task at hand, describe in some detail the different components of our system, and present the experimental setting, while section 4 discusses the results of our evaluation experiments; we conclude the article with a summary and plans for future work in Section 5.

2 Related Work

Automatic text simplification has traditionally had a double purpose. It can be used as a preprocessing tool for other NLP applications [10], where it serves the purpose of improving their performance. On the other hand, it has been widely used to offer simpler reading material for target users, such as foreign language learners [11], readers with aphasia [12], low literacy individuals [13], etc. The first attempts are rule-based syntactic simplification systems [14]. Carroll et al. [15] contribute with an additional lexical simplification module, and introduce the paradigm, often repeated thereafter, of simplification based on synonym substitution. They use WordNet to obtain a set of potential synonyms of content words in the input text, and determine the simplest out of the set by looking up Kucera-Francis frequencies in the Oxford Psycholinguistic Database [16]. Word frequency

is, therefore, seen as a measure of lexical complexity, and this approach has been adopted in a number of works that follow. Bautista et al. [17] use a similar approach but introduce word length as an additional indicator of word difficulty. De Belder et al. [18] were the first to introduce a word-sense-disambiguation element to their lexical simplification system in order to account for numerous cases of polysemy, especially common among the more frequent words.

The availability of large parallel corpora, such as the "original" and the Simple English Wikipedia, has made recent approaches to automatic text simplification more data-driven. Biran et al. [19] apply an unsupervised method for learning pairs of complex and simple synonyms from a corpus of texts from the original Wikipedia and Simple English Wikipedia. Their approach is called context-aware because they calculate cosine similarity between the given context of a lexical item and the context vector of that item from a trained model. Recently, text simplification has been likened to machine translation, and techniques traditionally used in the latter have been exploited for the purpose of developing novel automatic text simplification systems [20], [21].

3 Methodology

Our methodology consists of: (1) an analysis of a parallel corpus of original and manually simplified news articles, aimed at extracting types of simplification operations to be automated; (2) building our system accordingly; and (3) evaluating the output of the automatic simplification, with regards to the grade of simplification, the grammaticality of the output, and the preservation of meaning in the simplification process.

3.1 Corpus Analysis

We have compiled a corpus of 200 original and manually simplified news articles in Spanish, provided by the Spanish news agency Servimedia[3]. Simplifications have been applied by trained human editors, familiar with the particular needs of our target user (a person with cognitive disabilities) and following a series of easy-to-read guidelines suggested by Anula [22]. We examine the said corpus in order to target different types of simplification operations and, subsequently, prepare their possible computational implementation.

The simplification changes observed in the corpus can largely be grouped as follows:

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] `http://www.servimedia.es/`

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.

Even though we have explored the possibility of automating all four strategies employed by human editors, we have so far implemented the first two: (1) a syntactic simplification module, which conducts a series of transformations at the sentence level, based on operations observed in the parallel corpus; (2) lexical rule-based transformations, which transform certain phrases and expressions that cannot be simplified through the traditional synonym substitution, such as numerical expressions or ethnic adjectives. These operations are also a subset of operations applied by human editors when building the parallel corpus. We are also working on developing a second component of the lexical module, one based on synonym substitution, in which we employ a word vector model to find possible substitutes for difficult original words, and we compute the difficulty (or simplicity) of a word based on its frequency and length. Even though this is intended as a significant component of our system, this module is currently under development and will not be discussed in further detail in this paper.

3.2 Syntactic Simplification Module

We developed a rule-based system for syntactic simplification [23] which is dedicated to several types of sentence splitting operations. These operations turn subordinate and coordinate structures, such as relative clauses, gerundive constructions and VP coordinations into separate sentences, producing shorter and syntactically less complex outputs. The module operates on syntactic dependency trees and tree manipulation is modelled as graph transduction. The following pair of (1) original and (2) simplified sentences are an example of the simplification of a participle construction.[4]

1. The participants (...) will be presented with a book, edited by the town council (...)
2. The participants (...) will be presented with a book. This book is edited by the town council (...)

The grammar comprises five groups of rules, which are dedicated to different syntactic target phenomena. The grammar was previously evaluated, looking at correct rule applications, but so far it has not been evaluated for its contribution to the simplicity of its output. The evaluation in Section 4 includes this second aspect of evaluation.

[4] All examples in the paper are translated into English so as to make it more legible.

3.3 Lexical Rule-Based Transformations

Corpus analysis has revealed that human editors pay special attention to certain types of expressions and that they consistently apply simlification operations to them. Although the operations are applied at the lexical level of the text, synonym substitution is not sufficient in these cases, since the simplification strategies we have observed are somewhat more complex. After carefully examining all such cases, we eventually prepared the computational treatment of the following expressions:

1. **Numerical Expressions (NumExp).** We here define numerical expressions as consisting of a number, expressed either in figures or in letters (*in* [*2009*]) and additional elements, such as modifiers ([*around*] *370,000 children*), measurements (*120,000* [*square kilometres*]), quantified objects (*almost 700* [*crimes*]), etc. We treat nine different types of numerical expressions. Simplification operations include rounding, insertion of modifiers to account for the loss of precision, eliminating optional elements in dates, etc. Detailed description of all simplification operations is beyond the scope of this paper (see [24] for details), though some have been illustrated in Table 1, together with the nine types of Numerical Expressions we treat.

Table 1. Types of NumExp and examples of original and simplified expressions

Type of NumExp	Example Orig.	Example Simpl.
General quantities	*451 attacks*	*almost 500 attacks*
Decimal numbers and fractions	*1.5 million Pakistanis*	*almost 2 million Pakistanis*
Monetary expressions	*1,400 euros*	*more than 1,000 euros*
Percentages	*13.4% of the doctors*	*more than 13% of the doctors*
Dates	*from the 1st of February of 2011*	*from 2011*
Years	*2010*	*the year 2010*
Numbers in letters	*nine million*	*9 million*
Decades	*two decades*	*20 years*
Centuries	*four centuries*	*400 years*

2. **Parenthetical Expressions.** Any information contained in parentheses is eliminated, as it is seen as additional content not essential for the core message of the text. The following sentence is an illustration from our corpus, where the eliminated content is in boldface:
   'Ana María Matute had previously won the National Award for Children's Literature **("Just a bare foot", 1987)** and the Spanish Literature National Award **(2007)**.'
3. **Ethnic Adjectives.** We have observed that ethnic adjectives, such as *Tunisian*, have been substituted with the construction [from/of + <ORIGIN>] rather consistently in our corpus. So, for example, *the Tunisian authorities* has been transformed into *the authorities of Tunisia*. The same is true of nominalisations of these adjectives, where the combination of the definite article and the adjective, e.g. *the Pakistanis*, is substituted with the construction [person from/of + <ORIGIN>], e.g. *the people from/of Pakistan*.

4. **Reporting Verbs.** The various reporting verbs found in the original texts of our corpus have been repeatedly substituted with *decir* (*say*), which is perceived as the simplest option. Such decision is in accordance with the WCAG guidelines that indicate that one and the same term should be consistently used to express the same concept [8]. In order to apply this transformation, we have compiled a list of 32 reporting verbs, based on our corpus and using the Web as a resource. Although this list is by no means an exhaustive list of reporting verbs in Spanish, it serves the purpose better than looking up synonyms in a dictionary. In the Spanish OpenThesaurus dictionary[5], which we use for the development of the module based on synonym substitution, only a third of the verbs from our list is synonymous with *decir*.

These transformations have been implemented with JAPE rules [25], but given space constraints, we here cannot provide a full account of the implementation procedure.

3.4 Experimental Setting

We evaluate both the different components of our system and the system as a whole. The aim of the evaluation is to test (1) the degree of the simplification of the system and its components; and (2) the grammaticality of the output and the preservation of meaning with respect to the original. To achieve the former, we use a set of Spanish readability formulae, and we simultaneously carry out evaluation with human annotators, who rate the degree of grammaticality and meaning preservation in a Likert-scale type of questionnaire[6].

The first evaluation step consists of applying a series of readability formulae for Spanish [22], [26] to the original and simplified texts. The readability formulae intend to capture complexity at the syntactic and lexical levels and are presented in Table 2 (where N = number, w = words, s = sentences, cs = complex sentences[7], dcw = different content words, lfw = low-frequency words, cw = content words, dw = different words, rw = rare words, NumExp = numerical expressions, punct = punctuation marks, and char = characters). It is important to point out that, following Anula [22], we consider as *low frequency words* those words whose frequency rank in the Referential Corpus of Contemporary Spanish[8] is lower than 1,000. Similarly, *rare words* are, according to Spaulding [26], the words that do not appear on the list of 1,500 most commonly used Spanish words. Both lists 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).

The decision about the grammaticality of the output of our system and the meaning preservation in the process of simplification was entrusted to a group of

[5] `http://openthes-es.berlios.de/`

[6] `http://nil.fdi.ucm.es/surveysimp`

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

[8] `http://corpus.rae.es/lfrecuencias.html`

Table 2. Readability formulae applied to the data sets

Formula	Calculation
Average Sentence Length (ASL)	N(w)/N(s)
Index of Complex Sentences (ICS)	N(cs)/N(s)
Sentence Complexity Index (SCI)	(ASL+ICS)/2
Lexical Density Index (LDI)	N(dcw)/N(s)
Index of Low-Frequency Words (ILFW)	(N(lfw)/N(cw))*100
Lexical Complexity (LC)	(LDI+ILFW)/2
Spaulding Density (SD)	N(w)/N(rw)
Spaulding Spanish Readability (SSR)	1.609*ASL+331.8*SD+22.0
Average Word Length (AWL)	N(char)/N(w)
Number of NumExp (NUM)	N(NumExp)
Number of punctuation marks (PUNC)	N(punct)

25 human annotators. They were presented with a questionnaire consisting of 38 pairs of original (O) sentences taken from the corpus of 100 texts used to test the formulae (see Section 4.1), and their simplified (S) equivalents obtained by our system. Every O-S pair contained at least one syntactic and one lexical change. The order of O and S sentences in the 38 pairs was alternated randomly. For every pair of sentences, three questions had to be answered choosing the degree of agreement on a scale from 1 (completely disagree) to 5 (completely agree): (1) *Paragraph A is grammatical*; (2) *Paragraph B is grammatical*; (3) *Paragraphs A and B have the same meaning*[9]. All annotators were native speakers of Spanish and did not include the authors of this paper. Inter-annotator agreement was not calculated, given the elevated number of annotators, and a wide range of options to choose from when grading (five-point scale).

4 Results and Discussion

The results of the two evaluation experiments are discussed separately in the sections that follow, upon which a joint conclusion is presented in Section 5.

4.1 Evaluating the Degree of Simplification

In the first instance, we applied the formulae to the pairs of original and manually simplified texts in our corpus (see Section 3.1), in order to test whether the formulae are a good indicator of the degree of simplification. The results of this experiment are presented in Table 3 (where higher values indicate higher complexity, and the individual formulae that combine into a single complexity index are left out). Differences between all features were reported to be statistically significant at a 0.001 level of significance (paired t-test implemented in SPSS).

After we confirmed the validity of all formulae as indicators of text complexity, the formulae were applied to 100 randomly chosen news articles from the

[9] We used the word "paragraph"since some original sentences were transformed into two simplified ones, and these could not be called a sentence.

Table 3. Formulae applied to the original and manually simplified texts

	LC	SSR	AWL	ASL	SCI	PUNC	NUM
Original	9.71	184.20	4.97	33.42	17.09	12.90	5.20
Simplified	5.28	123.82	4.75	13.69	7.14	1.61	1.80
Rel. diff.	-46.25%	-32.60%	-4.27%	-57.15%	-57.88%	-46.95%	-87.97%

categories of national news, international news, culture, and society, which had been simplified in three stages:

- applying only lexical rule-based transformations (*Lexical*);
- applying only syntactic simplification (*Syntactic*);
- applying both components of our system (*Both*).

We thus obtain three different data sets to be evaluated in comparison with the original texts (Table 4). Differences between the outputs of automatic simplification systems and the original texts which are statistically significant at a 0.001 level of significance (paired t-test implemented in SPSS) are shown in bold. Those marked with an '*' are statistically significant at a 0.002 level of significance, which is still a reasonable result. One important observation is that both original sets (Table 3 and Table 4) achieve practically identical scores on all formulae, meaning that the 100 texts used to test the system are structurally close to the 200 texts simplified manually and used to test the formulae. We can, therefore, expect the selected formulae to be a reliable indicator of complexity of the output produced by our system.

Table 4. Comparison of the original texts and the three simplified text sets

Corpus	LC	SSR	AWL	ASL	SCI	PUNC	NUM
Original	10.10	182.21	4.93	33.48	17.14	13.92	6.41
Lexical	10.08	**174.85**	**4.81**	33.65	17.22	**10.18**	5.73*
Syntactic	**9.92**	**174.40**	4.94	**28.15**	**14.43**	**13.50**	6.41
Both	9.90	**167.21**	**4.82**	**28.36**	**14.54**	**10.64**	5.73*

Averaged relative differences between the corresponding text pairs are given in Table 5. Two general conclusions can be made: (1) both the syntactic simplification and the lexical transformations generally produce simpler output with respect to the original; (2) the combination of the two simplfication processes generally produces a simpler output than either one individually. We have to acknowledge the considerable distance between the relative differences of automatically simplified texts and the ones simplified manually. This, however, is to a large extent due to the fact that manual simplification employs summarisation and paraphrases as most common simplification operations (44%), which results in the loss of a large number of structural elements of the original, among them the punctuation marks and numerical expressions taken into account by

the formulae. Our system in its current state does not perform comparable content reduction. We have previously investigated the possibility of using some summarisation techniques for the purpose of text simplification [27], and intend to accordingly expand the system in the future. However, it is important to point out that manual transformations applied to the original texts in our corpus are often highly idiosyncratic and dependent on world knowledge, and, as a result, it would be difficult to expect to achieve the same grade of simplification automatically.

Table 5. Averaged relative differences between the corresponding text pairs

Comparing	LC	SSR	AWL	ASL	SCI	PUNC	NUM
Original vs. Lexical	+1.31%	-3.97%	-2.55%	+0.65%	+0.66%	-25.22%	-6.66%
Original vs. Syntactic	-1.94%	-4.25%	+0.16%	-14.97%	-15.08%	-2.54%	0
Original vs. Both	-0.36%	-8.13%	-2.27%	-14.22%	-14.34%	-19.86%	-6.66%

4.2 Evaluating Grammaticality and Meaning Preservation

The obtained results were grouped in such a way so as to measure: (1) the annotators' attitude towards the grammaticality of original sentences; (2) the annotators' attitude towards the grammaticality of simplified sentences; and (3) the annotators' attitude towards the differences in meaning between O and S sentences. For each of the sets we measured the average, mean and median value, as indicators of central tendency, and frequency distribution, as an indicator of variability [28]. Table 6 contains the said data. We combined the two lower scores (1-2) into one, to indicate a generally negative attitude towards the grammaticality/meaning, the higher two scores (4-5) into the one indicating a generally positive attitude towards grammaticality/meaning, while the central score (3) represents a neutral attitude.

Table 6. Grammaticality and meaning preservation – central tendency and variability

Measure	Gramm. of O	Gramm. of S	Meaning
Average	4.60	3.58	3.83
Mode	5	4	4
Median	5	5	5
1	2.00%	10.53%	7.47%
2	2.63%	15.26%	10.74%
3	5.47%	16.53%	14.11%
4	13.26%	21.37%	26.53%
5	76.63%	36.32%	41.16%
Negative	4.63%	25.79%	18.21%
Neutral	5.47%	16.53%	14.11%
Positive	89.89%	57.69%	67.69%

Even though the grammaticality of original texts was, expectedly, rated more positively than that of their simplified equivalents (though not at the expected rate of 100%), the latter were also rated rather positively on the whole (average score for the entire set of simplified sentences being in the neutral category). The sentences that received individual average score lower than 3 (i.e. the grammatiacality of these sentences was generally negatively rated), contained 18 grammatical errors, ten resulting from poor syntactic simplification and the remaining eight from bad application of lexical transformation rules. The most recurrent grammatical error was incorrect treatment of different types of coordinate structures, such as coordination of relative clauses. The following pair of sentences is an illustration, with the coordination in question highlighted:

1. The Defence Minister announced that the museum (. . .) is going to achieve wider presence *in Spain and outside our borders* establishing itself as (. . .)
2. The Defence Minister said that the museum (. . .) is going to achieve wider presence in Spain. Outside our borders establishing itself as (. . .)

We found that one third of the errors were traceable to previous parsing errors. Correcting this bad input is beyond the scope of our system. Another third of the errors were attributed to slight errors of the grammar which can be reliably remedied with minor changes in the rules. The remaining errors were related to more complicated syntactic phenomena, which could, in principle, be treated by syntactic rules, but which would require more extensive grammar engineering.

As for the lexical errors, all but one resulted from poor inclusion of the output structure into the existing context. When rounding numbers and using modifiers to account for the loss of precision, we sometimes obtain an ungrammatical combination consisting of a determiner and an adverb, as in *another almost 30 houses*. Given that the majority of numerical expressions from the 100 text set (see Section 3.4) are accompanied by some kind of determiner, restricting the application of the rule to cases other than these would result in considerable drop in recall. What could be done is round the number without the use of modifier, since the loss of precision in meaning is seen as less problematic for our target user than is the actual complexity of the content (see Section 3.1). What is significant is that these two types of errors account for 80% of the S sentences with poor grammaticality, and addressing these two issues in the future should considerably improve the performance of the system.

Meaning preservation was quite positively rated, with the annotators stating that the meaning of the two sentences in the O-S pair was the same in almost 70% of the cases. Only three pairs of sentences were rated negatively (1 or 2). In all three cases, the distortion of meaning is due to syntactic simplification errors, similar to the one previously discussed. The said syntactic errors in combination with the previously mentioned lexical error, account for 60% of the pairs rated neutrally. Therefore, meaning preservation is seen as directly dependent on grammaticality, and the latter is perceived as more important than the loss of precision, even for users without cognitive disabilities (i.e. the participants in the questionnaire). With that in mind, future fine-tuning of certain elements of

our system, such as the aforementioned rounding of numerical expressions, seems like a feasible task, and one to favourably affect overall system performance.

5 Conclusions and Future Work

In this paper we presented two components of an automatic text simplification system for Spanish, and we evaluated them from two perspectives: (1) employing seven readability measures developed for Spanish, we tested the degree of the simplification of our system and its components; (2) in a Likert-scale type of questionnaire, we asked 25 human annotators to rate the grammaticality of the automatic simplification output and the grade to which meaning was preserved in the process.

Our results indicate that both components of our system (syntactic simplification module and rule-based lexical transformations) produce simpler output compared to the original, and that the combination of the two achieves a higher degree of simplification than either of the elements individually. Our system does not reach the simplification degree of manual transformations, but this is largely due to the fact that summarisation and paraphrases are two most commonly applied techniques in the process of manual simplification (they account for as much as 44% of all manual transformations), and as a result, a significant portion of the original content is eliminated. Given that easy-to-read guidelines for people with cognitive disabilities indicate that complexity reduction has preference over the preservation of informational precision, we intend to incorporate a summarisation component into future versions of our system, with the aim of increasing the degree of simplification.

As for linguistic accuracy of the output, our system was rather positively rated by the annotators, 60% of whom considered the simplified sentences to be grammatical, while around 70% of them agreed on the fact that the meaning was preserved reasonably well in the process of simplification. The qualitative analysis of the results revealed that most common errors that result in poor grammaticality of the output were bad treatment of coordinate structures in the syntactic simplification stage, and infelicitous treatment of context when applying lexical transformations. Meaning was seen as directly dependent on the grammaticality of the output, so addressing the two previously mentioned aspects of our system components in the future, should positively influence its overall performance. Nevertheless, the problems resulting from parsing errors remain out of our control for the time being.

Acknowledgments. We present this work as part of a project entitled Simplext: An automatic system for text simplification, with file number TSI-020302-2010-84[10] and partially supported by the Spanish Ministry of Economy and Competitiveness (Project Number TIN2012-38584-C06-03). We are also grateful to the fellowship RYC-2009-04291 from *Programa Ramón y Cajal 2009, Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación,*

[10] `http://www.simplext.es`

Desarrollo e Innovación, Spain. Part of this research is funded by the Spanish Ministry of Education and Science (TIN2009-14659-C03-01 Project), *Universidad Complutense de Madrid*, Banco Santander Central Hispano (GR58/08 Research Group Grant), and the FPI grant program.

References

1. Medero, J., Ostendorf, M.: Identifying Targets for Syntactic Simplification. In: Proceedings of Speech and Language Technology in Education Workshop (2011)
2. Aluísio, S.M., Specia, L., Pardo, T.A.S., Maziero, E., De Mattor Fortes, R.P.: Towards Brazilian Portuguese Automatic Text Simplification systems. In: Proceedings of the ACM Symposium on Document Engineering (2008)
3. Inui, K., Fujita, A., Takahashi, T., Iida, R., Iwakura, T.: Text Simplification for Reading Assistance: A Project Note. In: Proceedings of the 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications, IWP (2003)
4. Aranzabe, M.J., Díaz De Ilarraza, A., González, I.: First Approach to Automatic Text Simplification in Basque. In: Proceedings of the First Natural Language Processing for Improving Textual Accessibility Workshop, NLP4ITA (2012)
5. Rybing, J., Smithr, C., Silvervarg, A.: Towards a Rule Based System for Automatic Simplification of Texts. In: The Third Swedish Language Technology Conference (2010)
6. Saggion, H., Gómez Martínez, E., Etayo, E., Anula, A., Bourg, L.: Text Simplification in Simplext: Making Text More Accessible. In: Revista de la Sociedad Española Para el Procesamiento del Lenguaje Natural (2011)
7. Freyhoff, G., Hess, G., Kerr, L., Menzel, E., Tronbacke, B., Van Der Veken, K.: Make it Simple, European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability; for authors, editors, information providers, translators and other interested persons (1998)
8. Cooper, M., Reid, L., Vanderheiden, G., Caldwell, B.: Understanding WCAG 2.0. A guide to understanding and implementing Web Content Accessibility Guidelines 2.0. In: World Wide Web Consortium, W3C (2010)
9. Bott, S., Rello, L., Drndarević, B., Saggion, H.: Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In: Proceedings of Coling 2012: The 24th International Conference on Computational Linguistics (2012)
10. Chandrasekar, R., Doran, D., Srinivas, B.: Motivations and Methods for Text Simplification. In: COLING, pp. 1041–1044 (1996)
11. Burstein, J., Shore, J., Sabatini, J., Lee, Y.W., Ventura, M.: The Automated Text Adaptation Tool. In: HLT-NAACL (Demonstrations), pp. 3–4 (2007)
12. Devlin, S., Unthank, G.: Helping aphasic people process online information. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (2006)
13. Specia, L.: Translating from complex to simplified sentences. In: Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language, Berlin, Heidelberg, pp. 30–39 (2010)
14. Siddharthan, A.: An Architecture for a Text Simplification System. In: Proceedings of the Language Engineering Conference (LEC 2002), 64–71 (2002)
15. Carroll, J., Minnen, G., Canning, Y., Devlin, S., Tait, J.: Practical Simplification of English Newspaper Text to Assist Aphasic Readers. In: Proc. of AAAI 1998 Workshop on Integrating Artificial Intelligence and Assistive Technology (1998)

16. Quinlan, P.: The Oxford Psycholinguistic Database. Oxford University Press (1992)
17. Bautista, S., Gervás, P., Madrid, R.: Feasibility Analysis for SemiAutomatic Conversion of Text to Improve Readability. In: Proceedings of the Second International Conference on Information and Communication Technologies and Accessibility (2009)
18. De Belder, J., Deschacht, K., Moens, M.F.: Lexical simplification. In: Proceedings of the 1st International Conference on Interdisciplinary Research on Technology, Education and Communication (2010)
19. Biran, O., Brody, S., Elhadad, N.: Putting it Simply: a Context-Aware Approach to Lexical Simplificaion. In: Proceedings of the ACL (2011)
20. Zhu, A., Bernhard, D., Gurevych, I.: A Monolingual Tree-based Translation Model for Sentence Simplification. In: Proceedings of The 23rd International Conference on Computational Linguistics, Beijing, China, pp. 1353–1361 (2010)
21. Coster, W., Kauchak, D.: Simple English Wikipedia: a New Text Simplification Task. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (2011)
22. Anula, A.: Tipos de Textos, Complejidad Lingüística y Facilicitación Lectora. In: Actas del Sexto Congreso de Hispanistas de Asia, pp. 45–61 (2007)
23. Bott, S., Saggion, H.: Automatic Simplification of Spanish Text for e-Accessibility. In: Proceedings of the 13th International Conference on Computers Helping People with Special Needs, pp. 54–56 (2012)
24. Bautista, S., Drndarević, B., Hervás, R., Saggion, H., Gervás, P.: Análisis de la Simplificación de Expresiones Numéricas en Español mediante un Estudio Empírico. Linguamática 4 (2012)
25. Maynard, D., Tablan, V., Cunningham, H., Ursu, C., Saggion, H., Bontcheva, K., Wilks, Y.: Architectural Elements of Language Engineering Robustness. Journal of Natural Language Engineering – Special Issue on Robust Methods in Analysis of Natural Language Data 8, 257–274 (2002)
26. Spaulding, S.: A Spanish Readability Formula. Modern Language Journal (1956)
27. Drndarević, B., Saggion, H.: Reducing Text Complexity through Automatic Lexical Simplification: an Empirical Study for Spanish. SEPLN Journal, 13–20 (2012)
28. Boone Jr., H., Boone, D.: Analizing Likert Data. Journal of Extension (2012)