

Text Simplification Tools for Spanish

Stefan Bott, Horacio Saggion and Simon Mille

Universitat Pompeu Fabra
Departament of Information and Communication Technologies
C/ Tanger, 122-140
08018 Barcelona, Spain
{stefan.bott,horacio.saggion,simon.mille}@upf.edu

Abstract

In this paper we describe the development of a text simplification system for Spanish. Text simplification is the adaptation of a text to the special needs of certain groups of readers, such as language learners, people with cognitive difficulties and elderly people, among others. There is a clear need for simplified texts, but manual production and adaptation of existing texts is labour intensive and costly. Automatic simplification is a field which attracts growing attention in Natural Language Processing, but, to the best of our knowledge, there are no simplification tools for Spanish. We present a prototype for automatic simplification, which shows that the most important structural simplification operations can be successfully treated with an approach based on rules which can potentially be improved by statistical methods. For the development of this prototype we carried out a corpus study which aims at identifying the operations a text simplification system needs to carry out in order to produce an output similar to what human editors produce when they simplify texts.

Keywords: Text Simplification, Monolingual Parallel Corpora, Dependency Tree Transduction

1. Introduction

In the last decades the amount of available information has dramatically increased. But while the average reader has benefited from the new communication technologies to a large extent, for people with cognitive disabilities or other comprehension problems the situation is different. For them the real barrier is often not the availability of textual information, but the fact that many of the available texts are still too hard to read and to understand. Most texts are written for a general public with no special needs and the quantity of available simple text, which is specifically addressed to this reader group, is still rather limited. This is especially true in the case of news. News texts are only relevant for a limited time period and it is hard for human editors to keep pace with the constant production of new content and create simplified versions for even a relevant subset of them. Various organizations are dedicated to the production of textual material adapted to the needs of specific user communities. One methodology commonly used by these organizations to produce such adaptable textual material is the easy-to-read method (Petz and Tronbacke, 2008). But its dependence on human expertise and resources seriously limits the amount of simplified content which can be produced based on existing texts. There are also notable efforts to produce encyclopedic material in a simple language, like the Simple English Wikipedia¹, but such texts usually do not cover news and the efforts are necessarily limited to languages which have a large enough user community or enough financial backup to compensate the lack of volunteer work.

Automatic text simplification is a technology to produce simplified texts, aiming at reducing, at least in part, the efforts required by manual simplification. Text simplification can reduce the syntactic, semantic, and lexical complexity

of a given text by producing a quasi-paraphrase which will contain simple sentences, expressed in a common vocabulary. Our research is concerned with the development of a text simplification system for Spanish. Text simplification has been studied for some years in computational linguistics with research undertaken for English, Portuguese, and Japanese, but, to the best of our knowledge there is no research on simplification for Spanish.

As it has happened with other NLP tasks, the first attempts to tackle the problem were rule-based (Chandrasekar et al., 1996; Siddharthan, 2002), while now the focus is gradually shifting to more data driven approaches (Petersen and Ostendorf, 2007). Some researchers pursue a hybrid solution. The PorSimples (Aluísio et al., 2008; Gasperin et al., 2010) project used a methodology where a parallel corpus was created and this corpus was used to train a decision process for simplification based on linguistic features. Siddharthan (2011) compares a rule-based simplification system with a simplification system based on a general purpose generator. Some approaches have concentrated on specific constructions which are especially hard to understand for readers with disabilities (Carroll et al., 1998; Canning et al., 2000), others focused on text simplification as a help for other linguistic tasks such as the simplification of patent texts (Mille and Wanner, 2008; Bouayad-Agha et al., 2009). Recently the availability of larger parallel or quasi-parallel corpora, most notably the combination of the English and the Simple English Wikipedia, has opened up new possibilities for the use of more purely data-driven approaches. Zhu et al. (2010) use a tree-based simplification model which uses techniques from statistical machine translation (SMT) with this data set and Coster and Kauchak (2011) use standard software from the SMT field and apply these to the problem of text simplification. Their evaluation is based on BLEU, a standard SMT metric. Such ways of treating the problem are very attractive, especially because they are in principle

¹<http://simple.wikipedia.org>

language independent, but they depend on a large amount of data.

For the development of a Spanish text simplification system, we opted for a hybrid approach which largely relies on rule-based components, but tries to integrate data driven methods whenever it is possible. The main reason for this is the lack of large parallel data resources for this language. But we also found that there are some phenomena which make text simplification (in general) different from other NLP problems like machine translation (MT). On the one hand, the fact that simplification is a monolingual task seems to make it easier to treat than bilingual translation tasks, but on the other hand problems like sentence splitting require relatively complex copying operations, which do not necessary occur in other tasks and which are not trivial to model as an empirically learnable problem.

This paper describes a Spanish text simplification system which is still under development. We justify the approach we took and also describe some of the problems we encountered. Finally, we report on first results we obtained for a structural simplification module. The rest of the paper is organized as follows: Section 2. describes the data set we use for system development. Section 3. describes the current state of the simplification system we are developing, including an evaluation of the core simplification component and section 4. gives some outlook on the future work we are planing.

2. Data

We are preparing a parallel corpus of 200 news paper articles with their manually simplified counterpart from the topic domains of national news, international news, society and culture. The simplified versions are provided by trained experts. So the quality of the simplified part of the corpus is controlled for and these texts correspond to our specific needs. The texts of the corpus were automatically aligned on the sentence level with a tool we created for this purpose (Bott and Saggi3n, 2011b) and the automatic alignments were then manually reviewed and corrected.

A special trait of the corpus is that not only the same state of affairs is expressed in two different texts but each simplified sentence has a direct correspondence in the original text. There are only two exceptions to this full correspondence: in some cases new material has been added to the simplified texts, such as definitions of difficult words, and in other cases full sentences have been deleted in the simplification process. The corpus is thus fully parallel and not only comparable or quasi parallel, like other text resources which are available for some languages, for example for English.

The size of the corpus is, however, clearly not big enough to make pure machine learning techniques a promising option for text simplification as a global problem. There are no other large parallel text resources for simplified Spanish, either, which could serve as an empirical basis for data-driven methods, such as the Simple English Wikipedia which we mentioned above. We are investigating if some of the sub-problems of text simplification can be solved with data driven techniques, even if the amount of parallel data is not very large. Other sub-problems may not require par-

allel data and can be trained on larger monolingual corpora. For example the application of rules can be restricted by a statistical classifier which identifies "bad" target structures, which should not be manipulated. Another example is content reduction: it is very hard to decide if sentences or parts of them may be deleted only on the basis of syntactic information, while a statistical classifier may carry out this task. As we already mentioned, the lack of parallel data made us opt for the use of hand crafted rules. But in order to take advantage of the empirical data, we first carried out a corpus study in which we tried to classify the editing operations human simplifiers carried out in order to create rules which come as close as possible to human production. This approach also allowed us to identify those operations which can only be performed by humans because they are too complex to be emulated by computers and/or require complex inferences on the basis of context information and world knowledge.

We could find 8 major operation types, which are described in more detail in (Bott and Saggi3n, 2011a). The most frequent operation types were *change* (39.02%), *delete* (24.80%), *insert* (12.60%) and *split* (12.20%) operations. (1)/(2) is a complex example which combines a split operation with several lexical simplifications and a deletion.

- (1) La muestra ofrece al p3blico la oportunidad de acercarse a la fauna, la bot3nica y la cultura de esta inmensa regin selv3tica americana, al tiempo que recuerda las amenazas a que debe enfrentarse.

"The show offers the public the opportunity to get close to the fauna, the plant life and the culture of this immense American jungle region, and at the same time it reminds us of the dangers it is exposed to."

- (2) La exposici3n nos muestra la cultura de esta gran selva americana. Tambi3n nos muestra sus animales y plantas y las amenazas a las que se enfrenta.

"The exhibition shows us the culture of this big american jungle. It also shows us its animals and plants and the dangers they face."

There have been earlier classifications of simplification operations (Chandrasekar et al., 1996; Zhu et al., 2010; Coster and Kauchak, 2011), but we found that for our needs these taxonomies were not detailed enough. Since our automatic simplification approach necessarily involves manual creation of simplification rules, we were interested in a more detailed classification which would allow to create a list of problems which we had to cover. So we further divided the 8 top-level classes of editing operations into more detailed subclasses, according to the linguistic level at which they applied and the constructions they target. The annotation scheme was also orthogonal to prescriptive or suggestive simplification guidelines for human editors that carry out text simplification, for example the easy-to-read method mentioned above. Simplification guidelines are often very vague and leave room for interpretation, so they usually do

not directly predict directly what a human editor will do when she creates a simplification.

3. Automatic text simplification

3.1. Overview

In the creation of a prototype we concentrated on the most productive simplification operations and those which affect text complexity in the largest degree. We also ruled out operation types that were beyond the possible scope of automatic simplification. The kind of inference which is necessary for these operations is often defeasible and computationally too complex to be viable.

The current version of our simplification system concentrates on structural simplification. We are also developing a lexical simplification module for the substitution of words, a statistical module for content reduction, which will be able to delete phrases and clauses and a statistical support system for rule application, which filters out bad application of rules in the structural simplification module. These modules will form part of a future version of the simplification system.

For the representation of syntactic structures we use dependency trees. The trees are produced by the Mate-tools parser (Bohnet, 2009), trained on the ANCORA corpus (Taulé et al., 2008), and the syntactic simplification rules are developed within the MATE framework (Bohnet et al., 2000). MATE is a graph transducer which uses hand written grammars and one of its advantages is that it allows to create rules which operate on structures with an arbitrary embedding depth.

Figure 1 shows the general architecture of the automatic simplification system, including those modules which are still under development (shown in a lighter colour and gray font). The nucleus of the system in its current state is the structural simplification system, implemented as a MATE grammar, which consists of various layers.

Structural simplification is carried out in two steps: first a grammar looks for suitable target structures which could be simplified. Such structures are then marked with an attribute that informs subsequent levels of the grammar. In a second step the structural manipulations are carried out. This can combine deletions, insertions and copying of syntactic nodes or subtrees. The two steps, marking and manipulation, are carried out directly one after the other in the current version of the simplification system. But, as can be seen in figure 1, we plan to integrate two more modules between the two steps. One of the modules aims to statistically filter out bad applications of the syntactic rules. This is necessary because many of the syntactic trees are ambiguous and cannot be disambiguated only on the basis of morphosyntactic information. A particular case of such ambiguity is the distinction between restrictive and non-restrictive relative clauses. Only non-restrictive relative clauses can be turned into separate sentences and the distinction between the two types is usually not marked by syntax in Spanish². We also plan to integrate a statistical

module for content reduction, which decides whether certain syntactic units, such as adverbial phrases and subordinate clauses, can be deleted without harming the general output. This step is similar to sentence compression. Finally, we are also developing a module dedicated to lexical simplification which replaces "difficult" words with easier ones, where difficulty is measured as the frequency of the word in an open domain corpus.

3.2. Structural simplification

Within structural simplification we treat three groups of problems: sentence splitting, lexical substitution of functional multi-word units and the re-ordering of syntactic units.

Lexical substitution is a subtype of the family of *change* operations. It constitutes 17.5% of all edit operations and is the single most frequent specific operation type we could observe. It also has an important influence on the difficulty or ease with which texts can be understood. At the moment we only treat a small subset of these operations, namely those which require the substitution of functional multi-word units. For example, there are many ways to express contrast between clauses, with expressions like *but* (*pero*), *on the contrary* (*por lo contrario*) or *nevertheless* (*sin embargo*). In ordinary texts a rich vocabulary is considered good for stylistic reasons. In simplified text, on the contrary, there is a clear preference for the use of the simplest form, in this case the word *pero* (*but*). Many of these expressions span over more than one word. In some cases a substituted unit must be inserted in a different place of the sentence (as for example in the case intrasentential *sin embargo* (*however*) is substituted by a sentence initial *pero* (*but*)) or the substituted unit is discontinuous (for example the substitution of *tener X en cuenta* (*to keep X in mind*) by *considerar X* (*to consider X*)). Also the substitution of *al tiempo que* (*at the same time*) by *y* (*and*) in example (1)/(2) falls into this category of lexical substitutions. These cases are hard to treat as simple string substitution operations and for this reason we model them as operations on syntactic trees. But the cases we handle in this way are within a limited range and most other cases of lexical simplification will be treated in a dedicated module which is under development (cf. figure 1).

Sentence splitting is the second simplification operation we treat in the structural simplification module. This operation, which constitutes 12.2% of the editing operations we observed (cf. section 2.), reduces the structural complexity of a text in terms of embedding depth and sentence length. We also considered that this operation is typical for the whole group of syntactic manipulations, since it implies a complex manipulation of a syntax tree. Sentence splitting can target various constructions which subordinate clauses to a matrix clause. There are various cases which are covered by our simplification grammar:

- Relative clauses: we distinguish between simple relative clauses which are only introduced by a bare relative pronoun (e.g. *a question which is hard to answer*) and complex relative clauses which are introduced by a preposition and a relative pronoun (e.g. *a question to which there is no answer*)

²In English it is mandatory to place non-restrictive relative clauses between commas, but in Spanish comma-placement is only a stylistic recommendation.

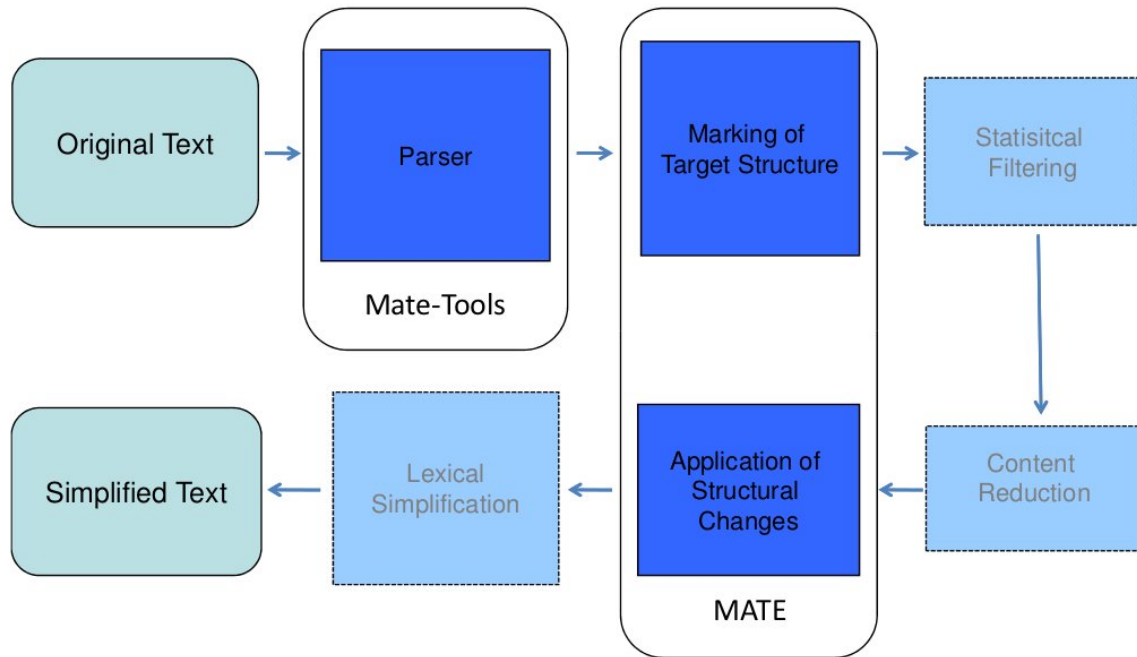


Figure 1: The architecture of the simplification system

- Gerundive constructions and participle constructions (e.g. *the elections scheduled for next November*)
- Coordinations of clauses (e.g. *the problem is difficult and there is probably no right answer*) and verb phrases (e.g. *The problem is difficult and has no easy solution*).
- Coordinations of objects clauses (e.g. *... to get close to the fauna, the plant life and the culture of this immense American jungle region in (1)*)

Finally, we found that there is one particular reordering type which is rather frequent in simplified text and included it in our simplification grammar. We dubbed this operation *quotation inversion*, since it occurs in the context of direct speech. In normal text the pattern $\langle \textit{Quoted_speech}, \textit{said X} \rangle$ is very common and in all of the cases we observed in manually simplified text this pattern was changed to $\langle X \textit{said: Quoted_speech} \rangle$.

A typical case of sentence splitting are relative clauses like the one in (3), and we will illustrate the functioning of the splitting rules with this example. As a matter of fact, this example also contains an object coordination (*have been revised and are being offered ...*), which our grammar can handle, but which we will not discuss here.

- (3) Estos pisos son inventariados (...) y se ofrecen a los jóvenes solicitantes, a los que se acompaña en la visita.
 "These flats have been revised and are being offered to young applicants, which will be accompanied during the visit."

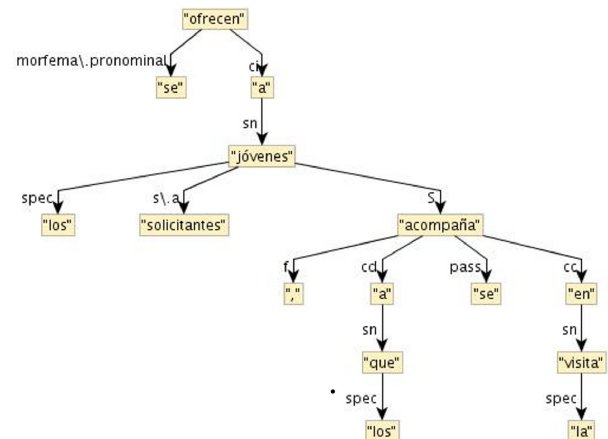


Figure 2: A target structure containing a relative clause

This is a case of a complex relative clause where the relative pronoun is combined with the preposition *a*. The preposition in this case has a case-marking function in Spanish and signals that the phrase *a los que* is the object of the impersonal clause, despite the fact that it occurs in a clause initial position. The sentence can be simplified by expressing the content of the relative clause in a separate sentence, as shown in (4).

- (4) Estos pisos son inventariados (...) y se ofrecen a los jóvenes solicitantes.
 A los jóvenes se acompaña en la visita.
 "These flats have been revised and are being offered to young applicants.
 The young persons will be accompanied during the"

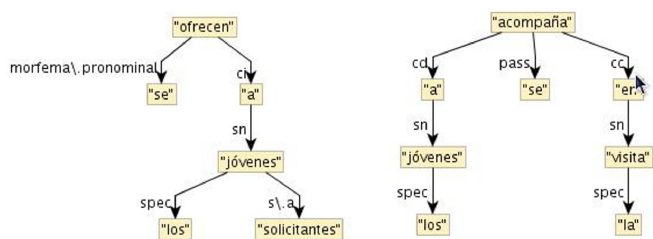


Figure 3: A target structure containing a relative clause

visit.”

The first step towards simplification is the identification of a target structure. In the case of relative clauses we look for a syntax subtree where a relative pronoun is preverbal and directly or indirectly depends on an inflected verb. A target structure for this simplification type is represented by the dependency tree shown in figure 2. The information on linear ordering between words is present in the structures we work with, but it has been omitted here for the sake of readability. In figure 2 the verb *acompaña* (*accompanied*) is the head of the relative clause. It depends on the noun *jovenes* and the label of the dependency is *S* for a sentential subordination. The sequence *a los que*, which contains the preposition and the relative pronoun, in turn, forms a subtree which depends on this verb. This allows the grammar to identify the structure as a target for simplification.

In a next step the relative pronoun is replaced by the full noun form of its antecedent. In this case the noun *jovenes* already carries an article but in many cases the article has to be added, respecting agreement. Finally, the dependency branch between the main verb of the relative clause and the superordinate clause is cut, resulting in two independent clauses. This can be seen in figure 3.

The simplification operations shown here are implemented as syntactic rules within the MATE framework (Bohnet et al., 2000). The syntactic rules in MATE identify a target structure in what is called the left side of a rule and maps them to new nodes and relations on the right side of the same rule. The rule for the manipulation of (3), expressed informally, requires three nodes: a noun, a verb and a relative pronoun. Between noun and verb a sentential (subordinating) relation must hold. Then three corresponding nodes in the output structure are created, corresponding to the three input nodes. The space between the noun and the subordinate verb is marked as cutoff point. Finally, the relation which still holds between matrix clause and the former relative clause is cut in a second step, which requires a further, quite generic, rule. In the example we show here, word order does not interfere with the operation, but in many cases relative clauses occur sentence internally, so they have to be moved out of the matrix clause.

Note that there is also an implicit *change* operation involved in the treatment of this example. The relative pronoun has to be substituted by a lexical NP. In this case the lexical content of the noun can be copied in the place of the former relative pronoun. Here this only involves the copying of lexical content from one syntactic node into another, but in

some cases the presence of an article must be checked, or a tree fragment must be copied from the matrix clause into the place of the relative pronoun.

The current version of the simplification component cover those phenomena which are most frequent and those, which we consider to have the strongest influence on text simplicity. The approach is also extensible to other cases, which represent similar problems, like various types of delete operations or the treatment of nominal appositive constructions.

3.3. Evaluation

We carried out an evaluation of the different rules of the simplification grammar over 886 sentences of our corpus. We considered every case a rule had been applied and decided if the rule had been applied in a correct context and if the rule was able to produce a felicitous output. Minor errors in the grammatical output were ignored when it was apparent that the rule identified the correct target structure and that the grammaticality of the output can be resolved with further fine-tuning of the grammar rules. The results are given in table 1. The precision here is defined as the percentage of correct cases of the application of each rule. For the calculation of recall we hand annotated 262 sentences for structures which contain a target structure which could be simplified.³ The frequency of rule application is given as the percentage of sentences affected by a rule. Whenever we could observe an error, we tried to attribute it to a category of error sources.

Quote inversion is the operation type which the simplification grammar handles most reliably. This is because the syntactic configuration on which it operates can be defined in a very concise way. *Relative clauses splitting* is the most frequent operation. At first sight the application of the corresponding rules does not seem to be very precise. But at the moment we do not perform any filtering. One of the main problems is that 57.58% of the bad rule applications are due to the restrictive nature of the clause. Restrictive clauses usually cannot be turned into individual sentences and, as we explained above, Spanish usually does not represent the distinction orthographically. For this reason we are developing a statistical filter for such cases. We are confident that the combination of the simplification grammar with a statistical filter can improve the overall performance significantly. An additional problem is that in 18.18% of the cases the error occurred because the rule was applied to structures which are structures different from relative clauses, such as verb-dependent object clauses. Also these cases can possibly be filtered out by a statistical classifier. Finally, 16.16% of the errors are caused by idiomatic expressions which can be confounded with relative pronouns. Fortunately these errors can also be detected easily and be eliminated by further grammar development.

In order to decide if a simplification rule should be applied to a target structure we have trained a number of classifiers. The methodology is as follows: the simplification rules that identify target constructions (e.g., relative

³We did not annotate simple and complex relative clauses separately and for this reason we can only list the recall for the set of all relative clause types.

Operation	Precision	Recall	Frequency
Relative Clauses (all types)	39.34%	66.07%	20.65%
Simple Relative Clauses	37.06%	-	19.18%
Complex Relative Clauses	69.23%	-	0.90%
Gerundive Constructions	63.64%	20.59%	2.48%
Quotation Inversion	78.95%	100%	2.14%
Object coordination	42.03%	58.33%	7.79%
VP and clause coordination	64.81%	50%	6.09%

Table 1: Percentage of right rule application and frequency of application per rule type

clauses, participial phrase) were run over a dependency-parsed corpus of Spanish news articles. Sentences in which rules have been fired were identified, collected, and manually annotated with classes “yes” or “no” depending on whether the human annotator considered the actual simplification should take place (between 100 and 150 sentences where annotated depending on the task). A number of features have been implemented, all of them computed based on the context surrounding the target structure to be simplified (e.g., a participle or a relative pronoun). Features include the actual syntactic information of the target structure, parts of speech surrounding the target, and features based on linguistic intuitions such as the depth of the node in the dependency tree, the number of sibling structures, etc. The classifiers are based on support vector machines (SVMs) (Li et al., 2005) distributed with the GATE system (Maynard et al., 2002). Results are still preliminar, but for classification of simple relatives the classifier achieves 86% F-score in cross-validation experiments which we consider a reasonable figure. This improves over a *select-all* baseline F-score of 78%. There is of course still the question of whether the application of the filter after structure identification and marking would actually improve the effectiveness of rule-application, a question we are addressing right now and results of which will be reported elsewhere.

4. Conclusion and Outlook

In this paper we described the development of a text simplification system for Spanish. The development is part of an ongoing research project⁴ and many of the results represent a snapshot of a current state of work in progress. We have argued in favour of a rule based system with support from data-driven techniques. We argued for that mainly because we have not enough data available for Spanish which would allow for the statistical treatment of the global problem. But we also argued that a statistical approach would have to confront problems which are not trivial. These problems mainly are present in the form of complex copy operations which are involved in sentence splitting and such problems would require a careful formulation of the learning task.

We have shown that many structural simplification operations can be handled successfully and that many of the errors of the structural simplification module can be identified in a way which makes their solution feasible. This shows

that there is still much room for improvement without making the required machinery too complex.

There is still a lot of work to be done. In particular, we plan to integrate a statistical filter in the system, which will reduce the number of bad applications of simplification rules. Preliminary experiments have shown that statistical filters of this kind can achieve good results in isolation, so we are confident that they will improve the system as a whole when they are integrated. Content reduction is a problem which requires a similar treatment. We are developing a statistical classifier which decides for different target units (sentences, clauses, adverbial phrases, etc.) whether it can be deleted or not without harming the understandability of the output. Finally, we are trying to combine the structural simplification described here with lexical simplification, which presents a different problem setting, but which is extremely important to make texts easier to read.

5. Acknowledgements

The research described in this paper arises from a Spanish research project called Simplext: An automatic system for text simplification (<http://www.simplext.es>). Simplext is led by Technosite and partially funded by the Ministry of Industry, Tourism and Trade of the Government of Spain, by means of the National Plan of Scientific Research, Development and Technological Innovation (I+D+i), within strategic Action of Telecommunications and Information Society (Avanza Competitiveness, with file number TSI-020302-2010-84). We are grateful to fellowship RYC-2009-04291 from Programa Ramón y Cajal 2009, Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación, Desarrollo e Innovación, Spain.

6. References

- Sandra M. Aluísio, Lucia Specia, Thiago Alexandre Salgueiro Pardo, Erick Galani Maziero, and Renata Pontin de Mattos Fortes. 2008. Towards brazilian portuguese automatic text simplification systems. In *ACM Symposium on Document Engineering*, pages 240–248.
- Bernd Bohnet, Andreas Langjahr, and Leo Wanner. 2000. A development environment for mtt-based sentence generators. *Revista de la Sociedad Española para el Procesoamiento del Lenguaje Natural*.
- Bernd Bohnet. 2009. Efficient parsing of syntactic and semantic dependency structures. In *Proceedings of the Conference on Natural Language Learning (CoNLL)*,

⁴Simplext: <http://www.simplext.es>

- pages 67–72, Boulder, Colorado. Association for Computational Linguistics.
- Stefan Bott and Horacio Saggión. 2011a. Spanish text simplification: An exploratory study. In *XXVII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2011)*, Huevla, Spain.
- Stefan Bott and Horacio Saggión. 2011b. An unsupervised alignment algorithm for text simplification corpus construction. In *Workshop on Monolingual Text-To-Text Generation, co-located with ACL 2011*, Portland, Oregon.
- Nadjet Bouayad-Agha, Gerard Casamayor, Gabriela Ferraro, and Leo Wanner. 2009. Simplification of patent claim sentences for their paraphrasing and summarization. In *FLAIRS Conference*.
- Yvonne Canning, John Tait, Jackie Archibald, and Ros Crawley. 2000. Cohesive generation of syntactically simplified newspaper text. In *TSD*, pages 145–150.
- John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of english newspaper text to assist aphasic readers. In *In Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10.
- Raman Chandrasekar, Christine Doran, and Bangalore Srinivas. 1996. Motivations and methods for text simplification. In *COLING*, pages 1041–1044.
- William Coster and David Kauchak. 2011. Learning to simplify sentences using wikipedia. In *Proceedings of Text-To-Text Generation*, Portland, Oregon. Association for Computational Linguistics.
- Caroline Gasperin, Erick Galani Maziero, and Sandra M. Aluísio. 2010. Challenging choices for text simplification. In *PROPOR*, pages 40–50.
- Yaoyong Li, Katalina Bontcheva, and Hamish Cunningham. 2005. Using Uneven Margins SVM and Perceptron for Information Extraction. In *Proceedings of Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*.
- Diana Maynard, Valentin Tablan, Hamish Cunningham, Cristian Ursu, Horacio Saggion, Katalina Bontcheva, and Yorik Wilks. 2002. Architectural Elements of Language Engineering Robustness. *Journal of Natural Language Engineering – Special Issue on Robust Methods in Analysis of Natural Language Data*, 8(2/3):257–274.
- Simon Mille and Leo Wanner. 2008. Making text resources accessible to the reader: The case of patent claims. Marrakech (Morocco), 05/2008.
- Sarah E. Petersen and Mari Ostendorf. 2007. Text simplification for language learners: a corpus analysis. In *In Proc. of Workshop on Speech and Language Technology for Education*.
- Andrea Petz and Bror Tronbacke. 2008. People with specific learning difficulties: Easy to read and hci. In *IC-CHP*, pages 690–692.
- Advait Siddharthan. 2002. An architecture for a text simplification system. In *In LREC’02: Proceedings of the Language Engineering Conference*, pages 64–71.
- Advait Siddharthan. 2011. Text simplification using typed dependencies: A comparison of the robustness of different generation strategies. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*, pages 2–11, September.
- Mariona Taulé, Maria Antònia Martí, and Marta Recasens. 2008. AnCorà: Multilevel Annotated Corpora for Catalan and Spanish. In *Proceedings of the LREC-2008*, Marrakesh, Morocco.
- Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of The 23rd International Conference on Computational Linguistics*, pages 1353–1361, Beijing, China, Aug.