

UNIVERSITAT AUTÒNOMA DE BARCELONA

Facultat de Filosofia i Lletres
Departament de Filologia Francesa i Romànica



José Camacho Collados

Syntactic Simplification for Machine Translation

A Project submitted as part of a programme of study for the award of
MA Natural Language Processing & Human Language Technology



Supervisors:

Richard Evans

Sandrine Fuentes

Sylviane Cardey-Greenfield

May 2013

Abstract

This research presents a new rule-based Syntactic Simplification system for the Spanish language. The system basically consists of splitting complex sentences into several simple ones preserving its original meaning. It is envisioned to be a preprocessing tool for other Natural Language Processing applications such as Text Summarization, Information Extraction, Parsing and Machine Translation. This research includes a comparison between direct Machine Translation and Machine Translation using the Syntactic Simplification prior to the translation. According to the evaluation, the application of the syntactic rules proposed clearly improves Machine Translation from Spanish to far target languages such as Korean and Chinese.

A parallel corpus of Spanish original and simplified texts has been manually built for this research. The corpus has been validated by native Spanish speakers. The results extracted from the validation suggest that the meaning of the original and simplified sentences is preserved without damaging its grammaticality. In addition, an experiment was carried out among two groups of Spanish learners comparing their reading comprehension of original and simplified texts. Their answers show a null correlation between simplification and text comprehension.

The automatic Syntactic Simplification system has also been intrinsically evaluated on the simplification of complex sentences containing two conjugated verbs. The results are quite promising as more than 90% of the evaluated sentences were positively handled by the system.

Keywords: Text Simplification, Syntactic Simplification, Natural Language Processing, Spanish, Machine Translation.

Resumen

Esta investigación presenta un nuevo sistema de Simplificación Sintáctica para el español. El sistema consiste básicamente en dividir una oración compleja en varias simples preservando el significado original. El sistema se ha desarrollado con el objetivo de servir como preprocesador de otras aplicaciones del Procesamiento del Lenguaje Natural como el Resumen Automático de Textos, la Extracción de Información, el Análisis Sintáctico o la Traducción Automática. Según las evaluaciones llevadas a cabo en esta investigación, la Simplificación Sintáctica propuesta mejora la traducción automática del español a lenguas lejanas como el coreano o el chino.

Asimismo, se ha construido un corpus paralelo español-español simplificado para esta investigación. El corpus ha sido después validado por españoles nativos. Los resultados extraídos de la validación sugieren que el significado de las oraciones originales y las simplificadas es prácticamente idéntico. Además, se ha llevado a cabo un experimento entre dos grupos de estudiantes de español para comparar su comprensión lectora de textos originales y simplificados. Sus respuestas muestran una nula correlación entre la simplificación y la comprensión de textos.

El sistema automático de Simplificación Sintáctica que ha sido desarrollado en esta investigación ha sido también evaluado intrínsecamente en la simplificación de oraciones complejas con dos verbos conjugados. Los resultados son prometedores ya que más del 90% de las oraciones evaluadas fueron positivamente simplificadas por el sistema.

Résumé

Ce travail présente un nouveau système de Simplification Syntactique pour la langue espagnole. Le système, basé en règles, essaye de diviser des phrases complexes en d'autres plus simples mais en conservant le sens de la phrase originale. Le système est développé en tant qu'outil de prétraitement pour d'autres applications du Traitement Automatique de Langues comme le résumé automatique de textes, l'extraction d'information, l'analyse syntactique ou la traduction automatique. Les résultats montrent que la Simplification Syntactique proposée améliore la traduction automatique de l'espagnol vers des langues très éloignées comme le coréen ou le chinois.

Un corpus parallèle simplifié espagnol-espagnol a été construit pour cette recherche. Ce corpus a été validé pour des espagnols natifs. Selon les résultats obtenus à partir de cette validation, la signification des phrases originales est préservée après la simplification. Cette recherche a également envisagé une étude avec des étudiants d'espagnol pour comparer la compréhension des textes originaux et simplifiés. Les réponses des étudiants présentent une nulle corrélation entre la simplification et la compréhension.

Un système automatique de Simplification Syntactique a été aussi développé et évalué concernant des phrases complexes de deux verbes conjugués. Les résultats sont encourageants puisque plus de 90% des phrases évaluées ont été simplifiées d'une façon positive par le système.

Acknowledgements

Firstly I would like to give a big thank you to Yunseo and Ángela, whose help and support was really important to me. Without them, this work would not have been possible.

Special thanks to the translators Prof. Zheng Quan Zhang and Mi Kyung Lee, whose generous and perfect translations from Spanish to Chinese and Korean were essential in the whole evaluation process. I would also to thank all the other Chinese and Korean collaborators who helped me with the evaluation.

I am also grateful for the help from the collaborators in the validation of the corpus and from the students and professors of *Facultad de Traducción e Interpretación (Universidad de Granada)* who agreed to collaborate in one of the experiments of the research.

I would like to acknowledge my supervisors: Prof. Sandrine Fuentes from *Universitat Autònoma de Barcelona*; Prof. Richard Evans from the *University of Wolverhampton*; and Prof. Sylviane Cardey-Greenfield from *Université Franche-Comté*. I would also like to thank to all the people who helped me in some way during my stay in the three universities.

This research was partially supported by the European Commission, Education & Training, Erasmus Mundus: EMMC 2008-0083, Erasmus Mundus Masters in NLP & HLT programme.

Last but not least, thank you to Obra Social “La Caixa”, which supported this masters by a “La Caixa” grant for master studies in Spain.

Contents

Abstract.....	iii
Acknowledgements.....	vi
List of Tables	xi
List of Figures	xii
List of Abbreviations	xiii
1.Introduction.....	1
1.1 Text Simplification for Machine Translation	3
1.2 Plan and goals of the project.....	5
1.3 Research outline.....	6
2. Text Simplification and Machine Translation: Literature review.....	8
2.1 First steps in Text Simplification	9
2.2 Siddharthan's Text Simplification architecture.....	12
2.3 Corpus-based Text Simplification	14
2.4 Spanish Text Simplification.....	17
2.5 Problems in Machine Translation	19
2.6 Controlled Languages and Machine Translation	20
3. Methodology	22
3.1 Syntactic Simplification splitting rules	22
3.2 Regeneration stage	25
3.2.1 Anaphoric references and noun phrase repetition.....	26

3.2.2 Sentence reordering.....	26
4. Corpus	28
4.1 Creation of the parallel corpus.....	28
4.2 Corpus validation	30
5. Automatic Syntactic Simplification system.....	32
5.1 Analysis stage	33
5.2 Transformation stage	34
5.3 Regeneration stage	35
5.3.1 Anaphoric references and noun phrase repetition.....	36
5.3.2 Sentence reordering.....	37
6. Evaluation	39
6.1 Syntactic Simplification for Machine Translation.....	39
6.1.1 Spanish-Korean.....	40
6.1.1 Spanish-Chinese.....	44
6.2 Foreigners learning Spanish.....	48
6.3 Automatic Syntactic Simplification system.....	50
6.3.1 Analysis stage	51
6.3.2 Transformation stage.....	52
6.3.3 Regeneration stage	54
6.3.4 General results	58
7. Conclusions and future work	60
7.1 Main observations.....	60
7.2 Future research.....	61

References.....	64
Appendix.....	68
Appendix A: Original and simplified text in the parallel corpus (Example).....	69
Appendix B: Parallel corpus validation	72
Appendix C: Evaluation of the Syntactic Simplification for Spanish-Korean Machine Translation using Google Translate	76
Appendix D: Evaluation of the Syntactic Simplification for Spanish-Chinese Machine Translation using Google Translate	87
Appendix E: Reading comprehension test for foreigners learning Spanish	97
Appendix F: Evaluation test set for the SS system + Output sentences + Gold Standard.....	107

List of Tables

Table 1: Relation between number of languages and language pairs.....	4
Table 2: Corpus validation (grammaticality and preservation of meaning).....	31
Table 3: Spanish-Korean MT results, grammaticality.....	41
Table 4: Spanish-Korean MT results, preservation of meaning.....	43
Table 5: Spanish-Chinese MT results, grammaticality.....	45
Table 6: Spanish-Chinese MT results, preservation of meaning.....	47
Table 7: Summary of the participants in the reading-comprehension test.....	49
Table 8: Results from the reading-comprehension test.....	49
Table 9: Analysis stage results.....	51
Table 10: Regeneration stage results.....	55
Table 11: Symplification system general results.....	59

List of Figures

Figure 1: Architecture of a Text Simplification system (Siddharthan, 2002: 3).....	12
Figure 2: Transformation stage (Siddharthan, 2002: 5).....	13
Figure 3: Integration of Text Simplification in a Machine Translation system.....	21
Figure 4: Creation of the parallel corpus in text format (Camacho Collados, 2013).....	29
Figure 5: Sentence reordering, relative clause example.....	38
Figure 6: Spanish-Korean MT evaluators' graphic, grammaticality.....	42
Figure 7: Spanish-Korean MT evaluators' graphic, preservation of meaning.....	44
Figure 8: Spanish-Chinese MT evaluators' graphic, grammaticality.....	46
Figure 9: Spanish-Chinese MT evaluators' graphic, preservation of meaning.....	47

List of Abbreviations

- LS: Lexical Simplification
- MT: Machine Translation
- NLP: Natural Language Processing
- PoS: Part of Speech
- SMT: Statistical Machine Translation
- SS: Syntactic Simplification
- TS: Text Simplification

1. Introduction

“Text Simplification is a Natural Language Processing task that aims to rewrite sentences to reduce their syntactic or lexical complexity while preserving their meaning”

(Siddharthan 2002: 1)

Text Simplification (TS) is a Natural Language Processing (NLP) task consisting of reducing the complexity of a sentence in order to make it simpler, but without changing the general meaning of it. Depending on the purpose of the simplification, there are two main types: Lexical Simplification (LS) and Syntactic Simplification (SS). This research focuses on the syntactic level. The syntactic simplifications considered are basically based on splitting a complex sentence into several simpler ones without damaging its original meaning.

The motivations for Text Simplification are basically divided in two main groups. The first one is the one which uses TS in order to make a text accessible for certain people with problems in reading complex structures. Amongst these people, we can find language learners and people suffering some kind of language disability such as aphasia. People suffering this disorder become unable to read and understand some structures and TS attempts to modify the text so that complex structures disappear. This research includes an evaluation of the system regarding this application but, more specifically, for foreigners learning Spanish.

The second main group within TS includes the systems aiming to improve other NLP applications. The rules from this group differ in some way from the rules proposed to make a text more understandable for certain groups of the population. Even though there is an evaluation within language learners included in this research, the SS proposed in this thesis is mainly envisioned to be used as a preprocessing tool for other NLP

applications. The applications which may be improved from this approach are Text Summarization, Information Extraction, parsing or Machine Translation (MT):

-Text Summarization: The idea for the TS in Text Summarization is to reduce the information extracted by each sentence, keeping just the relevant one. Since Text Summarization is sometimes based on the information extracted from different sentences, having the information better divided in sentences might be extremely useful for the task (Chandrasekar et al., 1996; Siddharthan, 2002).

-Information Extraction: As far as Information Extraction is concerned, automatic systems usually work better when the complexity of a sentence is low. They are usually based on some kind of automatic parsers in order to extract the relevant information. Example (1) shows how TS may help store different types of information inside a complex sentence into several simpler sentences.

- (1) a. *Obama, who was born in Hawaii, is currently the US president.* (One sentence, two statements)
- b. – *Obama was born in Hawaii.* (One sentence, one statement)
– *Obama is currently the US president.* (One sentence, one statement)

-Parsing: The goal of TS in parsing is due to the improvement of performance if given a short sentence as input (Siddharthan, 2002). Longer sentences give more cases of ambiguity, so that TS aims to reduce the length of the sentences while preserving the meaning as much as possible.

-Machine Translation: It will be the NLP application evaluated on this research. The possible benefits of using TS for Machine Translation will be explained in more detail in the next Subchapter.

1.1 Text Simplification for Machine Translation

One of the evaluations of the SS system carried out on this research will be extrinsic and it will be done regarding the improvement of some MT systems taking Spanish as source language. The target languages taken into account in this research will be Chinese and Korean. The distant languages (such as Spanish and either Chinese or Korean) and long sentences are nowadays two of the main drawbacks of MT, so it would be quite useful to have a tool improving current MT systems on this matter regardless of the target language. It will be especially relevant for MT systems which are not largely developed, as we can observe in Example (2), where Spanish-Korean Google Translate has been used as MT system.

(2) a. *La policía, que últimamente ha sido duramente criticada, ha detenido a dos secuestradores.*

[Police forces, who have been very criticized lately, have arrested two kidnappers]

Google Translate: 무겁게 최근 비판을 받아왔다 경찰은 두 납치범을 체포했다.
[Somebody (unknown) has been heavily criticized lately police have arrested two kidnappers.]

b. – *La policía últimamente ha sido duramente criticada.*

Google Translate: 경찰은 무겁게 최근 비판을 받아왔다.
[Police have been heavily criticized lately.]

– *La policía ha detenido a dos secuestradores.*

Google Translate: 경찰은 두 납치범을 검거했다. [Police have arrested two kidnappers.]

The output sentence given by direct MT is not quite grammatical in the Korean language and there are several issued damaging the original meaning of the Spanish sentences. The output sentences obtained by using SS prior to the Google Translate are grammatical and accurate translations from the original Spanish sentence.

SS is therefore used in order to solve some of the MT problems handling complex sentences. Firstly, as we have seen in the previous chapter, parsing could be benefited by this SS system. Since parsing is usually an essential tool to preprocess an input sentence, especially for rule-based MT approaches, the MT system would be directly improved. Secondly, Rule-based MT systems tend to backfire when the complexity of a given sentence is quite high and work better when this complexity is low. Therefore a preprocessing tool to reduce the complexity of a sentence such as a SS system could be quite appropriate. The split of the given sentence into several simple ones may lead to a better performance of these MT systems.

As far as Statistical Machine Translation (SMT) is concerned, the length of a sentence is inversely proportional to the accuracy of the system. This is not only due to the increasing number of possible alignments in long sentences, but also to the reliability of the language model, which mainly relies on n-grams found on a monolingual text. The combination of these two factors makes a SMT system be quite more reliable in short and medium size sentences.

Last but not least, it is important to notice that SS is carried out just on the source language no matter the target language. This kind of monolingual approach is quite requested nowadays because of a purely mathematical reason. The total number of language pairs (and consequently the MT systems to be developed) is much larger than the total number of languages in the world, as we can appreciate in the Table 1. The difference is from order n^2 , being n the total number of languages. That is the reason why the MT tasks tend to be divided as much as possible and focus on source and target language separately, i.e., monolingual tasks are computationally less expensive than working on specific MT systems directly.

Table 1: Relation between number of languages and language pairs

Number of languages	N	n
Number of language pairs	$n*(n-1) / 2$	$O(n^2)$
Number of MT systems	$n*(n-1)$	$O(n^2)$

Monolingual tasks other than using a reliable language model could be applied on the target language too. If necessary, specific target language models could be used to improve the fluency of simplified sentences translations, which will be often split sentences from the original one. This could be done by packing the information of the group of output sentences produced by the simplification into a single complex one, taking the inverse way of the SS carried out in the source language.

1.2 Plan and goals of the project

This project aims to create an automatic syntactic simplification system which is envisioned to be used as a preprocessing tool for other NLP applications. The research includes a first version of this automatic system for the Spanish language which applies on sentences with two conjugated verbs at this stage. By conjugated verbs, we are referring to the verbs on the Spanish language which need a subject to be completed.

To fulfill this goal, a small parallel corpus of Spanish original and simplified sentences is also provided. This corpus has been developed by following all the desirable rules for the simplification. The corpus will be useful in order to be used as an analysis tool prior to the creation of an automatic system. It has also been used for several evaluation tasks in this research. Simplified sentences have been used as Gold Standard for the SS system and as reference to evaluate the benefits of this approach.

This project, apart from the main goal of improving a MT system, will also study the effect of simplification on foreigners learning Spanish. This extrinsic evaluation will therefore concern to the readability and possible improvements on the understanding of a text by this part of the population which do not have a perfect domination of the language. An experiment has been carried out with intermediate and advance Spanish

learners in order to find out whether this approach makes the text more accessible to certain ranges of the population or not.

To conclude with, this research will have as main purpose to evaluate the improvements of a MT system by using this SS approach. The way to implement it and some preliminary thoughts about the improvement of the MT system have already been explained in the previous Subchapter. It will also be evaluated taking the Spanish language as the source language and far languages as the target language. The target languages considered in this research will be Korean and Chinese.

1.3 Research outline

This project is organized in seven chapters. Chapter 2 shows previous researches related to Text Simplification and Machine Translation. Subchapter 2.1 includes the first researches about TS and Subchapter 2.2 shows an important state of the art TS system developed by Siddharthan. Some corpus-based TS approaches are commented on whereas Subchapter 2.4 focuses on TS for the Spanish language. Finally, Subchapter 2.5 treats some problems of MT and Subchapter 2.6 shows works improving MT by using controlled languages.

Chapter 3 explains the methodology followed along this research including details of the SS splitting rules considered (Subchapter 3.1) and the final decisions taken at the last stage of the simplification, the regeneration (Subchapter 3.2).

Chapter 4 introduces the Spanish parallel corpus of original and syntactically simplified sentences created by following the rules explained in Chapter 3 (Subchapter 4.1) and its validation (Subchapter 4.2).

Chapter 5 shows the development and architecture of the automatic SS system by applying the rules presented in the methodology. It is divided in three stages: the analysis

stage (Subchapter 5.1), the transformation stage (Subchapter 5.2) and the regeneration stage (Subchapter 5.3).

Chapter 6 includes both the extrinsic and intrinsic evaluations of the system. As for the extrinsic evaluations will be handled on Subchapters 6.1 and 6.2 while the intrinsic evaluation of the system will be done on Subchapter 6.3. Subchapter 6.1 evaluates one of the main applications of this system, Machine Translation. Section 6.1.1 will consider Korean as the target language while Section 6.1.2 will take Chinese as the target language. The other extrinsic evaluation about improving the comprehensibility of a text for foreigners studying Spanish is handled on Subchapter 6.2. Finally, Subchapter 6.3 includes the results concerning the evaluation of the automatic SS system.

To end with, Chapter 7 gives the conclusions of the work (Subchapter 7.1) and addresses some possible directions for further research on the subject (Subchapter 7.2)

2. Text Simplification and Machine Translation: Literature review

Different researches about Text Simplification (TS) have been carried out in recent years. Although it has been referred the potential of using TS for MT purposes (Chandrasekar et al., 1996; Siddharthan, 2002 and 2004), the effectiveness of this approach has not been proved yet. TS has also been used for other Natural Language Processing (NLP) applications such as Information Extraction (Klebanov et al., 2004; Evans, 2011). However, it has been mainly used for human with reading problems or adults learning a language so far (Petersen and Ostendorf, 2007; Aluísio et al., 2008; Specia et al., 2009; Specia, 2010; Bott and Saggion, 2011a, 2011b). Therefore, most of the simplification rules used on these approaches follow this direction.

This chapter starts with the presentation of the first attempts on TS (Subchapter 2.1). This is followed by an important state of the art TS system (Subchapter 2.2) whose architecture has been used as a model for this research. More recently, some machine learning approaches have been developed. These approaches are presented in Subchapter 2.3.

As far as Spanish TS is concerned (Subchapter 2.4), the need of a Spanish parallel corpus with original and simplified sentences has been addressed in previous papers (Bott and Saggion, 2011a, 2011b). Therefore the creation of a manually aligned corpus would be useful not only to build the SS system for this project, but also for future researches on this area. However, like previous attempts on TS, Spanish TS has been focused on improving the readability and understandability of texts but not as a preprocessing tool for other NLP applications.

In order to build a reliable TS system helpful for MT, it would be essential to create some specific simplification rules. That is the main reason why these previous researches have not been able to check whether MT could be improved with TS or not. There are other researches handling the problems of MT for the natural language

(Subchapter 2.5) and proposing some alternatives such as the controlled languages (Subchapter 2.6). Controlled languages are artificially created languages with a low complexity which have been proved to be quite reliable for MT purposes. This could suggest an improvement of MT in natural languages if a proper TS system is used as a preprocessing tool.

2.1 First steps in Text Simplification

Chandrasekar et al. (1996) was the first serious attempt in order to create a general architecture for a TS system. In the research, they comment the usefulness of the TS for a wide variety of applications. The main applications have been explained in Subchapter 1.1. Apart from helping NLP applications such as MT, parsing and Text Summarization, Chandrasekar et al. (1996) introduce the possibility that TS could help Information Retrieval. The reason is quite simple, TS reduces the length of the sentences, so the query could be solved in a more appropriate way, as all the information searched could be found in a simple sentence instead of a long sentence. Not only are NLP applications addressed as possible beneficiaries of the TS, but also TS could make the text clear and easier to understand. This could be particularly useful for manuals when creating an artificial language is not such a difficult task. TS could be used as a preliminary step in order to create this artificial constrained language like controlled languages, which will be deeply studied in Subchapter 2.6. Nevertheless, the research was focused on improving the performance of a parser, which consequently would be beneficial for the other applications.

Regarding this task of improving a parser, the main problem was to reduce the complex syntactic structures that we can find in long sentences. This was achieved by opportunistically splitting the complex sentences into two or more simple ones. To do so, as a first step it was necessary to decide whether the sentence is a candidate to be split or

not. At the beginning of their research there is an example (Chandrasekar et al., 1996: 1) showing how the system works – see Example (3).

(3) a. *The embattled Major government survived a crucial vote on coal pits closure as its last-minute concessions curbed the extent of Tory revolt over an issue that generated unusual heat in the House of Commons and brought the miners to London streets.*

b. *The embattled Major government survived a crucial vote on coal pits closure. Its last-minute concessions curbed the extent of Tory revolt over the coal-mine issue. This issue generated unusual heat in the House of Commons. It also brought the miners to London streets.*

However, there is a lack of a general explanation of how the system decides the simplifications and a general overview of the main SS rules carried out. This is an important lack as we cannot even guess the real power of the system. They state that the starting point is to find a possible articulation point. As an articulation point, punctuation marks, coordinating and subordinating conjunctions, beginning and end of a phrase, and relative pronouns are included. It is also explained that these articulations points could be modify or augmented depending on the domain. Therefore they implicitly claim that their TS system could work more accurately in a domain-specific framework. They apply their system on a corpus of news stories. However, there is not a clear experiment of the system within the domain so it becomes hardly difficult to keep the same direction of research. It would have also been quite appropriate to have studied what to do where a parser fails to recognize these punctuation marks, or even correctly recognizing them, what to do if the punctuation mark is not exactly adjusted to the structure searched. Therefore a general set of patterns and rules should have been commented in order to help further research.

There are two methods carried out in Chandrasekar et al. (1996): the first one is based on a Finite State Grammar (FSG) and the second one is based on syntactic

information given by a dependency representation. FSG is used in order to obtain the chunks of the sentences, which are the basis of this first method. From there, a set of simplification rules based on those chunks was decided. This set of simplification rules are built manually and just general patterns are taken into consideration. This is one of the main problems of this system. As the rules are not deeply developed, the system fails to simplify a wide range of sentences. This system was evaluated in a corpus of news stories of 224 sentences. The evaluation gave as a result several weaknesses of the system. Apart from having a small and not very complex set of rules, they pointed out that it is necessary to rely on a richer syntactic structure in order to obtain better results.

The second method attempts to solve these problems of the first FSG approach. It is based on parsing and particularly, on the dependency information of the sentence. The dependency representation is given by Lexicalized Tree Adjoining Grammar (LTAG). In this case, a much richer syntactical representation is taken into consideration, the dependency trees. The identification of a sentence candidate to be simplified is also taken from the trees, identifying coordination, relative clauses and other parts previously selected.

The evaluation was made in a corpus of news. However, it is not clear how this was carried out. It is claimed that the Dependency-based Model clearly outperforms the FSG. This is due to the much richer syntactic information captured by the first one, while the FSG just relies on chunking and a not quite big set of rules. It would have been interesting to have seen the exact rules to be really able to compare the two methods, as the FSG could perform better if higher developed. In this case, the Dependency-based Model was expected to work better as it takes advantage of a more complete parser. The evaluation gave also promising results for some specific structures. These successful structures are relative clauses and appositives, the latter getting a 100% of accuracy in the corpus. A new direction for further research is suggested, as several problems related to the discourse are addressed.

2.2 Siddharthan's Text Simplification architecture

Siddharthan (2002) proposes a new architecture for a TS system. It follows the work of Chandrasekar et al. (1996) but introducing quite remarkable improvements. He keeps the same line and gives similar applications to his TS system, yet without any external evaluation to verify it. The first difference that we could notice from the previous work is the introduction of a new stage on the TS process: regeneration. This is the main improvement of this research and where Siddharthan tries to focus on. The discourse level problems raised by Chandrasekar et al. (1996) are partially solved by this new stage in the system. The new architecture is then based in three stages: analysis, transformation and regeneration. LS is also included on his system. More specifically, he suggests two ways for LS: paraphrasing and synonym substitution. The paraphrasing involves dictionary lookup and the substitution consists of replacing a word for a simpler synonym. Both could be a useful tool to solve some kind of ambiguity within the phrase. The full architecture is summarized in the Figure 1, taken from his own work.

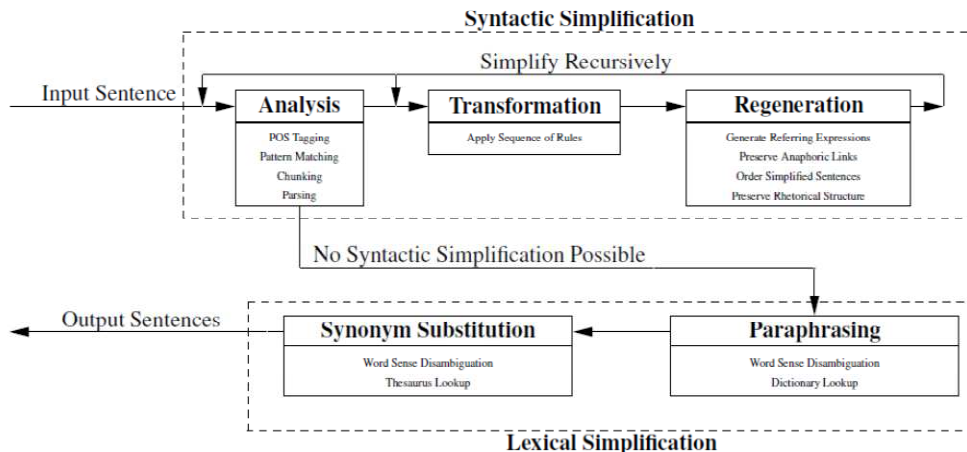


Figure 1: Architecture of a Text Simplification system (Siddharthan, 2002: 3)

As far as SS is concerned, Siddharthan (2002) considered several syntactic phenomena such as relative clauses, adverbial clauses, coordinated clauses, subordinated clauses, correlated clauses, participial phrases, appositive phrases and passive voice. He gave specific examples of how the system works for each of them. From them we can appreciate that he keeps quite similar transformation to the previous Chandrasekar et al. (1996). Afterwards, he analyzes and evaluates each stage separately, giving more emphasis to the new regeneration stage.

The analysis stage relies on shallow parsers such as Part of Speech (PoS) tagging and chunking. This is due to the slow analysis given by full parsers and the mistakes that these also produce in long and complex sentences, which are the ones pretended to be simplified. However, he uses different techniques regarding the syntactic phenomenon present in the sentence. For instance, for relative clauses, the use of WordNet and machine learning are suggested. The evaluation shows that his system can identify correctly relative clauses with accuracy higher than 95%. In addition, the mistakes of the system are correctly detected and it is present in approximately 25% of the cases of the corpus, whereas the remaining 75% sentences get an accuracy of almost 100%.

The transformation stage is just a set of manually created rules which are directly applied as indicated in Figure 2 for the transformation of a relative clause.

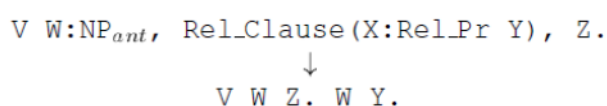


Figure 2: Transformation stage (Siddharthan, 2002: 5)

Finally, he gives a deep explanation about the last stage, regeneration. There are two main reasons why he creates this new stage. The first is purely stylistic, which is essential to keep some readability in the text, and the second one is to preserve some anaphoric links. This second reason is in some cases essential, otherwise the simplified sentence could present a quite different meaning from the original one. It is necessary to

create an anaphora resolution procedure in order to preserve the anaphoric links. The research uses an algorithm which successfully solves many problematic cases. The evaluation gives 81% of correct results, 13% of acceptable results and a 7% of wrong results. These 7% of the mistakes are mainly due to the incorrect analysis in complicate structures. So this lead to a huge improvement by this regeneration stage and directions of how to improve the process are in many cases external to this stage. The order of the simplified sentences is also evaluated giving quite good results. 100% of the cases have an acceptable order. From them, just 8% could have had a better order than the one given by the system.

Although there are different evaluations for each different stage, there is a lack of an evaluation of a general evaluation of the system, and even some extrinsic evaluations. As SS is done as a tool to be used prior to other applications, it would have been quite appropriate to assess the best external applications of this system and an evaluation of these applications.

Siddharthan (2011) compares two different approaches for the regeneration stage using dependency representation given by the Stanford parsers. The goal is to find a suitable approach to deal with one of the main problems in building a TS system: the reliability of parsers. Therefore the robustness of two TS systems towards parsing errors is evaluated. In the end, he concludes that the system becomes more robust if making use of the word order and morphology of the original sentence.

2.3 Corpus-based Text Simplification

Several rule-based systems were created after Siddharthan's work. Some of them already gave some external evaluation depending on the purpose of the system (Petersen and Ostendorf, 2007; Evans, 2011). There are also works addressing the need of a parallel corpus of original and simplified sentences (Petersen and Ostendorf, 2007;

Aluísio et al., 2008; Specia et al., 2009; Specia, 2010). The corpus, apart from potentially being used to develop a *Machine Learning* algorithm, would be useful to carry out a deeper analysis of the task leading to some new ideas or improvements of rule-based TS systems.

Specia et al. (2009) followed this direction and worked on building a Brazilian Portuguese parallel corpus of original and simplified sentences with both lexical and syntactic simplifications. The main goal of the corpus is to help people with low level of literacy or some other cognitive disabilities. In the corpus, they even separate two kinds of simplification changes: natural and strong. This could be particularly useful as the purposes for the TS could be very different, so that depending on the nature of the goal of the system, either strong or natural will be more appropriate than the other. The corpus consists of 104 texts from a corpus of news, which were simplified by a linguist expert in TS. This work includes as well a Simplification Annotation Editor, which could be used in future similar works for other corpus of different languages.

Specia (2010) experimented afterwards with a quite simple corpus-based TS approach. The goal of this system is the same as in Specia et al. (2009), for people with some problems at reading. She basically used a Statistical Machine Translation (SMT) method to deal with the TS problem and check the results. The SMT was carried out without making many changes, so that the approach could be easily improved by adapting the framework to the particular TS problem. The corpus used was taken from two Brazilian newspapers. The corpus was manually annotated by native Brazilian speakers. In such annotation, the kind of the simplification was also annotated on the simplified sentence. The operations taken into consideration were eleven, including LS, SS and non-simplification in some cases. One of the operations was called strong rewriting and it includes any other kind of simplification not covered by any other.

The SMT framework was a current state of the art system. Moses was used in this case. The features used in this new approach included the language model, phrase translation model, word penalty, distortion limit and a model covering the reordering of

the lexical items. The training to decide the weights of the features was given by the Minimum Error Training, as explained in Och (2003).

The evaluation was performed in two different ways: automatic with scores such as BLEU and NIST, and manual evaluation. The automatic evaluation gave a score of BLEU = 0.6075 and NIST = 9.6244. A BLEU score of 0.6 means that the translation quality is not quite bad, so these results, though not perfect, are quite promising for future similar experiments. The manual evaluation was made in just 20 segments from the total test corpus. It was done by comparing the target sentences given by the system with a Gold Standard created by a human. Fluency, adequacy and simplifications performed were the characteristics measured. The scores were 1 to 3, being 1 none of the simplifications was performed and 3 all the simplifications expected were performed. The results were that fluency got a 2.5, adequacy 2.5 and simplification 2.3, which seem not bad at all at a first glance. However, a deeper analysis stated that most of the SS were not well captured, as it hardly ever was performed by the system. Therefore this method should be highly improved in order to perform some SS tasks. The most promising results were given by LS, which let us think that this method could be a better approach for this task. The future work should be focused on building a larger parallel corpus as this usually easily improves the performance of a SMT system. This contained just around 4000 sentences pairs, which is not a high number compared to the normal amount of sentence pairs for a SMT system. Furthermore, it will be necessary to adapt the SMT framework in order to make it more suitable for the TS framework, especially for SS, where it did not perform quite well.

2.4 Spanish Text Simplification

Bott and Saggion (2011a) studies the problem of TS for the Spanish language. The goal of Bott and Saggion (2011a) is to create a text easy to read for people with learning disabilities, so it differs in this point from the main goal of this research. It is a preliminary study where they analyze a corpus of news and the respective simplified one. They address the need of getting a parallel corpus in order to be able to use it for the creation of a reliable TS system.

The corpus was firstly parsed and then aligned (the original text with the simplified one). The alignment was carried out automatically and it attempts to be a tool for future creation of parallel corpus of original and simplified sentences. This could be quite useful, as in Spanish it is possible to find these parallel corpora but it is impossible to find them manually aligned. With the help of this system, obtaining a parallel corpus would be a task much easier than in the present. The method is carried out by using TF*IDF measures and sentence similarity, where the last one was proved to perform better for this particular task.

The operations considered for the research are the usual ones, present in Siddharthan (2002): change (usually for LS), delete, split... In Bott et al. (2011a), it is highlighted the need of keeping a track of the changes made throughout the corpus. This could be automatically, by creating a hand-crafted set of rules or even with an opportunistic combination of both. It is also addressed the need of having a reliable anaphora resolution system, as many of the operations could be done just by recognizing some anaphoric components within the sentence or sentences. They also emphasize on the need of getting a larger parallel corpus in order to apply machine learning techniques, especially for SS. To conclude, it is addressed the need of getting a global evaluation metric for TS, which is actually one of the main problem of this topic. They indicate some ways to do so but indicating the actual difficulty to solve this. This should be one of

the main directions of research of the near future, as it is necessary to keep a track of the improvements in this area.

Bott et al. (2011b) follows their last paper and explain in detail an algorithm to align a parallel corpus of news and their simplified ones at a sentence level. As there is not training data for task (there is no manually aligned parallel corpus), they rely on unsupervised learning. The alignment is done applying a Hidden Markov Model given by two starting points: simplified sentences and original ones usually share similar words and the order does not change in many cases. Using these two intuitions they build the machine learning algorithm to find the alignments within the corpus. The results suggest that this approach has improved previous alignments algorithms at a sentence level, though it is not directly comparable as this algorithm attempts to solve a very specific problem whereas the previous ones had a different goal. The system that they use to compare to is the given in Nelken and Shieber (2006), relying on TF*IDF scores and supervised learning. They took advantage of a training corpus that Bott et al. (2011b) could not get.

However, as explained before, this method may be modified regarding the transformations taken into consideration. The transformation from an original text to a text easy to read as in Bott et al. (2011b) will differ from the SS considered in this research. The difference between an original sentence and a simplified sentence affected by SS for MT will share much more similar words than in Bott et al. (2011b). This will lead to some modifications in the intuitions which will be relevant, even though this algorithm could be used as a starting point.

Drndarević et al. (2013) presents a two-component (syntactic and lexical) automatic text simplification system for the Spanish language in order to make the text easier to read for people with cognitive disabilities. The system managed to get simpler sentences without seriously damaging their grammaticality and preservation of meaning with the original sentences. However, they do not propose any solution to the parsers errors, which caused most of the errors found in the system.

2.5 Problems in Machine Translation

Cardey et al. (2004) addressed the problem for the MT between an Indo-European language such as French and two languages belonging to a far language such as Chinese and Arabic. The MT method considered in the research is transfer, though the conclusions stated here could be extended to other rule-based approaches. A conclusion taken from the article is that semantics of linguistic units is a very special issue which should be treated with a special care. Another main conclusion is that many of the issues concerning a pair of language studied separately for each different language pair, which makes these MT systems really language-independent. This is a problem as many pairs of minority languages will not receive the needed attentions to solve all these kinds of problems. SS could partially solve this problem, as it could be seen as an attempt to reduce the complexity of the source language. Although the complexity from a language pair could differ from another language pair with the same source language, a reduction on the complexity of the source language could always lead to an improvement on the translation quality.

This could be also extensible for SMT. SMT also improves the quality while translating simple sentences (Koehn, 2010). SMT tends to backfire when the source sentence is long and have several verb phrases, no matter the target language. SS consists mainly of reducing the complexity of a complex sentence, usually by splitting the sentence into several simple ones. So an appropriate SS system created with the goal of improving a MT system for a specific source language could be of great help. Considering a source language with a SS system, it would improve the translation quality of both rule-based and statistical MT for no matter what target language.

2.6 Controlled Languages and Machine Translation

The problems of MT often come, apart from the length of the sentence, from the ambiguity that such sentence could have. MT systems should deal with this problem, which is one of the main ones for every system. Even deeply treated, ambiguity is really complicated to be totally solved in order to create a high quality MT system. That is one of the main reasons why controlled languages were developed (Kaji, 1999; Mitamura, 1999; Cardey, 2011).

Controlled languages are artificially created subset of the natural language where the ambiguity is eliminated and complexity is reduced. This obviously improves MT and it is especially useful where used for different target languages. In controlled languages, vocabulary, grammar and even style are restricted. The restriction in vocabulary aims to eliminate the lexical ambiguity within the sentence. The words will no longer be polysemous, as just a meaning from a specific part of speech will be kept. The grammar restrictions have several goals. The first one is to reduce the syntactic ambiguity of the sentence. The relation within the phrase should be clear so some constructions and problematic orders are avoided. Another goal is to keep the simplicity of the sentence. Therefore there should be a limit of words in the sentences and coordinated and subordinated sentences are in some cases reduced. The restrictions on the style are on the same line with the grammar restriction. The punctuation and the structures used should be clear and easier to process for a MT system.

Although controlled languages belong to natural language, it is artificially created. This is one of the main problems of the controlled language. They need to be created and people using this language should know all the restrictions in order to be able to use it. As explained in Mitamura (1999), this is not an easy task because usually there is a big restriction in both grammar and vocabulary, so it should be deeply studied before using it. This is the point where TS could be quite useful. TS treats natural language and aims to reduce the complexity and ambiguity of the sentences. LS could be compared to the

vocabulary restriction of the controlled languages and SS the grammar restriction. On this way, TS could be seen as a preliminary step to the MT for controlled languages. In Figure 3, we can appreciate how TS could be integrated in the MT process as preprocessing tool on the source language.

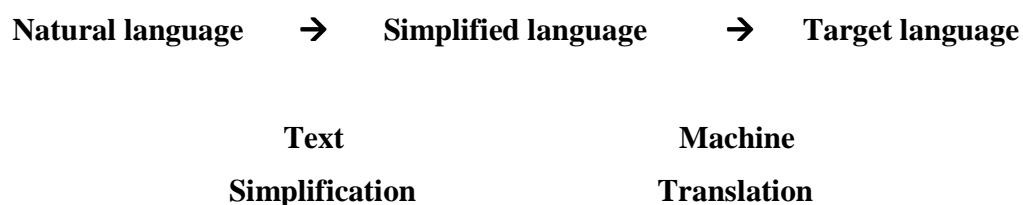


Figure 3: Integration of Text Simplification in a Machine Translation system

As controlled languages have been proved to be a useful approach for MT (especially for systems which include a source language and several target ones), TS could be used as a tool to reach the desirable controlled language. TS could be used to as a preprocessing tool on the whole MT process. Temnikova (2012) shows how the post edition by a human translator becomes an easier task by applying TS prior to the MT. This research will evaluate the output given by the MT directly and compared it with the output given by using SS. It is obviously not an easy task to build a TS which could take natural language as a starting point and give a controlled language as result. Some of the ambiguity and complexity of the source sentence will not be solved as completely as with controlled language but the advantage of starting from natural language could be worth trying.

3. Methodology

As it has been commented in previous chapters, the main goal of this project is the creation of an automatic Syntactic Simplification system to be applied on other NLP applications such as Machine Translation (evaluated in Chapter 6), Information Extraction, Information Retrieval or Parsing. Therefore the rules have been built taking into consideration these two premises: rules easily implementable on an automatic system and useful for these NLP applications. The rules are basically based on splitting complex sentences into several simpler sentences preserving its original meaning. All the rules summarized on this chapter have been decided by following this premise.

This Chapter will therefore explain the basis of this SS approach. Chapter 3.1 presents the general splitting rules considered for this research and Subchapter 3.2 handles the regeneration stage, which is a more complex stage of the syntactic simplification process. The rules presented in this Chapter will be the ones used in Chapter 4 for the creation of a parallel corpus of original and simplified sentences (Camacho Collados, 2013). The automatic SS system presented in Chapter 5 will be also based on these transformations.

3.1 Syntactic simplification splitting rules

To begin with, the first condition to split a sentence is the number of conjugated verbs. The sentences which are to be simplified must have at least two conjugated verbs. However, not all this kind of sentences will be simplified, as there are some kind of structures which do not allow splitting and preservation of meaning simultaneously. It is also important to notice that in most sentences there are different types of structures to be

simplified simultaneously, which makes the task harder. The main structures (one per sentence in the following examples¹) allowing the simplification are summarized below:

➤ **Coordination:** The original sentence is split on the position of the coordination nexus or articulation point as called by Chandrasekar et al. (1996). It is sometimes recommendable to repeat some noun phrases in order to improve the understandability and a best processing by other NLP applications. Example (4) shows an illustrative example of how the simplification is handled on a coordination structure.

- (4) a. *Fishes swim in the sea and butterflies fly in the sky.*
 b. – *Fishes swim in the sea.*
 – *Butterflies fly in the sky.*

➤ **Subordination:** The simplification is done by taking into consideration the subordination type. It is also worth noticing that there certain kinds of subordination which do not allow a syntactic simplification without change of meaning. Therefore every case has been studied separately and some cases of subordination and its simplification are the following: some kinds of relative clauses; cause and effect; indirect speech to direct speech...

• Non-restrictive relative clauses: This kind of structure is split without adding any other element which might cause a mistreatment by other NLP applications. The non-restrictive relative clauses including in the simplification process are the one which are introduced as an explanation of a certain noun phrase. This is easily recognizable by

¹ In this section, several examples are giving in English for a better understanding of the reader when possible. The examples given in English keep the same syntactic phenomena as the original Spanish sentences though. When given in Spanish, the English translation is provided if necessary.

either a human or a machine because it is introduced between commas within the sentence. The relative clause begins with a connector such as *que* or *quien*, which are the other indicators to let us classify the construction as relative subordinate. What can cause some trouble to an automatic SS system is the correct identification of the noun phrase and the ending of the relative clause. We can see a simple and representative example of this type of construction on the Spanish language and its simplification in the following Example (5):

- (5) a. *Juan, que es aún muy joven, consiguió el premio.*
 [Juan, who is still very young, got the prize.]
- b. – *Juan es aún muy joven.*
 [Juan is still very young.]
- *Juan consiguió el premio.*
 [Juan got the prize.]

- Effect: In this type of structure contains the relation cause-effect. The *cause* appears at the first part of the sentence and the *effect* appears afterwards. They are connected by a conjunction which indicates the end of the cause and the beginning of the effect. Therefore the splitting of the sentence will happen exactly at the conjunction's position and another effect connector such as *therefore* in English is introduced at the beginning of the second sentence as we can appreciate in Example (6). In Spanish the connector introduced will be "*Por lo tanto,*" at every *effect* case.

- (6) a. *The cat ate poisoned food, so it died.*
- b. – *The cat ate poisoned food.*
 – *Therefore, the cat died.*

- Cause: The same kind of relation cause-effect appears in this structure. However, the relation appears in the opposite order. The *cause* appears in this case at the second part of the sentence whereas the *effect* appears at first place. They will be connected by a

different kind of subordination causative connector such as *because* on the English language - Example (7) - and *porque* on the Spanish language. The output sentences will include the *effect* in the first position and the *cause* or *reason* in the second sentence.

- (7) a. *Dogs can't fly because they don't have wings.*
b. – *Dogs can't fly.*
– *Reason: They don't have wings.*

- Indirect speech to direct speech: This structure is currently simplified just on very specific cases. More precisely, in the cases where a communication verb such as *decir*, *comunicar* or *explicar* are followed by the relative *que* and a sentence concerning the communication of the speaker. Most of the communication verbs included for this task were extracted from *Diccionario combinatorio del español contemporáneo* (Bosque, 2004). The transformation follows is to introduce “:” between the first sentence and the second sentence, as we can observe in the following Example (8):

- (8) a. *El jugador dijo que el presidente estuvo con el equipo antes del partido.*
[The player said that the president was with the team before the match.]
b. – *El jugador dijo:*
[The player said:]
– *El presidente estuvo con el equipo antes del partido.*
[The president was with the team before the match.]

3.2 Regeneration stage

This is the stage concerning the last part of the simplification process. It was firstly introduced by Siddharthan (2002) in the creation of an automatic TS system. Once the original sentence has been split into several ones, there are still some unresolved

issues such as the simplified sentences order and anaphoric references which need to be tackled in this last stage.

3.2.1 Anaphoric references and noun phrase repetition

Since preserving the meaning of the original sentence is one of the main premises in TS, some noun phrases which should appear on an anaphoric way in the simplified sentences will be repeated instead and the use of pronouns will be considerably reduced. We can appreciate in the Example (6) on page 38 how the noun phrase “the cat” is repeated in both sentences. Although the readability of the sentence could be damaged, this will lead to a better understanding of the simplified sentences with no ambiguity. To improve the readability, these anaphoric references, albeit repeated, will appear as shortened as possible.

This noun phrase repetition is especially important in the sentences with non-restrictive clauses, as the noun phrase where the relative clause is attached must be correctly identified and repeated in the simplified sentence. However, this noun phrase often counts with a long amount of words, so that a reduction of the noun phrase is in those cases necessary (without substituting it by a pronoun).

3.2.2 Sentence reordering

The split phase gives us a set of new sentences with no order as output. Giving them a right order is in some cases essential to fully understand the message of the original sentence. Although most of the sentences keep the order as found in the original sentences, there are some cases where a sentence reordering is needed. This reordering is not a difficult task for a native Spanish but an automatic system might find it extremely

difficult. Therefore an algorithm is also included in Chapter 5 regarding the sentence order of the output sentences.

What is usually natural for a native speaker is to leave the order as it appears on the original sentences. However, there are certain embedded syntactic structures which need to be dealt more carefully. This is usually the case of relative clauses, which are a complement of a noun phrase in the middle inside the main sentence. It is the decision of the native speaker then to give a right order to these sentences. It is not a random decision though, as it is also related to the relative position of the relative clause within the sentence.

4. Corpus

A parallel corpus of Spanish original and simplified sentences is been built for the research. This was done manually by following the simplification rules given in the previous Chapter. Subchapter 4.1 explains in detail the parallel corpus creation process and Subchapter 4.2 shows the validation of the corpus by native Spanish speakers. This validation also concerns the rules considered in Chapter 3, as they are the same applied during the building of the corpus.

This corpus is the result of a long process of selecting the adequate rules for the simplification (Camacho Collados, 2013). It was necessary to decide the set of simplification rules prior to the creation of the final corpus. Once these rules were decided, the corpus was modified according to this set of rules. Since the set of rules is still extendable, the corpus could suffer more modifications in the future. However, this extension will not cause too much problem because of the nature of the simplification rules. They are mainly based on keywords rather than patterns, which makes the automatic search of rules within the corpus quite accessible.

The parallel corpus is envisioned to be used as a tool to create an automatic SS system. It is especially prepared and aligned to use supervised *Machine Learning* techniques. As the corpus is expected to be extended in the future, *Machine Learning* will be encouraged and will possibly improve the SS system. The corpus has also been a useful tool to create the Gold Standard for the evaluation experiments (Chapter 6).

4.1 Creation of the parallel corpus

The corpus chosen as a reference was the AnCora Corpus (Taulé et al., 2008), which consists of Spanish newspaper texts annotated at syntactic and morphological

level. Newspaper texts are quite representative of the natural language and complex enough for the simplification task. Once the corpus was transferred from the original XML format to text format, the sentences from the AnCora Corpus were manually simplified as explained in Chapter 3 in separate text files. On this way a parallel corpus of original and simplified texts was built. As we can see in Fig. 1, the original sentences and the simplified ones are easily aligned, as each original sentence is separated from each other on a new line, which is respected on the simplified part (right part on Figure 4) no matter how many new sentences have been created.

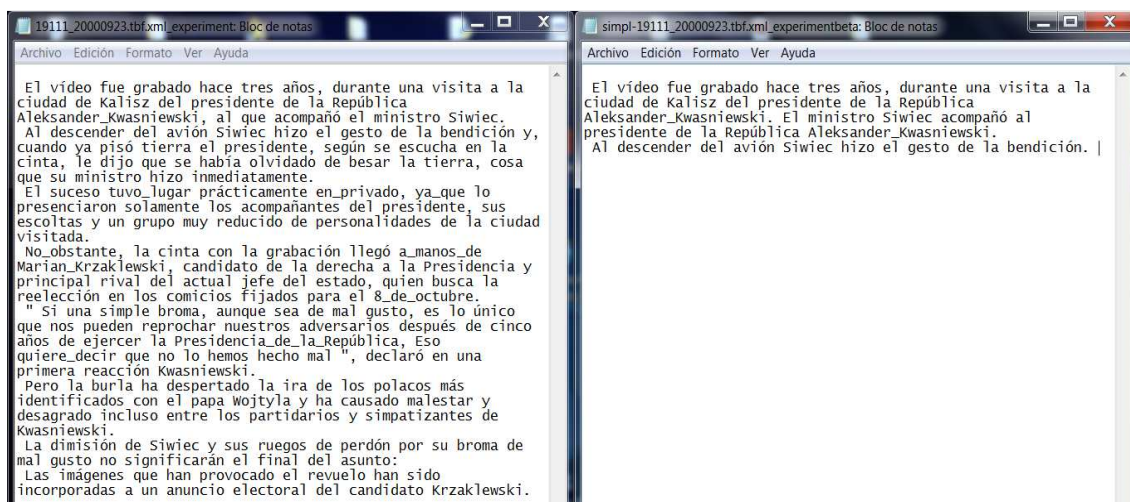


Figure 4: Creation of the parallel corpus in text format (Camacho Collados, 2013)

The corpus currently counts with 3000 original sentences and their respective simplified ones. It is already available for research purposes if required. Appendix A includes an example of an original text and its simplification from the parallel corpus.

4.2 Corpus validation

The validation of the corpus has been done regarding two issues (grammaticality and preservation of meaning), similar to the one used in Drndarević et al. (2013) for the evaluation of their SS system. To carry out this, some preliminary results were obtained from six native Spanish speakers (three of them linguists and three of them holding a non-related university degree). They required to fill three excel sheets (see Appendix B). The first one was to evaluate the grammaticality of the original sentences from AnCora corpus; the second one about the grammaticality of the simplified sentences from AnCora Corpus; and the third one concerning the preservation of meaning of original and simplified sentences. To do so, thirty complex sentences were randomly selected from thirty different texts of the corpus.

All the evaluations were done on a 1-5 scale. For the grammaticality measure, 1 means that sentence is completely a grammatical and the 5 that the sentence is completely grammatical. For the preservation of meaning, 1 means that there is no preservation of meaning at all and 5 that the meaning of the simplified sentence and the original is identical. Table 2 shows the total average for each evaluation and from the results we can appreciate how grammaticality is not damaged on the simplified sentences (4.74 without simplification – 4.66 with simplification). A paired two-sample t-test at the 0.05 level suggests that the difference is not statistically significant ($t(179)=1.513$; $p\text{-value}=0.132$)². There are even a few cases where the simplification actually improves the grammaticality of the original sentences.

² The results from the corpus validation have been analyzed by using the statistical program IBM SPSS Statistics 20. All the statistical tests and graphics throughout this research have been produced by this statistical program.

Table 2: Corpus validation (grammaticality and preservation of meaning)

	Grammaticality Original Sentences	Grammaticality Simplified Sentences	Preservation of meaning
Average	4.74	4.66	4.8
Positive (4-5)	95.6%	97.2%	98.9%
Neutral (3)	3.3%	2.8%	1.1%
Negative (1-2)	1.1%	0%	0%

Considering 4-5 as a positive attitude of the annotator towards the grammaticality /preservation of meaning, 3 a neutral attitude, and 1-2 a negative attitude, table 2 also shows the general results of the annotator’s attitude. The grammaticality and preservation of meaning from simplified sentences get an incredibly positive attitude by the annotator (over 97% in both cases).

Regarding the inter-rater reliability for the preservation of meaning task, a statistical test at a 0.05 level was carried out among the six annotators. The results ($F(5,174)=1.577$, $p\text{-value}= 0.169$) show an inter-annotator agreement quite high.

5. Automatic Syntactic Simplification system

Once the corpus and the rules were clearly defined, a small automatic SS system was developed on the programming language *Python*. Only sentences containing two conjugated verbs are covered so far by the system. Further improvements need to be added to the system in order to be able to tackle sentences with more than two conjugated verbs.

The SS system is built following a three-stage architecture (analysis, transformation and regeneration) following the system developed by Siddharthan (2002). Each stage will be explained and evaluated separately. In general, the system is built in a way that it does not rely neither on any full parsers nor dependency structures, which has been the main origin of errors in previous TS attempts (Siddharthan, 2011; Drndarević et al., 2013). This is also coherent with one of the goals of the system to improve the performance of parsers. It is just based on word searches of the articulation points (Chandrasekar et al., 1996) and in a few particular cases, as we are going to see in following chapters, on shallow PoS taggers and verb lemmatizers. In the regeneration stage, system relies on noun chunking in most cases. Therefore the performance of this automatic system may be generally improved regardless of the performance of automatic taggers.

Another advantage of this approach is the adaptability to other close languages such as French, Italian or Portuguese. The way the system is developed without relying on specific language-related peculiarities allows an implementation for those languages relatively easy. All the syntactic phenomena treated in this research and the subsequent transformations are shared by all these Romance languages.

This Chapter includes the explanations of how the system was built. Subchapter 5.1 focuses on the analysis stage, Subchapter 5.2 shows the rules considered for the transformation stage and Subchapter 5.3 deals with the last phase, the regeneration stage.

5.1 Analysis stage

This stage takes the original sentence as input and decides whether simplifying it or not. It mainly relies on Part of Speech (PoS) tagging, word searches and, in a specific case, lemmatization. PoS taggers are only necessary in order to find the number of verbs in the sentence. As explained before, this research will include the simplification with no more than two conjugated verbs. To be more precise, this research intends to simplify sentences with two conjugated verbs, since sentences with only one conjugated verb will be considered as simples already.

Once this is done, the analysis stage will rely on raw word searches which was proved to be reliable indicators of whether simplifying or not in Medero and Ostendorf (2011). More specifically, conjunctions will usually be the indicator of simplifiable and also non-simplifiable sentences. The conjunctions allow us to detect the structures described in chapter 3 which are to be simplified. On a specific rule (indirect to direct speech), lemma searches would be also necessary in order to find the simplification pattern. In this stage we are not worried yet of where the splitting should be done and how it should be done. All these concerns will be studied in more detail in the following stage (transformation), which appears in the next subchapter.

This stage was based on a decision-tree model taking into consideration conjunctions indicator of possible simplification but also indicators suggesting not to simplify (*cuando, donde...*). This was possible because of the first assumption of only two conjugated verbs per sentence. That is the reason why indicators of no simplification become essential at this stage, especially for the coordinate sentences. Coordinate sentences make use of ambiguous conjunctions such as commas or “y” (and), which are used in many different kind of structures. Therefore identifying correct coordinate sentences is a relatively difficult task (Goldberg, 1999). The decision was to include these coordinators conjunctions at the very bottom of this decision-tree model even after the *no simplification* indicators. This decision may damage the recall of the analysis stage as

certain true coordinate sentences will not be recognized. However, it will surely improve the precision, which is actually most important than the recall for the simplification task.

Concerning the indirect to direct speech rule, a search of a bag of communications verbs is needed so that lemma search is essential. In this case, we use the AnCora XML documents provided by the corpus. These lemmas were semi automatically captured and exported to XML as explained in Taulé et al. (2008).

In the next Chapter, an experiment is carried out in order to find how the stage works in terms of precision and recall.

5.2 Transformation stage

Once the analysis stage has decided to simplify the input sentence, the transformation stage comes into place. At this stage, the system basically chooses a position to split the sentence and introduces new connectors if necessary. This differs a bit from the transformation stage of Siddharthan's text simplification architecture (Siddharthan, 2002), where the analysis stage were the one selecting the boundaries between clauses and the main sentence. This stage does not take into account the relation within the sentences nor the anaphoric references to be considered in the output sentences. This will be considered later on the regeneration stage.

In the particular example of the non-restrictive relative subordinates, the splitting is produced just after the relative *que*, *quien* or *quienes* where the comma is placed. These relatives must be preceded by a noun or a noun followed by an adjective. In the cases where this relative clause is not the end of the sentence, the simplified sentence split will usually end in the following comma of the sentence as we can appreciate in the Example (9). However, there are a few considerations to be taken about this rule. The first is that the main sentence should continue with a conjugated verb and even so, if the comma is followed by nouns, adjectives or prepositions (no more than four in total) and a comma,

the second comma will be taken as split indicator. To do so, PoS tags from the XML document are used.

- (9) a. *Los **jugadores** excluidos, **que** no se comportaron como profesionales, no jugarán más esta temporada.*
[The excluded players, who didn't behave as professional players, will not play anymore this season]
- b. – *** *no se comportaron como profesionales.*
[*** didn't behave as professional players.]
- *** *no jugarán más esta temporada.*
[*** will not play anymore this season]

On the other cases (cause, effect and indirect to direct speech), the sentences are split just once, i.e., the first output sentence appears in the first part of the original sentence and the second one appears in the second and last part of the original sentence. There are not embedded structures like in the sentences including relative clauses structures.

5.3 Regeneration stage

This is the last phase of the SS process. It is also the technically most difficult stage to handle automatically. After the splitting of the sentence on the transformation stage, many assumptions need to be taken in order to make these new simplified sentences readable and understandable. These assumptions are tackled in Chapter 4 but in this Subchapter we will treat them from the automatic point of view.

As in Subchapter 4.2, we are going to divide this task in two subtasks: anaphoric references and noun phrase repetition in Subsection 5.3.1 and sentence reordering in Subsection 5.3.2.

5.3.1 Anaphoric references and noun phrase repetition

This is the most challenging task in the whole SS process. As the original sentence is split, there are some links within the original sentence which may be lost while the simplification. Anaphoric references may appear at the end of the original sentence but may refer to a certain element of the beginning of the sentence and, after SS, this appears at a different sentence. There might be other problems at the discourse level as well. For instance, some anaphoric references at the original sentences referring to some element at a previous sentence may cause trouble to the simplification. However, at this stage this research is going to focus on the problems inside the sentence and not at the discourse level.

The structure with more problems at a practical point of view is again the relative subordinate. It is the one this research has focused on at this stage, as the noun phrase repetition is essential for the well understanding of the output sentences. The main decision is was to consider the closer noun phrase to the conjunction *que or quien*, unless there are other indicators which may lead to take a better decision. To consider noun phrase, no shallow parsers are used, as this caused more trouble from the beginning than deciding the noun phrase by other specific approaches. A noun phrase was considered to be combinations of adjectives, nouns and maximum a preposition such as *de* with a maximum length of eight words. This simple assumption seems to be enough in the vast majority of the cases of the parallel corpus. For the repetition of the noun phrase in the output sentence corresponding to the relative clause, it is exactly copied in most cases. In some other cases (where some proper nouns on capital letters are placed at the beginning), these nouns appear alone without its noun complements on the output sentences. In the Example (10), which we have previously studied for the transformation stage, the noun phrase *los jugadores excluidos* [the excluded players] is repeated without any change in both output sentences.

- (10) a. Los jugadores excluidos, que no se comportaron como profesionales, no jugarán más esta temporada.
[The excluded players, who didn't behave as professional players, will not play anymore this season]
- b. – Los jugadores excluidos no se comportaron como profesionales.
[The excluded players didn't behave as professional players.]
- Los jugadores excluidos no jugarán más esta temporada.
[The excluded players will not play anymore this season.]

5.3.2 Sentence reordering

Transformation stage gives a set of sentences with no order. The regeneration stage attempts to give a correct order to the output sentences. For the coordinated sentences, as well as the cause-effect subordinates and indirect to direct speech, this does not cause any problem in this matter. The logical decision of giving the order of the output sentences as they appear on the original sentences work perfectly.

For some non-restrictive relative subordinates containing embedded structures, the reordering of the output sentences is not a trivial issue. This issue becomes sharper as the number of conjugated verbs increase, so it is essential to start treating it at this simple stage of sentences with no more than 2 conjugated verbs. The decision to solve this issue was purely mathematical and it is based on the relative clause position. As we can see in Figure 5, the formula takes into account where the relative clause starts (including the noun phrase which complements), where it finishes and the length of the sentence.

If (position beginning relative clause + position ending relative clause < length (sentence)) →

Relative clause comes in the first place.

Las vacaciones de Luis, que no se presentó a trabajar ayer, le van a costar la destitución.

Rel. clause beg.: 3

Rel. clause ending: 11

Length sentence: 17

$3+11=14 < 17 \rightarrow$ a. *Luis no se presentó a trabajar ayer.*

b. *Las vacaciones de Luis le van a costar la destitución.*

Figure 5: Sentence reordering, relative clause example

We can appreciate in the example of Figure 5 how the relative clause ends up being at the first place while the rest of the sentence comes afterwards. If the condition were not fulfilled, the position would be inverted. This basic approach to handle the problem of sentence reordering on embedded structures gives a good understanding of the resulting output sentences.

6. Evaluation

This Chapter presents the evaluation in three different ways, each of them corresponding to a different goal. Subchapters 6.1 and 6.2 tackle two extrinsic evaluations whereas Subchapter 6.3 evaluates the automatic simplification system. Subchapter 6.1 treats the first main goal of this research: SS for MT. As explained in previous Chapters, simplification is carried out on the source language (Spanish) and the target language will be Korean and Chinese in this case. The second evaluation will be about improving the understandability of text for foreigners learning Spanish. To end the evaluation, Subchapter 6.3 will attempt to give a first evaluation of the automatic simplification system for sentences with two conjugated verbs. It is also important to notice that all the evaluations include in some way the parallel corpus.

6.1 Syntactic Simplification for Machine Translation

In this Subchapter, the effect of SS in MT will be evaluated by a comparison between direct MT systems and MT systems which previously have used SS on the source languages. There will be two different evaluations concerning different target languages, the first one taking Korean as the target language and the second one taking Chinese as the target language. Spanish will be in both cases the source languages.

The experiments settings are similar for both experiments. Thirty complex sentences have been randomly selected from thirty different texts of the parallel corpus created along this research (see Chapter 4), including their respective simplified ones. The thirty original sentences were translated respectively by a Spanish-Korean and a Spanish-Chinese translator, which was considered as Gold Standard.

The Spanish-Korean and Spanish-Chinese Google Translate was the MT system for the evaluation. First, the original sentences were introduced directly on the automatic

translator and the output sentences were stored, what we will call simply MT. Second, SS was used prior to the translation (see Subchapter 1.2 to understand the implementation of SS in the MT process). This second output will be called SS+MT. The simplified part of the thirty original sentences (taken from the parallel corpus) was introduced on Google Translate to produce three new outputs.

An evaluation of both ways was then carried out among native Korean and Chinese. The native speakers were asked to evaluate the grammaticality of the output sentence and compare each output with the Gold Standard regarding its preservation of meaning. The grammaticality was evaluated by following a 1-5 scale where 1 indicates a totally non grammatical sentence and 5 indicates a fully grammatical sentence. The preservation of meaning was also done by following the 1-5 scale in a similar way. 1 means that the output sentence does not preserve at all the meaning of the Gold Standard sentence and 5 indicates a total preservation of meaning. Appendix C shows a summary of the Excel files which the native Korean evaluators were asked to fill in, whereas Appendix D includes the native Chinese evaluators' files.

Subchapter 6.1.1 gives a small description of the peculiarities of the Korean language and shows the results from the experiment explained above. Subchapter 6.2.2 focuses on these concerns but considering Chinese as the target language in this case.

6.1.1 Spanish-Korean

Korean language, so-called *Hangeul*, uses a phonemic writing system which is composed of 14 consonants and 10 vowels. Unlike general phonemic writing systems such as Spanish, it combines consonants and vowels into syllabic units. The total number of linguistic units which can be made as a result of the combinations of consonants and vowels reaches 11,172. For example, *문자* is composed by three consonants *ㅁ, ㄴ, ㅈ* and two vowels *ㅏ, ㅑ*.

Korean was considered to be a Ural-Altaic language until the end of 20th century. It shares some characteristics with the Ural-Altaic languages: the Subject-Object-Verb structure, the morphologically richness and the agglutination among others. However, because of the lack of cognates with this language family and the highly amount of loanwords from the Chinese language, Korean language is currently defined as an isolated language (The National Institute of the Korean Language, 2008). Spanish and Korean are therefore considered to be far language with very few characteristics in common.

The MT experiment setting is prepared as explained at the beginning of Subchapter 6.1. There are some preliminary results obtained from three native Korean evaluators holding at least a university postgraduate degree. They were asked to filled two Excel files (see Appendix C) concerning the grammaticality and preservation of meaning from the output sentences given by the MT directly (MT) and output obtained by applying SS prior to MT (SS+MT).

Spanish-Korean MT systems have not been as developed as other pair of languages. According to the native Korean evaluators of this research, the output sentences given by the Spanish-Korean Google Translate MT system are poorly constructed and in many cases not understandable. This is reflected on the results which are summarized on Table 3.

Table 3: Spanish-Korean MT results, grammaticality.

	Grammaticality MT	Grammaticality SS+MT
Average	2.6	3.07**
Positive (4-5) ³	15.6%	31.1%
Neutral (3) ⁴	40%	44.4%
Negative (1-2) ⁵	44.4%	24.4%

** : difference statistically significant at the level 0,001

³ Positive attitude by the evaluator towards the sentence grammaticality (scores 4 and 5).

⁴ Neutral attitude by the evaluator towards the sentence grammaticality (score 3).

⁵ Negative attitude by the evaluator towards the sentence grammaticality (scores 1 and 2).

According to a paired t-test, the difference between the grammaticality score averages of MT and SS+MT output sentences is statistically significant ($t(89)=-5.205$, $p\text{-value}<0.001$). This is a considerable difference taking into account that the Gold Standard sentences have been translated directly from the original sentences and not the simplified ones. One of the reasons may be the usual sentence length of both languages. Whereas Spanish sentences accept long and complex sentences like the ones included on the test set, Korean language appreciates short and simpler sentences. This is one of the reasons why Spanish-Korean MT system backfires on these complex sentences. SS has therefore solved several problems posed by the MT system and the grammaticality average score has increased from 2.6 to 3.07 on a 1-5 scale, which is a significant improvement. We can see on Figure 6 how the average is clearly improved on every evaluator (S1, S2 and S3).

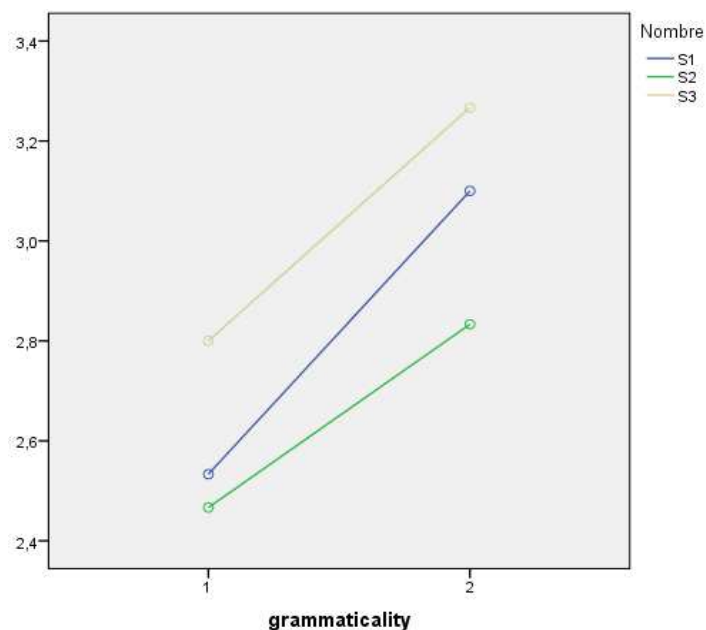


Figure 6: Spanish-Korean MT evaluators' graphic, grammaticality

Even though each evaluator has his own subjectivity about the grammaticality task, this is not relevant to compare direct MT and SS+MT. An ANOVA test carried out among the results of the experiment shows that the evaluators' attitudes toward this

comparison is extremely similar ($F(2,87)=0.409$, $p\text{-value}=0.666$), i.e., the grammaticality improvement is clear regardless the evaluator.

The results from the preservation of meaning task are summarized on Table 4. The first column (MT) shows the results from the preservation of meaning between direct MT output sentences and the Gold Standard. The second column (SS+MT) shows the results from the preservation of meaning between SS+MT and the Gold Standard. In this case, the meaning is better preserved by SS+MT. The difference is also statistically significant according to a t-test ($t(89)=-6.183$, $p\text{-value}<0.001$). As we can see from the Table 4, the poor meaning preservation obtained by MT (more than 50 % of the sentences got a negative attitude from the evaluator) are highly improved by using SS prior to MT (the sentences with a negative attitude from the evaluator is in this case quite below 50%).

Table 4: Spanish-Korean MT results, preservation of meaning

Preservation of meaning between Gold Standard and... ->	MT	SS+MT
Average	2.44	3.02**
Positive (4-5)	18.9%	28.9%
Neutral (3)	24.4%	35.6%
Negative (1-2)	56.7%	35.6%

** : difference statistically significant at the level 0,001

This significant difference between both tasks is also due to one of the main drawbacks of MT: handling long and complex sentences. The splitting rules considered in this research attempts to solve this MT problem, especially in MT systems poorly developed like the Spanish-Korean treated in this chapter.

As we have observed before in the grammaticality task, the results from each evaluator show a clear improvement on the meaning preservation task by applying SS, as we can appreciate from Figure 7. An ANOVA test shows how this improvement is similar among the three annotators regardless their particular results on either MT or SS+MT ($F(2,87)=1.919$, $p\text{-value}=0.153$).

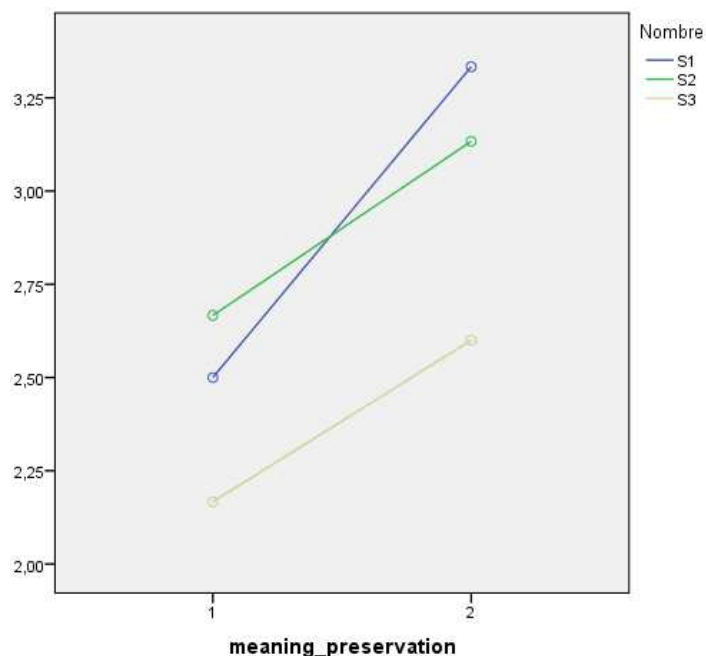


Figure 7: Spanish-Korean MT evaluators' graphic, preservation of meaning

6.1.2 Spanish-Chinese

Chinese language, so-called *Han-yu*, belongs to the Sino-Tibetan languages family. Around one-fifth of the world population uses Chinese as their native language. It is ideogramic but the characters have suffered a process of encoding with the time. Some of the main characteristics of the Chinese language in contrast to the Spanish language are the following ones: monosyllabic, tonal, use of measure words and Subject-Predicate-Object word order (Gong, 2001). Each Chinese character represents a monosyllabic Chinese word or morpheme (男(man), for instance). These characters have full meaning when found alone and the combinations with other characters enable the creation of new words (男 (man) +孩子 (kid) → 男孩子 (boy)). Therefore, as with the Korean language, Spanish and Chinese are very distant languages in terms of syntaxis and morphology.

As far as the Spanish-Chinese MT experiment is concerned, the general setting is the same as the one prepared for the Korean language explained in Subchapter 6.1 (see Appendix D). Some preliminary results were obtained from three graduated native Chinese speakers. They were asked to perform two tasks: First, evaluating the grammaticality and preservation of meaning of the output sentences given by the Google Translate Spanish-Chinese MT system taking original sentences as the input (MT) and second, evaluating the same for the output sentences given by the Google Translate Spanish-Chinese MT system considering the syntactically simplified sentences as the input (SS+MT). The results obtained from the evaluation about the grammaticality task are summarized on Table 5.

Table 5: Spanish-Chinese MT results, grammaticality

	Grammaticality MT	Grammaticality SS+MT
Average	2.41	2.62*
Positive (4-5)	12.2%	12.2%
Neutral (3)	31.1%	44.4%
Negative (1-2)	56.7%	43.3%

*: difference statistically significant at the level 0.05

As we can observe from Table 5, the grammaticality is improved when applying SS prior to MT. The difference is statistically significant at the 0.05 level according to a paired t-test ($t(89)=-2.074$, $p\text{-value}=0,041$). This difference is slightly smaller than the one obtained in the Korean experiment. More evaluators will be needed in the future (or finding an accurate automatic evaluation metric for the task) in order to obtain the real improvement on this task. As we can see on Figure 8, the task was handled in a different way by the three evaluators and the comparison between the grammaticality of MT and SS+MT may not be totally accurate. An ANOVA test let us confirm this graphical impression: $F(2,87)=11.727$, $p\text{-value} < 0.001$. This test suggests getting more evaluators to obtain more reliable results.

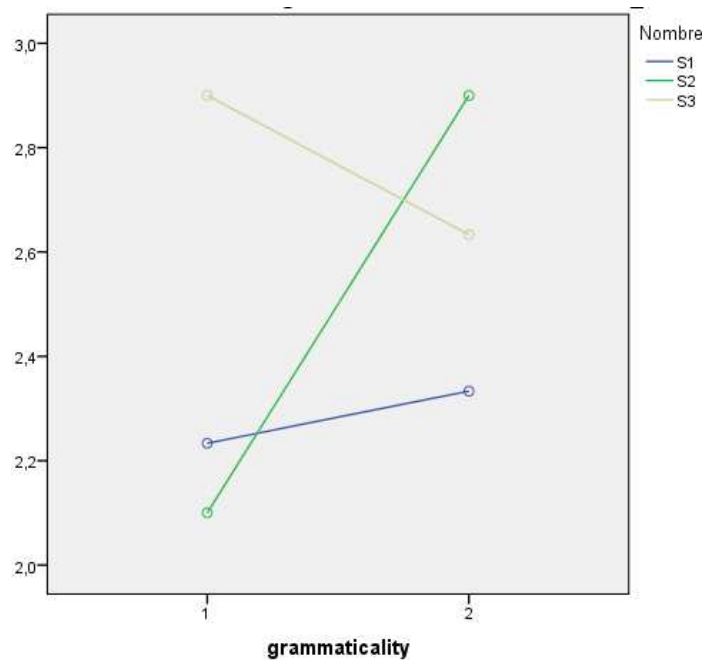


Figure 8: Spanish-Chinese MT evaluators' graphic, grammaticality

The meaning preservation results, which are the most important source to check the improvement of a MT system, were even more promising than the one regarding the grammaticality task. These results are summarized on Table 6. The meaning preservation average obtained by applying direct MT got the extremely low score 1.82, with 81.1% of the sentences obtaining a negative attitude from the evaluator. These results, even being still low for a MT system, show an important improvement by applying SS prior to the translation. The average reaches a 2.18 score and the percentage of sentences with a neutral or positive attitude by the evaluators is considerably raised, whereas the percentage of sentences with a negative attitude from the evaluator decreases until 65.6%. The average difference between MT and SS+MT for the meaning preservation task is even statistically more different than in the grammaticality task according to a paired t-test ($t(89)=-3,734$, $p\text{-value}<0.001$).

Table 6: Spanish-Chinese MT results, preservation of meaning

Preservation of meaning between Gold Standard and... ->	MT	SS+MT
Average	1.82	2.18**
Positive (4-5)	4.4%	8.9%
Neutral (3)	14.4%	25.6%
Negative (1-2)	81.1%	65.6%

** : difference statistically significant at the level 0.001

In this case, the results of each evaluator show an improvement (neutral correlation in one case) of the MT system by using SS (see Figure 9). However, the improvement is not regular according to an ANOVA test ($F(2,87)=11.600$, $p\text{-value} < 0.001$). Therefore the inter-annotator agreement on the Spanish-Chinese experiment is not quite high in general. Even though the improvement of the Spanish-Chinese MT is clear by using SS, more evaluators will be needed in the future to state the actual power of SS for this pair of languages.

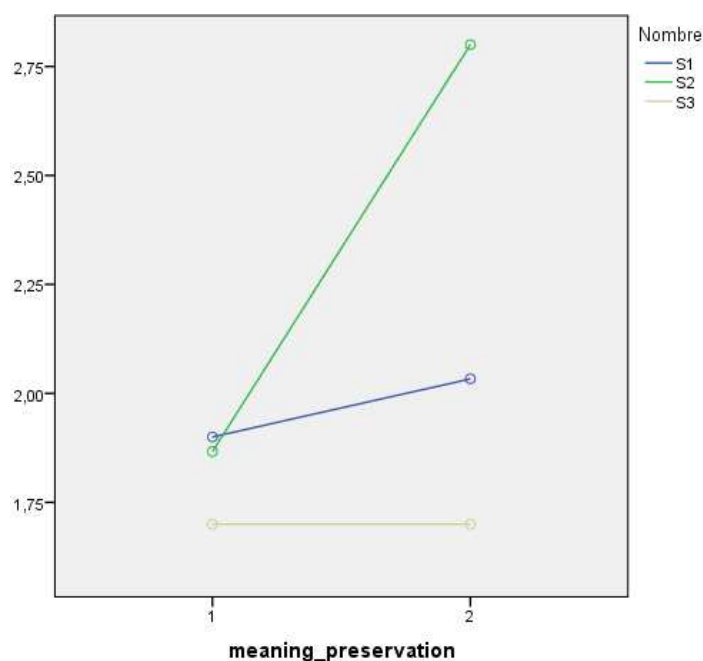


Figure 9: Spanish-Chinese MT evaluators' graphic, preservation of meaning

6.2 Foreigners learning Spanish

The research includes an extrinsic evaluation to check whether this approach may improve the understanding of a text by Spanish learners or not. Since the simplification rules intend to make sentences simpler, it would be interesting to know if it could also make texts more understandable for people with difficulty at reading. The main drawback of the approach is the readability, as there are some noun phrases which are repeated within sentences and it makes the reading sometimes slower. However, sentences become shorter and simpler after the simplification, which for certain parts of the population is a huge advantage in order to fully understand a text. For foreigners learning Spanish, there are certain structures which are difficult to understand as they usually differ from their native constructions.

Twenty four Spanish learners at the Faculty of *Traducción e Interpretación* in Granada University agreed to collaborate in this project. The participants had a native language other than Spanish. These languages were German, Bulgarian, Russian and French. They were divided in two groups of twelve participants each: A and B. Group A participants were given two original texts from AnCora Corpus and were asked to answer eight multiple-choice questions about them. It was a reading-comprehension test. Group B students were asked to do the same but from the group A's simplified texts. These simplified texts belong to the parallel corpus created along this research. Appendix E includes the tests given to both Group A and B students. The participants of each group had a similar level of English: five participants with an advanced level (C1-C2 according to the Common European Framework of Reference for Languages) of Spanish and seven with an Intermediate level (European Framework of Reference for Languages) Table 7 summarizes the characteristics about all the participants of the experiment.

Table 7: Summary of the participants in the reading-comprehension test

	A: Original text	B: Simplified text	Total
Participants	12	12	24
Intermediate	7	7	14
Advanced	5	5	10
Age average (years old)	24.2	22.6	23.4
Men	2	1	3
Women	10	11	21

The results obtained from the text were processed and summarized on Table 8. The results are given in percentages of correct answers. Surprisingly, the participants from group A (original texts) got a higher amount of right answers than the participants from group B (simplified texts). Group A obtained 82.3% of correct answers in contrast to the 79.2% obtained by group B. However, after applying an un-paired t-test at 0.05 level ($t(22)=-0.457$, $p\text{-value}=0.652$), we concluded that this difference is not statistically significant. Therefore these results suggest that the text comprehension has not been affected by SS. It does not improve the comprehension but it does not damage it either.

Table 8: Results from the reading-comprehension test

Correct answers	A: Original text	B: Simplified text
Total	82,3%	79,2%
Intermediate	82,1%	74,5%
Advanced	82,5%	87,5%

Regarding the Spanish level within the participants, it is interesting to notice how the simplification affects in a different way to advanced and intermediate students. It is quite unexpected the way the simplification damage the comprehension of intermediate level readers, whereas it slightly improves the performance of advanced readers. This may be due to the similar results obtained by intermediate and advanced readers in the group A. Group B (simplified text) obtained in their task more predictable results, as

intermediate level students got a 74.5% of correct answers in contrast to the 87.5% obtained by the advanced.

6.3 Automatic Syntactic Simplification system

For the automatic SS system evaluation, the AnCora Corpus XML documents provided were used. As explained in Taulé et al. (2008), the corpus was semi automatically tagged. These documents were fully parsed so PoS and lemmas were able to be extracted in the few cases when it was necessary. Since the automatic SS system currently just tackles sentences containing two conjugated verbs, a set of 40 sentences of two conjugated verbs were selected (see Appendix F). The selection process was done by following a random process. Every sentence was randomly selected from an AnCora Corpus text. There is no more than one sentence selected from a specific text. Therefore 40 different texts were necessary for the extraction of all the sentences to be used as a test set.

Since these 40 texts belong to the original part of the parallel corpus, the simplified part of the corpus was taken as a reference (Gold Standard) in order to evaluate the system. The clear alignment between sentences in the parallel corpus (as explained in Section 4), enabled an easy preparation of this test set and Gold Standard sentences. From the simplified sentences of the corpus, we found that 30 sentences were in some way simplified and 10 sentences were not simplified.

The evaluation was then divided in three parts. Firstly, the analysis stage is evaluated on Subsection 6.3.1 and it consists of checking whether the simplification should be carried out or not. Secondly, whether the sentences chooses the right case to simplify and it chooses the right position is evaluated, which is the transformation phase (Subsection 6.3.2). Thirdly, the results from the regeneration stage are shown on Subsection 6.3.2. This is done by following a 0-2 scale in contrast to the two other phases. To end with, Subsection 6.3.4 summarizes all these results considering the system

as a whole. It shows total averages taking an original sentence as input and giving the final simplified sentences as an output.

6.3.1 Analysis stage

As explained in the Subchapter 5.3, the analysis stage is the module in charge of deciding whether simplifying the input sentence or not. It mainly relies on finding the simplification articulation indicators and no simplification indicators. It also relies on PoS tagging to find out the number of verbs inside the sentences (which was not necessary in this case because of the test set selection process) and in a specific case, verb lemmatization as well.

The analysis stage takes an individual sentence from the test set as input and decides whether simplifying the sentences or not. The output of a single sentence takes only two values then: True for a sentence to be simplified and False for a sentence which does not need simplification. The results from the analysis stage are shown on Table 9.

Table 9: Analysis stage results

	Simplification	No Simplification
Gold Standard (n. of sentences)	30	10
Correct	28	10
Wrong	2	0
Precision:	100%	
Recall:	93.3%	

As we can appreciate from Table 9, there were 12 sentences which were not simplified by the system and 28 correctly simplified. From the 12 simplified sentences, two of them should have been simplified. This makes the recall not be perfect and gets a score of 93.3%, which is still extremely high. Since the main applications of this system

are mainly extrinsic, the mistakes of recognizing simplifiable sentences by the system are not quite important because they will not damage the performance of external applications such as MT (obviously in those cases they will not improve it either). In terms of precision, the analysis stage obtains a remarkable 100%. This means that there is no a single sentence simplified when it should not be simplified.

Therefore these results show that identifying simplifying patterns is not a difficult task and it can be done by basically using raw word searches. There were only three cases where verb lemmatizers were also needed (indirect to direct speech structures). However, this does not cause any problem to the automatic lemmatizers due to the rich variety of endings in verbs in the Spanish language making the automatic detection of the lemma quite an easy task. This means that the performance of the analysis stage does not rely on the performance of automatic parsers, which is quite encouraging to adapt this system to other Romance languages such as French, Italian and Portuguese.

As the automatic system is expected to be extended in order to handle sentences containing more than two conjugated verbs in the future, these results show that an induction from the 2-conjugated verb case is quite likely to succeed. Whether simplifying a sentence or not may be a not so difficult task regarding these results, which is a solid beginning for future improvements. The development of this complete automatic system may need further modifications in the transformation and regeneration stage though.

6.3.2 Transformation stage

The splitting of the original sentence is carried out on the transformation stage. The correct identification of the position where the splitting should be done and the simplification case are crucial at this point. This would lead to a clean transformation and a well understanding of the output sentence, as only a few corrections will be made later on the regeneration stage.

The same test set is used at this stage for the evaluation. We have therefore 28 sentences to be split at this stage. The system correctly selects the positions of the clauses and the simplification cases in the 100% of the sentences from the test set and makes just a single mistake in the transformation of the sentences. Therefore, the accuracy of the transformation stage reaches 96.43%. This result is already quite promising regardless the regeneration stage. It is also important to notice that the mistake produced by the system does not damage drastically the meaning of the original sentence –see Example (11). This is due to the correct identification of the case but the incorrect application of the transformation rule. In Example (11), we can appreciate how the subject ellipsis of the Spanish language on the subordinate clause has lead to the error. This is one of the main problems to handle automatically the Spanish language that should be studied deeply. However, the system usually handles well this subject ellipsis on the regeneration stage and in this particular case the general meaning of the output sentence is not seriously damaged.

- (11)
- a. Symplification system general resula. *Ivanov también afirmó que estaba "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.*
 - b. *Ivanov también afirmó: estaba "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.*
 - c. (Gold Standard) *Ivanov también afirmó estar "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.*

It is important to highlight the impressive detection of the coordination by the system so far. Having a set of words and combination of words which are the indication of each case is proved to be a simple but reliable approach to handle coordination. Especially significant for this case is the bag of words which are indicators of no simplification, as the coordinators connectors are in most cases ambiguous (“,”,”y”) and do not often indicate any coordination.

6.3.3 Regeneration stage

Once the splitting phase is done, regeneration stage attempts to fix some missing or broken links between the output sentences. This stage is the one causing most practical problems to the simplification. Even though it seems that all the transformation are already done in the previous stage, there a lot of important issues on the output sentences which need to be fixed. It is not an easy task for an automatic system to decide whether to use anaphoric references or not and the position where they should be included. It is also quite hard to correctly identifying the boundaries of the relative clauses which are attached to noun phrases (the boundaries of the noun phrase are easily found either). These are the reason why this stage is the most challenging in the whole SS process.

We continue with the same test set used in previous stages for the evaluation. 27 sentences which are already split on the transformation phase and need to be fixed on some cases. Each of the original sentences taken as an input counts now with one and most cases two output sentences like in the Example (12), extracted from the parallel corpus:

- (12) a. *El rival por el oro va a ser Rusia, que se impuso a Portugal.*
 b. *–El rival por el oro va a ser Rusia.*
 – **** se impuso a Portugal.*

There is an empty space in the second output sentence (***) which needs to be filled by an appropriate noun phrase. This will be done in the regeneration stage. In this particular example, there are two tasks to be done by the regeneration module of the system. The first one is correctly identified the noun phrase which is attached to the relative clause “que se impuso a Portugal” on the original sentence. In Subchapter 5.3 it is explained how the system deals with this in complicate cases where the noun phrase is

not easily identified. However, this case is quite simple as the noun Rusia is not approached by other nouns or adjectives which could make the task harder. Therefore, the system correctly selected Rusia as the noun phrase to be repeated on the second output sentence. Rusia will act as the subject of the new sentence then.

The second task which needs to be done on the regeneration stage is the reordering of the sentences. By applying the simple algorithm explained on Subchapter 5.3.2, the regeneration module correctly decided to keep the order as it appears in Example (12). Comparing the output results to the Gold Standard, this reordering task actually gets a 100% of accuracy on the test set. The 27 output sentences were given the correct order by the system. This means that the system tends to select the most natural order in all cases.

A general evaluation was done for the regeneration stage. Each output from the 27 already split sentences was given a 0-2 score. 2 means that the output is exactly same as the Gold Standard. 1 means that the output has the same meaning as the Gold Standard but something has been modified on a different way, just affecting the readability. 0 means that the meaning of the original sentence is on some way changed by the regeneration. The general results obtained by the regeneration module (including all the subtasks) are shown on Table 10.

Table 10: Regeneration stage results

Score	0	1	2
Number of sentences	1	8	18
Percentage	3.7%	29.6%	66.7%
Average:	1.63		

As we can appreciate on Table 10, 18 out of the 27 sentences (66.7%) were correctly regenerated according to the Gold Standard, which is a remarkably high percentage taking into account the complexity of this module. The score average of the system is 1.63. This result is quite promising, as a few improvements may lead to an

almost perfect score (2) by the system. Only one sentence got a 0 score, which is shown in the Example (13) where we can observe how the identification of a noun phrase attached to a relative clause was incorrectly handled.

(13) a. *Varios centenares de personas acudieron al palacio Brongniart, sede simbólica de la Bolsa parisiense. *** amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.*

b. *Varios centenares de personas acudieron al palacio Brongniart, sede simbólica de la Bolsa parisiense. La Bolsa parisiense amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.*

c. (Gold Standard) *Varios centenares de personas acudieron al palacio Brongniart, sede simbólica de la Bolsa parisiense. El palacio Brongniart amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.*

The noun phrase selected by the system to be repeated on the second sentence was La Bolsa parisiense while the correct one was El palacio Brongniart. The system failed to identify the link between the relative clause and the main sentence. What failed was the recognition of the apposition sede simbólica de la Bolsa parisiense from the noun El palacio Brongniart. However, the system, as the score states, found correctly or almost correctly the rest of the noun phrases of the test set.

In Example (14) below, we can appreciate how the second output sentence lacks a subject, which is totally right for the Spanish language. The module should improve in this matter though, as the subject will give a much better understandability of the sentence.

(14) a. *Patrick Kluivert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato. *** ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.*

b. *Patrick Kluivert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato. Ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.*

c. (Gold Standard) *Patrick Kluivert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato. **Patrick Kluivert** ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.*

Example (15) identifies correctly the noun phrase and repeated in the output sentence. Obviously the preservation of meaning between the Gold Standard and the output sentence is total. However, in this case, where a proper noun is included, it is suggested reducing the length of the noun phrase in order to improve the fluency of the sentence. These minor changes are not only stylistic, but also help sentences to be better processed by other NLP applications.

(15) a. *El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. *** hoy cumple 59 días en huelga de hambre.*

b. *El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. **El francés** Daniel Derguy hoy cumple 59 días en huelga de hambre.*

c. (Gold Standard) *El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. Daniel Derguy hoy cumple 59 días en huelga de hambre.*

As we have seen in the previous Examples (14) and (15) with score 1, the meaning of Gold Standard and output sentences is preserved, as in the remaining sentences with score 1. Therefore we can consider sentences with a score of 1 or 2 to be positively simplified. As we could observe from Table 10, there is a huge percentage of the test set with a score of 1 or 2. There are exactly 26 sentences out of 27 (96.3%) which are positively handled by the system at this stage, which is a high rate.

6.3.4 *General results*

Considering the original sentences as the input and the simplified sentences after the regeneration stage as the output, we obtain the following results shown in Table 11. Perfectly simplified are the sentence obtaining perfect scores at each stage of the system whereas positively simplified include the sentences which are correctly handled by the analysis and transformation stage and obtain a 1-2 score at the regeneration stage. Positively simplified sentences are the ones which have been simplified without damaging the meaning of the original sentence. The incorrectly simplified sentences include the sentences which have not been well simplified in either analysis or transformation stage or have obtained a 0 score at the regeneration stage. To summarize, incorrectly simplified sentences are the one which either does not preserve the meaning of the Gold Standard or have not been simplified when they were supposed to.

Table 11: Symplication system general results

	Total	Percentage
Input sentences	40	100%
Sentences simplified by the system	28	70%
Positively simplified	26	92.9%
Incorrectly simplified	2	7.1%
<i>Perfectly simplified</i>	18	64.3%
Sentences not simplified by the system	12	30%

The general results obtained are quite promising, as 92.9% of the simplified sentences were positively handled by the system. These are the sentences which suffered the splitting process of the simplification process and preserve the meaning with the original sentence (comparing it to the Gold Standard). There were just two sentences out of forty which were actually bad simplifications by the system. One was incorrectly handled on the transformation stage and the other one incorrectly regenerate at the last stage. This last stage is the most complex stage within the simplification process and the one which needs to get more adjustments in the future. However, there were a 64.3% of output sentences corresponding word by word to the respective Gold Standard at the end of the simplification process. This is an incredibly huge percentage considering the complexity of this task and the early stage of this project.

These results and the way the system has been developed suggest that an automatic syntactic simplification system is possible if a careful selection of the rules is done. One the main advantage of this approach is its adaptability, so further researches should focus on adapting it to be able to handle longer and more complex sentences.

7. Conclusions and future work

A new SS system based on splitting complex sentences is proposed. The general architecture of the system was inspired by the TS architecture proposed in Siddharthan (2002). The system is envisioned to be used as a preprocessing tool for other NLP applications such as Text Summarization, Information Extraction, Parsing and Machine Translation. The latter is the NLP application evaluated on this research, taking Spanish as source language and Korean and Chinese as target language of the MT system. The performance of the MT system evaluated has been improved by using the SS proposed prior to the translation. This approach has partially solved one of the main drawbacks of MT in handling long and complex sentence. A first version of the automatic SS system has also been analyzed and evaluated intrinsically. This SS system is expected to be extended in order to properly handle longer and more complex sentences.

This Chapter presents the main results obtained from the research together with a general overview of the project on Subchapter 7.1. Subchapter 7.2 addresses some directions for future research concerning the improvement of the system and a deeper analysis of evaluation metrics.

7.1 Main observations

The approach of this research differs from previous TS approaches in how the system is built and its applications. The SS system proposed relies on noun chunking rather than dependency structures, unlike previous works (Klebanov et al., 2004; Siddharthan, 2011; Drndarević et al., 2013). Its adaptability is one of the main advantages. The system is adaptable to close Romance languages such as Catalan, French, Italian or

Portuguese by just making a few adjustments of the rules to the specific language. Although it currently handles only complex sentences of two conjugated verbs, it is easily extensible to process longer sentences by applying an induction process. New rules can be also added to the system without damaging the remaining ones.

The results from the evaluation state that a reliable syntactic simplification system is feasible by considering these splitting rules. A few adjustments need to be made especially on the regeneration stage, focusing on the subject ellipsis problem on the Spanish language and the improvement of the specific noun chunkers.

A parallel corpus of Spanish original and simplified sentences has been built on this research. It is envisioned to be used as a tool in order to create an automatic Syntactic Simplification system. It has been later validated by native Spanish speakers with the expected results. Neither the grammaticality nor the preservation of meaning on newswire texts are damaged by the syntactic modifications. This validation was essential prior to carry out other extrinsic evaluations.

As far as Machine Translation is concerned, Syntactic Simplification has been proved to be a quite reliable and easy to implement monolingual resource for underdeveloped systems. Spanish-Chinese and Spanish-Korean Statistical Machine Translation systems have been evaluated on this research. The preliminary results state that the application of the Syntactic Simplification proposed prior to the translation is really beneficial on both cases. However, more evaluators will be needed in order to state the real improvement, especially on the Spanish-Chinese case according to the inter-annotator agreement results. Other languages should be taken into consideration as target languages in the future. In the next Subchapter, some directions for future research on this matter are indicated.

7.2 Future work

As pointed out in the previous Subchapter, further research should focus on the development of the Syntactic Simplification system and extend it to handle every kind of sentence from the natural language. Different rules may be also added in the system and evaluate its impact on extrinsic applications other than Machine Translation (Text Summarization, Information Extraction and parsing, for instance). Modules at the lexical and semantic level could be implemented in the system in order to solve some ambiguity problems on these extrinsic applications.

One of the main drawbacks of this approach is the lack of solutions at the discourse level. This system does not include any anaphora and co-reference resolution modules, which could give more accurate simplifications in some cases. Further research should focus on how to implement these modules to the system.

A possible way to improve the system will be to extend the parallel corpus which would enable to apply Machine Learning techniques like in Specia (2010). The future system could be hybrid by using the syntactic rules and Machine Learning in some complicated cases.

As far as the evaluation is concerned, further researches on evaluation metrics would be desirable. Text Simplification is a field which has always lacked of a widely accepted evaluation metric. This is mainly due to the different external applications for what the Text Simplification system is built. In this case, extrinsic evaluation metrics would be more indicative of how the system works rather than evaluating the system itself. A reliable baseline for extrinsic applications such as Machine Translation would be necessary prior to future improvements of the system. However, Machine Translation metrics such as BLEU are not excessively reliable because of the splitting nature of the simplifications. Therefore other automatic or human Machine Translation evaluation metrics should be considered and investigated.

References

- Aluísio, S., Specia, L., Pardo, T., Maziero, E., and Fortes, R. (2008) Towards Brazilian Portuguese Automatic Text Simplification Systems. *ACM Symposium on Document Engineering*, pp. 240–248.
- Bosque, I. (dir.) (2004) *Redes. Diccionario combinatorio del español contemporáneo*, Ediciones SM, Madrid.
- Bott, S. and Saggion, H. (2011) Spanish Text Simplification: An Exploratory Study. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*.
- Bott, S. and Saggion, H. (2011) An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction. *Proceedings of the ACL Workshop on Monolingual Text-to-Text Generation*, pp. 20-26.
- Camacho Collados, J. (2013) Splitting complex sentences for Natural Language Processing applications: Building a Simplified Spanish, *V International Conference on Corpus Linguistics*, Alicante, Spain, *in press*.
- Cardey, S., Greenfield, P., Alsharaf, H., and Shen, Y. (2004) Problems and Solutions in Machine Translation Involving Arabic, Chinese and French. *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04)*.
- Cardey, S. (2011) Machine Translation of Controlled Languages for More Reliable Human Communication in Safety Critical Applications. *Proceedings of the 12th*

International Symposium on Social Communication - Comunicación Social en el Siglo XXI, Santiago de Cuba, Cuba, Vol. II, pp. 953-958.

Chandrasekar, R., Doran, C., and Srinivas, B. (1996) Motivations and methods for text simplification. *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*, pp. 1041–1044.

Evans, R. (2011). Comparing methods for the syntactic simplification of sentences in information extraction. *Literary and Linguistic Computing*, 26(4), pp.371-388.

Drndarević, B., Štajner, S., Bott, S., Bautista, S., and Saggion, H. (2013), Automatic Text Simplification in Spanish: A Comparative Evaluation of Complementing Modules. *Computational Linguistics and Intelligent Text Processing*, 7817, pp. 488-500.

Goldberg, M. (1999). An Unsupervised Model for Statistically Determining Coordinate Phrase Attachment. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*. College Park, Maryland.

Gong, S.G. (2001) *China, Chinese and Chinese culture*, Darakwon.

Kaji, H. (1999) Controlled languages for machine translation: state of the art. *Proceedings of MT Summit VII "MT in the Great Translation Era"*. Kent Ridge Digital Labs, Singapore, pp. 37-39.

Klebanov, B. B., Knight, K., and Marcu, D. (2004). Text Simplification for Information-Seeking Applications. *On the Move to Meaningful Internet Systems*. Berlin: Springer-Verlag, pp. 735–47.

Koehn, P. (2010) Statistical Machine Translation. *Cambridge University Press. Textbook*.

- Medero, J. and Ostendorf, M. (2011). Identifying Targets for Syntactic Simplification. *ISCA International Workshop on Speech and Language Technology in Education (SLaTE 2011)*.
- Mitamura, T. (1999) *Controlled Language for Multilingual Machine Translation. Proceedings of Machine Translation Summit VII*.
- Nelken, R., and Shieber, S.M. (2006) Towards robust context-sensitive sentence alignment for monolingual corpora. *11th Conference of the European Chapter of the Association for Computational Linguistics*.
- Och, F. J. (2003) Minimum error rate training in statistical machine translation. *Proceedings of the 41st ACL, pp. 160-167*. Sapporo, Japan.
- Petersen, S., and Ostendorf, M. (2007) Text simplification for language learners: a corpus analysis. *Proceedings of Workshop on Speech and Language Technology for Education*.
- Siddharthan, A. (2002) An Architecture for a Text Simplification System. *Proceedings of the Language Engineering Conference (LEC'02): Hyderabad, India, December 13-December 15, 2002 IEEE Computer Society, London, United Kingdom, pp. 64*, Hyderabad, India.
- Siddharthan, A. (2004). Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4(1), pp. 77-109.
- Siddharthan, A. (2011). Text Simplification using Typed Dependencies: A Comparison of the Robustness of Different Generation Strategies. *Proceedings of the 13th European Workshop on Natural Language Generation*. Nancy, France.

Specia, L., Caseli, H.M., Pereira, T.F., Pardo, T.A.S., Gasperin, C., and Aluísio, S.M. (2009). Building a Brazilian Portuguese parallel corpus of original and simplified texts. In Alexander Gelbukh (ed), *Advances in Computational Linguistics, Research in Computer Science, 41*, pp. 59-70. 10th Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2009). Mexico City.

Specia, L. (2010). *Translating from Complex to Simplified Sentences. 9th International Conference on Computational Processing of the Portuguese Language (Propor-2010). Lecture Notes in Artificial Intelligence, 6001, Springer, pp. 30-39.* Porto Alegre, Brazil.

Taulé, M., Martí, M., and Recasens, M. (2008) AnCora: Multilevel Annotated Corpora for Catalan and Spanish. *Language Resources and Evaluation-LREC 2008*.

Temnikova, I. (2012) Text Complexity and Text Simplification in the Crisis Management domain. *Ph. D. Thesis*. University of Wolverhampton, United Kingdom.

The National Institute of the Korean Language (2008) *Hangeul*, accessed April 2013, <http://www.korean.go.kr/eng_hangeul>.

APPENDIX

Appendix A: Original and simplified text in the parallel corpus (Example)

Original text:

La española Margarita_Fullana logró hoy, sábado, la medalla de bronce en la prueba de " mountain-bike " o bicicleta todo_terreno (BTT) de los Juegos de Sydney, en una jornada en la que la italiana Paola_Pezzo, ganadora en Atlanta, revalidó su título olímpico, el único que se había disputado hasta la fecha.

Fullana entró en meta a 33 segundos de la campeona italiana, que cubrió los 35,7 kilómetros del recorrido olímpico en un tiempo de una hora, 49 minutos y 24 segundos, con una ventaja de 27 segundos sobre Barbara_Blatter, que regresará a Suiza con una medalla de plata.

La mallorquina, de 28 años, doble campeona del mundo de la especialidad (1999 y 2000), aspiraba al oro, pero la falta de un mayor desnivel en el circuito y una caída que sufrió en la cuarta vuelta tras chocar con Pezzo en un rocoso descenso la obligaron a conformarse con un bronce olímpico que, en cualquiera de los casos, pone broche a una sensacional temporada.

Fullana se situó pronto en los puestos de cabeza de una prueba en la que las competidoras dieron cinco vueltas a un circuito de 6,9 kilómetros - más un bucle llano de salida de 1.200 metros -, situado en la Fairfield_City_Farm, una granja en la que el ganado vacuno, ovino y porcino alterna con fauna autóctona como canguros, koalas o emús y que está a unos 30 kilómetros del Parque_Olímpico de Sydney.

Pero la primera en destacarse fue la suiza Blatter, que cruzó en cabeza las dos primeras vueltas de un recorrido cuyos diseñadores tuvieron especial cuidado en no dañar la flora y fauna del entorno, en unos Juegos cuya organización presume de ser la más ecológica de la historia.

Blatter abrió un hueco que fue de 17 segundos tras el segundo giro y que llegó a rozar el medio minuto en su amplitud máxima, pero a_falta_de dos kilómetros para completar la tercera vuelta, la ciclista balear enlazó con la suiza y la superó colocándose al_frente_de la prueba.

La autoridad del pedaleo de Fullana dejaba entrever la posibilidad de que la española redactase, efectivamente, la crónica de un oro anunciado : tras la tercera vuelta, la isleña lideraba con 14 segundos de ventaja sobre Blatter y 18 sobre Pezzo.

Pero en la cuarta, la italiana efectuó un sensacional despliegue de fuerza, superando a la helvética y conectando con la rueda trasera de Fullana, poco antes de adelantarla en una de las zonas más peligrosas del circuito, la " Dusty_Chute " (" caída polvorienta ").

En un descenso con piso muy irregular, la italiana superó a Fullana tras chocar con ella y derribarla, en una maniobra que no estuvo bien vista por la delegación española, que quiso presentar una reclamación, algo que descarta el reglamento olímpico del " mountain-bike ".

En la zona de meta se vivieron unos instantes no exentos de cierta tensión, pero los jueces, que, al parecer, visionaron el vídeo del adelantamiento, no apreciaron la necesidad de alterar el desenlace de la prueba.

Fullana, que milita en el Specialized y cuyo entrenador es el ex ciclista holandés Gert_Jan_Theunisse, perdió un tiempo importantísimo tras su accidente; la italiana abrió hueco y Blatter enlazó desde atrás, superando a Marga en la quinta y última ronda.

La rusa Alla_Epyfanova acabó cuarta en la prueba, a 1:21 y por delante de la canadiense Alison_Sydor, que cruzó la meta con un retraso de 2:54, en una prueba en la que la catalana Silvia_Rovira entró la decimocuarta, a 9:43.

Mañana se disputará en este mismo circuito la prueba masculina de esta disciplina, en la que se darán dos vueltas más que las dadas hoy por Fullana y sus rivales.

A la balear quizá le haya sabido a poco el bronce, pero tiene motivos suficientes para sonreír : Fullana pasará a la historia, entre otras cosas, por haber sido la primera mujer que logra una medalla olímpica para el ciclismo español.

[AnCora Corpus (Taulé et al., 2008)]

Simplified text:

La española Margarita_Fullana logró hoy, sábado, la medalla de bronce en la prueba de " mountain-bike " o bicicleta todo_terreno (BTT) de los Juegos de Sydney. En la jornada la italiana Paola_Pezzo, ganadora en Atlanta, revalidó su título olímpico, el único que se había disputado hasta la fecha.

Fullana entró en meta a 33 segundos de la campeona italiana. La campeona italiana cubrió los 35,7 kilómetros del recorrido olímpico en un tiempo de una hora, 49 minutos y 24 segundos, con una ventaja de 27 segundos sobre Barbara_Blatter. Barbara_Blatter regresará a Suiza con una medalla de plata.

La mallorquina, de 28 años, doble campeona del mundo de la especialidad (1999 y 2000), aspiraba al oro. Sin embargo, la falta de un mayor desnivel en el circuito y una caída que sufrió en la cuarta vuelta tras chocar con Pezzo en un rocoso descenso la obligaron a conformarse con un bronce olímpico que, en cualquiera de los casos, pone broche a una sensacional temporada.

Fullana se situó pronto en los puestos de cabeza de la prueba. En la prueba las competidoras dieron cinco vueltas a un circuito de 6,9 kilómetros - más un bucle llano de salida de 1.200 metros -, situado en la Fairfield_City_Farm, una granja. En la granja el ganado vacuno, ovino y porcino alterna con fauna autóctona como canguros, koalas o emús y que está a unos 30 kilómetros del Parque_Olímpico de Sydney.

Pero la primera en destacarse fue la suiza Blatter. La suiza Blatter cruzó en cabeza las dos primeras vueltas de un recorrido. Sus diseñadores tuvieron especial cuidado en no dañar la flora y fauna del entorno. La organización de los Juegos presume de ser la más ecológica de la historia.

Blatter abrió un hueco que fue de 17 segundos tras el segundo giro y que llegó a rozar el medio minuto en su amplitud máxima. Sin embargo a_falta_de dos kilómetros para completar la tercera vuelta, la ciclista balear enlazó con la suiza. La ciclista balear la superó colocándose al_frente_de la prueba.

La autoridad del pedaleo de Fullana dejaba entrever la posibilidad de que la española redactase, efectivamente, la crónica de un oro anunciado : tras la tercera vuelta, la isleña lideraba con 14 segundos de ventaja sobre Blatter y 18 sobre Pezzo.

Pero en la cuarta, la italiana efectuó un sensacional despliegue de fuerza, superando a la helvética y conectando con la rueda trasera de Fullana, poco antes de adelantarla en una de las zonas más peligrosas del circuito, la " Dusty_Chute " (" caída polvorienta ").

En un descenso con piso muy irregular, la italiana superó a Fullana tras chocar con ella y derribarla, en una maniobra que no estuvo bien vista por la delegación española. La delegación española quiso presentar una reclamación, algo que descarta el reglamento olímpico del " mountain-bike ".

En la zona de meta se vivieron unos instantes no exentos de cierta tensión. Sin embargo, los jueces no apreciaron la necesidad de alterar el desenlace de la prueba. Los jueces, al_parecer, visionaron el vídeo del adelantamiento,

Fullana milita en el Specialized. Su entrenador es el ex ciclista holandés Gert_Jan_Theunisse. Fullana perdió un tiempo importantísimo tras su accidente; la italiana abrió hueco. Blatter enlazó desde atrás, superando a Marga en la quinta y última ronda.

La rusa Alla_Epyfanova acabó cuarta en la prueba, a 1:21 y por delante de la canadiense Alison_Sydor. Alison-Sydor cruzó la meta con un retraso de 2:54, en una prueba en la que la catalana Silvia_Rovira entró la decimocuarta, a 9:43.

Mañana se disputará en este mismo circuito la prueba masculina de esta disciplina. En la prueba se darán dos vueltas más que las dadas hoy por Fullana y sus rivales.

A la balear quizá le haya sabido a poco el bronce. Sin embargo, tiene motivos suficientes para sonreír : Fullana pasará a la historia, entre otras cosas, por haber sido la primera mujer que logra una medalla olímpica para el ciclismo español.

Appendix B: Parallel corpus validation

Three excel sheets summarized in one:

Sheet 1 (Grammaticality of original sentences) + Sheet 2 (Grammaticality of original sentences) + Sheet 3 (Preservation of meaning):

A	B	C	D	E
Oración original	Gramati calidad columna A (1-5)	Oración simplificada	Gramati calidad columna C (1-5)	Preserva ción del significa do A-C: 1-5
El director de TV3, Lluís Oliva, ha explicado hoy que el canal catalán "ha alargado la mano todo lo que ha podido", con el objetivo de que el resto del estado español pueda ver el primer fin de semana de mayo la quinta prueba de Fórmula Uno, que se celebrará en el circuito de Montmeló, en Barcelona.		El director de TV3, Lluís Oliva, ha explicado hoy: el canal catalán "ha alargado la mano todo lo que ha podido", con el objetivo de que el resto del estado español pueda ver el primer fin de semana de mayo la quinta prueba de Fórmula Uno. La quinta prueba de Fórmula Uno se celebrará en el circuito de Montmeló, en Barcelona.		
No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado, quien busca la reelección en los comicios fijados para el 8 de octubre.		No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado. Marian Krzaklewski busca la reelección en los comicios fijados para el 8 de octubre.		
El conductor, Antonio Muñoz Serrano, que ha sido condenado a una multa de 120.000 pesetas y prohibición de la licencia de circulación durante un año, sufrió el accidente la mañana del 5 de junio del pasado año cuando conducía su cuadriciclo - cuyo uso requiere solo de la licencia de circulación -, por una travesía de Sant Boi de Llobregat (Barcelona).		El conductor, Antonio Muñoz Serrano, ha sido condenado a una multa de 120.000 pesetas y prohibición de la licencia de circulación durante un año. Antonio Muñoz Serrano sufrió el accidente la mañana del 5 de junio del pasado año cuando conducía su cuadriciclo por una travesía de Sant Boi de Llobregat (Barcelona). El uso de su cuadriciclo requiere solo de la licencia de circulación.		
Los equipos en la primera columna juegan como locales y el horario es del Este.		Los equipos en la primera columna juegan como locales. El horario es del Este.		
Raúl, por su parte, sigue sin saltar al terreno de juego, pero todo apunta a que el delantero podrá jugar.		Raúl, por su parte, sigue sin saltar al terreno de juego. Sin embargo, todo apunta a que el delantero podrá jugar.		
Fullana entró en meta a 33 segundos de la campeona italiana, que cubrió los 35,7 kilómetros del recorrido olímpico en un tiempo de una hora, 49 minutos y 24 segundos, con una ventaja de 27 segundos sobre Barbara Blatter, que regresará a Suiza con una medalla de plata.		Fullana entró en meta a 33 segundos de la campeona italiana. La campeona italiana cubrió los 35,7 kilómetros del recorrido olímpico en un tiempo de una hora, 49 minutos y 24 segundos, con una ventaja de 27 segundos sobre Barbara Blatter. Barbara Blatter regresará a Suiza con una medalla de plata.		

Respecto a las opciones que su equipo puede tener para conseguir su primer título de Liga, Pauleta dijo: "nos lo merecemos y tenemos las condiciones necesarias para conseguirlo".		Respecto a las opciones que su equipo puede tener para conseguir su primer título de Liga, Pauleta dijo: "Nos lo merecemos. Tenemos las condiciones necesarias para conseguirlo".		
Manuela de Madre, que además de alcaldesa es diputada en el Parlamento catalán, y el alcalde de L'Hospitalet de Llobregat, Celestino Corbacho, son los dos representantes de los socialistas catalanes en la comisión política, presidida por Manuel Chaves, que debe conducir el partido hasta el congreso de julio.		Manuela de Madre además de alcaldesa es diputada en el Parlamento catalán. Manuela de Madre y el alcalde de L'Hospitalet de Llobregat, Celestino Corbacho, son los dos representantes de los socialistas catalanes en la comisión política, presidida por Manuel Chave. La comisión política debe conducir el partido hasta el congreso de julio.		
El jugador del Valencia viajó el martes a Barcelona acompañado por el jefe de los servicios médicos del club, Jorge Candel para ser sometido a una TAC (Tomografía Axial Computerizada) que posteriormente analizó el doctor Villarrubia, quien determinó el proceso a seguir.		El jugador del Valencia viajó el martes a Barcelona acompañado por el jefe de los servicios médicos del club, Jorge Candel para ser sometido a una TAC (Tomografía Axial Computerizada) que posteriormente analizó el doctor Villarrubia. El doctor Villarrubia determinó el proceso a seguir.		
El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy - que hoy cumple 59 días en huelga de hambre -.		El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. Daniel Derguy hoy cumple 59 días en huelga de hambre.		
"No veo a ninguno de esos dos países bloqueando hasta el final la directiva", comentó el diplomático, quien explicó que las últimas modificaciones introducidas en el borrador han satisfecho ya a otra delegación reticente, Holanda.		"No veo a ninguno de esos dos países bloqueando hasta el final la directiva", comentó el diplomático. El diplomático explicó que las últimas modificaciones introducidas en el borrador han satisfecho ya a otra delegación reticente, Holanda.		
El programa se iniciará con una feria "Info.com/2000" que se celebrará a partir del 17 de febrero en la que Microsoft lanzará el Windows 2000 y Telefónica Data y Bankinter presentarán sus productos y a partir de entonces se efectuará un recorrido por las 20 ciudades más importantes del país en las que difundirán las nuevas tecnologías para la empresa.		El programa se iniciará con una feria "Info.com/2000" en la que Microsoft lanzará el Windows 2000 y Telefónica Data y Bankinter presentarán sus productos. La feria "Info.com/2000" se celebrará a partir del 17 de febrero. A partir de entonces se efectuará un recorrido por las 20 ciudades más importantes del país. En las 20 ciudades más importantes del país difundirán las nuevas tecnologías para la empresa.		
El diputado socialista Ramón Jaúregui considera que "la presencia de Ibarretxe en La Moncloa ha sido utilizada para reclamar radicalidad nacionalista" y considera que el Gobierno del lehendakari es "insostenible".		El diputado socialista Ramón Jaúregui considera: "la presencia de Ibarretxe en La Moncloa ha sido utilizada para reclamar radicalidad nacionalista". Ramón Jaúregui considera: el Gobierno del lehendakari es "insostenible".		

El presidente ejecutivo explicó que representantes del banco irán al campo y preguntarán al tendero y arrendatario cómo paga las cuentas el solicitante del préstamo y "al cura (sacerdote) cómo se comporta".		El presidente ejecutivo explicó: representantes del banco irán al campo. Representantes del banco preguntarán al tendero y arrendatario cómo paga las cuentas el solicitante del préstamo y "al cura (sacerdote) cómo se comporta".		
Explicó que el proyecto de construcción del tren de alta velocidad entre Córdoba y Málaga cuenta ya con un estudio de impacto ambiental elaborado por la Junta, pero que el Ministerio de Fomento había solicitado una nueva declaración a Medio Ambiente.		Explicó: el proyecto de construcción del tren de alta velocidad entre Córdoba y Málaga cuenta ya con un estudio de impacto ambiental elaborado por la Junta. Sin embargo, el Ministerio de Fomento había solicitado una nueva declaración a Medio Ambiente.		
Ivanov también afirmó que estaba "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.		Ivanov también afirmó estar "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.		
"Una Rusia fuerte y unida está llamada a desempeñar un papel importante en un mundo multipolar y en la palestra internacional ", dijo el canciller cubano, quien destacó su coincidencia con Ivanov sobre "importantes problemas internacionales".		"Una Rusia fuerte y unida está llamada a desempeñar un papel importante en un mundo multipolar y en la palestra internacional ", dijo el canciller cubano. El canciller cubano destacó su coincidencia con Ivanov sobre "importantes problemas internacionales".		
El rival por el oro va a ser Rusia, que se impuso a Portugal.		El rival por el oro va a ser Rusia. Rusia se impuso a Portugal.		
El barco en el que viajaron como polizones, "Wuarta", con bandera de Bahamas, había zarpado de Guayaquil (Ecuador) el pasado día 15 con una carga de diversos productos con destino a Estados Unidos, e hizo el pasado fin de semana una escala de reabastecimiento en el puerto caribeño de Colón.		El barco en el que viajaron como polizones, "Wuarta", con bandera de Bahamas, había zarpado de Guayaquil (Ecuador) el pasado día 15 con una carga de diversos productos con destino a Estados Unidos. El barco hizo el pasado fin de semana una escala de reabastecimiento en el puerto caribeño de Colón.		
Las derrotas, que ya llegan a 18 en lo que va de los partidos de exhibición, han hecho que la paciencia de Steinbrenner se termine, por lo que ya quiere soluciones.		Las derrotas ya llegan a 18 en lo que va de los partidos de exhibición. Las derrotas han hecho que la paciencia de Steinbrenner se termine. Por lo tanto, Steinbrenner ya quiere soluciones.		
Mientras tanto, el principal índice de la Bolsa de París, el CAC-40, abrió la sesión de hoy con una fuerte bajada del 2,38 por ciento a 6.105 puntos, una tendencia que continuó a lo largo de gran parte de la sesión, aunque se enjugaron un tanto las pérdidas.		Mientras tanto, el principal índice de la Bolsa de París, el CAC-40, abrió la sesión de hoy con una fuerte bajada del 2,38 por ciento a 6.105 puntos. La tendencia continuó a lo largo de gran parte de la sesión. Sin embargo, se enjugaron un tanto las pérdidas.		
El secretario general del PSC-PSOE, Jaime Blanco, se mostró hoy " muy satisfecho " con el resultado del XXV Congreso del partido, aunque agregó que le queda una " pequeña insatisfacción " por la fórmula acordada para la elección de José Luis Rodríguez Zapatero como nuevo líder de los socialistas.		El secretario general del PSC-PSOE, Jaime Blanco, se mostró hoy " muy satisfecho " con el resultado del XXV Congreso del partido. Sin embargo, Jaime Blanco agregó: le queda una "pequeña insatisfacción" por la fórmula acordada para la elección de José Luis Rodríguez Zapatero como nuevo líder de los socialistas.		

José Hurtado recordó que "Barcelona se presentó en varias ocasiones" por lo que abogó en que "Sevilla esté en la misma línea, porque tarde o temprano, si seguimos en este camino, lo conseguiremos".		José Hurtado recordó: "Barcelona se presentó en varias ocasiones". Por lo tanto, José Hurtado abogó en que "Sevilla esté en la misma línea. Motivo: tarde o temprano, si seguimos en este camino, lo conseguiremos".		
Esa ventaja, que ya era de gran valía para el conjunto amarillo, se transformó en un auténtico tesoro cuando la amplió Alvaro, con un cabezazo que finiquitó una buena jugada de Pablo Lago por la banda derecha.		Esa ventaja ya era de gran valía para el conjunto amarillo. Esa ventaja se transformó en un auténtico tesoro cuando la amplió Alvaro, con un cabezazo que finiquitó una buena jugada de Pablo Lago por la banda derecha.		
En Teherán no habrá segunda vuelta electoral, pues 32 candidatos, dos más de los necesarios, han conseguido más del 25 por ciento de los votos requeridos para ser designados en la primera ronda, según el sistema de listas abiertas que se utiliza en Irán.		En Teherán no habrá segunda vuelta electoral. Motivo: 32 candidatos, dos más de los necesarios, han conseguido más del 25 por ciento de los votos requeridos para ser designados en la primera ronda, según el sistema de listas abiertas que se utiliza en Irán.		
En este sentido, aseguran que Ecuador es una nación pequeña y no es un "paraíso fiscal" que libera por completo su mercado monetario, pues se han adoptado una serie de normas para evitar ese tipo de problemas.		En este sentido, aseguran: Ecuador es una nación pequeña y no es un "paraíso fiscal" que libera por completo su mercado monetario. Motivo: se han adoptado una serie de normas para evitar ese tipo de problemas.		
El portavoz del PSOE en la Comisión de Interior del Congreso, Ramón Jáuregui, aseguró que aunque su partido desea que de la cita entre el presidente del Gobierno, José María Aznar, y el lehendakari, Juan José Ibarretxe, salgan "avances", su impresión es que se producirá un "desencuentro profundo".		El portavoz del PSOE en la Comisión de Interior del Congreso, Ramón Jáuregui, aseguró: su impresión es que se producirá un "desencuentro profundo". Sin embargo, su partido desea que de la cita entre el presidente del Gobierno, José María Aznar, y el lehendakari, Juan José Ibarretxe, salgan "avances".		
El ministro de Economía, Rodrigo Rato, afirmó hoy que el Ejecutivo actuará en la participación recíproca entre el BBVA y Telefónica "siempre en defensa de los intereses generales", ya que ha pasado la época en la que los acuerdos pasaban "por el poder político" e "incluso por la financiación del PSOE".		El ministro de Economía, Rodrigo Rato, afirmó hoy: el Ejecutivo actuará en la participación recíproca entre el BBVA y Telefónica "siempre en defensa de los intereses generales". Motivo: ha pasado la época en la que los acuerdos pasaban "por el poder político" e "incluso por la financiación del PSOE".		
Caja Rural de Asturias prevé obtener este ejercicio unos beneficios netos de 2.237 millones de pesetas, lo que supone un incremento del 5 por ciento respecto a 1999, con lo que volverá a batir su "récord" de rentabilidad, según indicó el presidente de la entidad, Román Suárez Blanco.		Caja Rural de Asturias prevé obtener este ejercicio unos beneficios netos de 2.237 millones de pesetas. Esto supone un incremento del 5 por ciento respecto a 1999. Con esto volverá a batir su "récord" de rentabilidad, según indicó el presidente de la entidad, Román Suárez Blanco.		
Unos días antes, el atleta nigeriano de 22 años Hyginus Anugo, que se encontraba en la ciudad austral a la espera de competir en los 400 metros, fue atropellado por un coche cerca del anillo olímpico, y murió a consecuencia de las heridas.		Unos días antes, el atleta nigeriano de 22 años Hyginus Anugo fue atropellado por un coche cerca del anillo olímpico. Hyginus Anugo se encontraba en la ciudad austral a la espera de competir en los 400 metros. Hyginus Anugo murió a consecuencia de las heridas.		

Appendix C: Evaluation of the Syntactic Simplification for Spanish-Korean Machine Translation using Google Translate

Sheet 1 (Direct MT without applying SS):

A	B	C	D
Gold Standard: Human translation	Google Translate	Grammaticality column B (1-5)	Preservation meaning A-B (1-5)
TV 3 의 감독 Lluís Oliva 는 오늘 카탈란 채널 바르셀로나 몬트멜로 라운드에서 열릴 포물러 원의 다섯번째 경기를 스페인 나머지 지역에서도 5 월 첫째 주말에 볼 수 있기 위해 “할 수 있는 모든 방법을 동원했다”라고 설명했다.	TV3 의 이사 Lluís Oliva 는 카탈루냐어 채널은 스페인 국가의 나머지의 첫 번째 주말에 볼 수 있도록 '된 가늘고 긴 손 모든 것을 갖추고 있습니다 "라고 현재 설명 다섯 번째 시험은 월 포물러 원은 바르셀로나의 회로 드 카탈루냐 (Catalunya)에서 개최합니다.		
그렇지만, 녹화 테이프는 대통령 우파의 후보자이고 10 월 8 일에 있을 선거에서 재선거를 원하는 현재 주지사의 라이벌인 Marian Krzaklewski 의 손으로 갔다.	그러나, 테이프 녹음 마리안 Krzaklewski, 대통령과 10 월 8 설정 선거에 재선을 찾고 있습니다 상태의 현재 머리의 라이벌에 맞는 후보의 손에서왔다.		
운전자 안토니오 무뇨스 세라노(Antonio Muñoz Serrano)는 작년 6 월 5 일 허가증이 요구되는 4 륜거 타다가 바르셀로나의 산 보이 드 요브레갓(Sant Boi de Llobregat)에서 사고를 당했고 120,000 페세타 벌금 및 1 년간 면허 정지에 처해졌다.	그의 쿼드 운전하는 동안 드라이버는 120,000 pesetas 의 벌금을 선고하고 년 동안 순환 라이선스를 금지 한 안토니오 무 노즈 세라노 (Serrano)는 지난 올해 6 월 5 일 아침에 사고를 고통 - 만 필요있는 사용증번호판 - 산티아고 드 Boi Llobregat (바르셀로나)의 여행 있습니다.		
첫 번째 줄의 팀 경기는 시간과 지역적으로 동쪽이다.	지역 및 시간은 동부만큼 팀은 첫 번째 열에서 재생됩니다.		
Raúl 이 여전히 필드로 뛰지는 않지만, 스트라이커가 경기를 할 수 있을 것이라고 모두 겨냥하고 있다.	라울 한편, 여전히 현장에 뛰어하지 않습니다,하지만 공격수가 재생됩니다 것으로 보인다.		

<p>Fullana 는 스위스로 은메달이 되돌아갈 Barbara Blatter 의 27 초 앞서 1 시간 올림픽 코스인 35,7 킬로미터를 49 분 24 초로, 이탈리아 챔피언의 33 초를 목표로 들어왔다.</p>	<p>폴라나의 목표는 스위스로 돌아갑니다 바바라 블래터, 27 초의 리드와 함께 한 시간, 49 분에서 24 초 35.7 km 실행할 올림픽 시간을 적용 이탈리아 챔피언으로 33 초 와서실버 메달.</p>		
<p>첫 수상자 리가 Pauleta 가 첫 타이틀을 얻기 위해 그 팀이 가질 수 있는 견해를 말했다: "우리는 잘 맞고 우승을 위해 필요한 조건을 가지고 있다."</p>	<p>컴퓨터가 처음으로 리그 우승을 할 수 있는 옵션에 관한, 파올레타는 말했다, "우리는 가치 그리고 우리는 그것을 위해 필요한 조건이었다."</p>		
<p>Manuela de Madre 카탈란 지방 의회에서 선출된 여자 시장 뿐만 아니라 L'Hospitalet de Llobregat 의 시장 Celestino Corbacho 둘 다 7 월 의회까지 당을 이끌어야 하는 Manuel Chaves 에 의해 이끌어지는 정치 위원회에서 카탈란 사회주의자들을 대표하고 있다.</p>	<p>마뉴 엘라 드도 카탈로니아 의회에 조역입니다 마드레, 그리고 르 호스 피탈 레 데 로브 레 가트, Celestino Corbacho 의 시장은 리드해야 마누엘 샤 위원장으로 정치위원회의 카탈로니아 어 사회주의의 두 대표,입니다 7 월까지 의회 파티.</p>		
<p>발렌시아 선수는 따르는 과정을 결정한 Villarrubia 박사가 후에 분석했던 TAC (컴퓨터 단층 촬영)를 담당하기 위해 클럽 의료 서비스의 팀장인 Jorge Candel 과 동행하여 화요일에 바르셀로나로 여행했다.</p>	<p>발렌시아 선수는 화요일에 바르셀로나 여행, 클럽 의료 서비스의 머리와 함께, 호르헤 CANDEL 우리가 연속적으로 따라야 할 절차를 결정 박사 Villarrubia 을 분석 CT 검사 (컴퓨터 단층 촬영)를 받아야합니다.</p>		
<p>지난 주 월요일, 형사법원은 "Pototo"와 오늘로써 단식 농성이 59 일째 되는 그의 예전 부책임자 프랑수아 Daniel Derguy 에게 테러리스트 목적과 관련 있는 범죄에 대한 10 년의 감옥형과 최대 벌금을 선고했다.</p>	<p>현재 단식 투쟁에 오십구일을 제공하는 -- 월요일, 형사 법원은 "pototo"와 그의 전 부, 다니엘 Derguy 프랑수아, 테러 목표와 범죄 단체에 대한 최대 벌금, 감옥에서 10 년 선고.</p>		
<p>네델란드 원고에 삽입된 최종 수정이 최근 다른 대표단을 이미 만족시켰다고 설명한 외교관은 "그 두나라 중 어떤 나라도 마지막까지</p>	<p>"나는 끝까지 지시어를 차단하는 두 나라 중 하나를 참조하지 않는다"초안과의 최신 개정 과목 다른 대표단, 네델란드 충족되었는지 설명</p>		

법률안을 차단하지 않을 것이라고 본다" 라고 언급했다.	외교관은 말했다.		
Microsoft 가 Windows 2000 과 Telefónica Data 그리고 Bankinter 를 착수하는 2 월 17 일 부터 열릴 "Info.com/2000" 박람회로 시작될 프로그램은 그 제품들을 소개할 것이고 그 이후로 그 회사를 위한 새로운 기술들을 퍼뜨릴 나라들 중 가장 중요한 도시들로 투어가 실현될 것이다.	이 프로그램은 마이크로 소프트가 자사의 제품을 제시하고 그 후 20 투어된다 Windows 2000 및 텔레포니카 데이터 및 Bankinter 해제 되는 2 월 17 일부터 개최되는 박람회 "Info.com/2000"로 시작합니다 국가의 도시는 어떤 회사를위한 새로운 기술을 보급.		
사회주의자 국회의원 라몬 하우레기(Ramón Jauregui)는 "몬클로아에서 Ibarretxe 의 입회(출정)는 민족주의의 극단성을 요구하기 위해 사용되었다"고 고려하고 lehendakari 의 정부는 "지속불가능"하다고 고려한다.	라몬 하우 레기 사회주의 MP 는 "라 몽클 로아의 Ibarretxe 의 존재가 급진적 민족주의를 주장하는 데 사용되었습니다 '바스크 대통령의 정부는"지속 가능 "하다고 생각 믿습니다.		
CEO 는 은행 대표자들이 농촌으로 가서 가게 주인 및 임차인들에게 어떻게 대출 계좌를 지불 할 것인지 물을 것이다. "성직자에게 어떻게 행동할지"를 설명했다.	CEO 는 은행의 대표자가 현장에 가서 가게 주인과 방법 계산서에게 대출하고 지불하는 임차인에게 설명했다 "그것이 작동하는 방법 성직자 (제사장)."		
코르도바와 말라기 간의 고속철도 건설 프로젝트는 안달루시아 정부가 제작한 환경충격에 관한 연구를 참조했지만, 개발부는 환경부에게 새로운 성명서를 신청했다.	그는 코르도바와 말라가 간의 고속 열차를 건설하는 프로젝트가 이미위원회에 의해 제조 환경에 미치는 영향 연구를 가지고 있다고 설명하지만, 국민의 정부가 작동하는지는 환경에 새로운 문을 요청했다.		
이바노프는 또한 Pérez Roque 와 "유익한" 대화 이후 "아주 만족했다"고 말했다.	이바노프는 또한 그가 페레즈 로크와 함께 "결실"대화 후 "매우 만족"라고 말했다.		
"강한 연방 러시아는 다극화 세계와 국제적 활동무대에서 중요한 역할을 한다." "중요한 국제적 문제점"에 대해 Ivanov 와	"강한 미국 러시아가 다극화 세계와 국제 무대에서 중요한 역할을하는 설정"에 이바노프와의 계약을 강조 쿠바 외무 장관은 말했다, "중요한 국제		

일치한다고 강조한 쿠바의 외무부 장관이 말했다.	문제."		
포르투갈을 이긴 러시아는 금메달을 위해 싸울 것입니다.	금 라이벌 포르투갈을 이길 러시아 될 것입니다.		
밀항자로 여행했던 바하마 국기가 있는 배 "Wuarta"는 미국을 목적지로 과야킬 (에콰도르)에서부터 지난 15 일 동안 다양한 제품의 짐들과 항해했고 지난주 콜론의 케리비안 항구에서 보급 기항했다.	밀항자로 여행 배, 바하마 플래그 "Wuarta"는 미국의 다른 제품의 부하 하루 15 과야 킬 (에콰도르)에서 항해와 주말에 한 규모 콜론의 카리브해 포트에서 보충.		
시범게임에서 합계 총 18 로 패배는 Steinbrenner 의 참을성을 다하게 만들어 이제는 해결책을 원한다.	지금은 시범 경기에서 지금까지 18 총 손실은 스타인브레너의 인내는 솔루션을 원하는, 그래서 부족 만들었습니다.		
파리 증권거래소의 주요 지표는 CAC-40 으로 2.38 퍼센트로 강하게 떨어져 6.105 포인트로 오늘 개회한 반면 손실을 없었지만, 대부분의 세션을 통해 계속되는 경향이었다.	한편, 파리 증권 거래소, CAC-40 의 주요 지수는 6,105 포인트 2.38 %의 급락 많은 세션을 통해 계속 추세에 오늘 세션을 연 다소 손실을 닦아 있지만.		
PSC-PSOE 의 서기장 Jaime Blanco 는 비록 사회주의자들 중에서 새로운 리더로서 José Luis Rodríguez Zapatero 의 선거를 위한 결의서에 "조금 불만족"임을 덧붙였지만, 오늘 당의 25 의회의 결과가 "아주 만족스러운" 모습을 나타냈다.	PSC-PSOE, 제임스 화이트의 사무 총장은 이제 파티 XXV 의회의 결과에 "매우 만족"이었다, 그러나 그는 같은 호세 루이스 로드리게스 사파 테로의 선거에 대해 동의 식으로 "약간 불만"이라고 추가사회주의의 새로운 지도자.		
"늦건 이르건 만약 계속 이 길로 갈다면, 그것을 얻을 것이기 때문에 세비야는 같은 라인에 있다."라고 언급하여 José Hurtado 는 "바르셀로나가 다양한 기회에 나타났다"고 기억했다.	호세 Hurtado 그래서 "우리는이 길을 계속하면 조만간, 당신은 얻을 것이다 때문에 세비야는 같은 라인에있다"라고 불리는 "바르셀로나는 여러 경우에 표시됩니다"라고 말했다.		

<p>노란 축구 팀에게 거대한 승리를 안겨줬던 그 경기는 우측 윙에서 Pablo Lago 가 펼쳤던 좋은 경기가 Alvaro 의 헤딩 슈트로 넓혀 졌을 때 진정한 진가로 바뀌었다.</p>	<p>그 장점은, 그것은 노란색 집합에 큰 값이었다 즉, 헤더가 정착 오른쪽 윙에서 파블로라고하여 좋은 움직임과 함께, 알바가 확장 보물이되었다.</p>		
<p>테헤란에서는 두번째 선거가 없을 것이다. 왜냐하면 필요 이상 보다 2 명 많은, 32 명의 후보가 이란에서 사용되는 열린 리스트 시스템에 따라 첫 번째 라운드에서 임명되기 위해 요구되는 25 퍼센트 이상의 표를 얻었기 때문이다.</p>	<p>테헤란에서이란에 사용되는 열기 목록 체계에 따라 필요 이상의 32 개 후보, 첫 라운드에서 선출해야하는 투표의 25 % 이상을 수상으로, 더 결선 투표가 없을 것입니다.</p>		
<p>이러한 의미에서, 에콰도르는 작은 국가이고 화폐시장의 완전히 자유로운 "조세 회피국"이 아니라고 확인한다. 왜냐하면 그런 형태의 문제점들을 피하기 위한 규범들을 채택했기 때문이다.</p>	<p>이런 의미에서, 그들은 이러한 문제를 방지하기 위해 일련의 규칙을 채택으로 에콰도르, 작은 나라가 아니라 완전히 돈을 시장 확보 "세금 천국"이라고 말한다.</p>		
<p>국회내부위원회에서 PSOE 대변인 Ramón Jáuregui 는 비록 그의 당이 "진보"로 나아가기 위해 정부 수상 José María Aznar 와 바스크자치정부의 수반 Juan José Ibarretxe 사이의 만남을 원한다 할지라도 그의 생각으로는 "깊은 오해"를 낳게 될 것이라고 언급했다.</p>	<p>의회 인테리어, 라몬 하우 레기의위원회에 PSOE 대변인은, 그의 당은 국무 총리, 호세 마리아 아스나르, 그리고 바스크 대통령 후안 호세 Ibarretxe 사이의 만남을 원하는 동안, "진행"인쇄를 얻을 수있다"깊은 불일치"가있을 것입니다.</p>		
<p>경제부 장관 Rodrigo Rato 는 오늘 정부가 "정치적 권력"과 "심지어 PSOE 의 자금조달"을 통한 합의의 시대는 지났기 때문에 "항상 공공이익을 지키 위한" BBVA 와 Telefónica 사이의 상호 참여를 실행시킬 것이라고 언급했다.</p>	<p>경제 장관 로드리고 라토는, 정치 권력은 "이 협정을 통과 시간을 통과로 정부가"일반적인 관심의 방어에 항상 "BBVA 와 텔레포니카 간의 상호 참여에 역할을했다고 말했다 "E"도 PSOE 금융".</p>		

<p>아스투리아스의 농업신용은행은 단체 회장인 Román Suárez Blanco 가 지적한 것에 따라 수익성에 대한 그의 "최대 결과"를 파괴할 1999 년 대비 5 퍼센트 상승한 것인 2.237 백만의 순이익을 이번 년도에 달성할 것이라고 예견했다.</p>	<p>올해 당신의 "레코드"성능을 이길 것입니다 1999 에 5 %의 증가를 나타내는 2237 만 유로의 순 이익을 얻을 것으로 예상 CAJA 농촌 드 아스투리아스 말했다 조직의 대통령 로마 수아레스 블랑코.</p>		
<p>며칠전, 남부도시에서 발견된 400 미터 경기를 희망한 22 살의 나이지리안 선수 Hyginus Anugo 는 올림픽 링 근처 차에 치어 상해로 인해 숨졌다.</p>	<p>몇 일전, 400m 에서 경쟁하기 위해 대기 남부 도시에서 있었던 나이지리아 선수 Anugo 22 Hyginus 은, 올림픽 링 근처에 차에 치, 상처 사망했다.</p>		

Sheet 2 (Direct SS+MT):

A	B	C	D
Gold Standard: Human translation	SS+Google Translate	Grammaticality column B (1-5)	Preservation meaning A-B (1-5)
<p>TV 3 의 감독 Lluís Oliva 는 오늘 카탈란 채널 바르셀로나 몬트멜로 라운드에서 열릴 포물러 원의 다섯번째 경기를 스페인 나머지 지역에서도 5 월 첫째 주말에 볼 수 있기 위해 "할 수 있는 모든 방법을 동원했다"라고 설명했다.</p>	<p>TV3 감독, 루이스 올리바, 오늘 설명 : 카탈루냐어 채널을 스페인어 국가의 나머지 월의 첫 주말, 다섯 번째 시험을 볼 수있는 목적으로 "했습니다 길쭉한 손으로 모든 것을 갖추고 있습니다" 포물러 원 포물러 다섯 번째 라운드는 바르셀로나 카탈루냐 서킷에서 개최됩니다.</p>		
<p>그렇지만, 녹화 테이프는 대통령 우파의 후보자이고 10 월 8 일에 있을 선거에서 재선거를 원하는 현재 주지사의 라이벌인 Marian Krzaklewski 의 손으로 갔다.</p>	<p>그러나, 테이프 녹음 마리안 Krzaklewski, 대통령과 국가의 현재 머리 라이벌에 맞는 후보의 손에서왔다. 마리안 Krzaklewski 년 10 월 8 설정 선거에 재선을 추구.</p>		

<p>운전자 안토니오 무뇨스 세라노(Antonio Muñoz Serrano)는 작년 6월 5일 허가증이 요구되는 4륜거 타다가 바르셀로나의 산보이 드 요브레갓(Sant Boi de Llobregat)에서 사고를 당했고 120,000 페세타 벌금 및 1년간 면허 정지에 처해졌다.</p>	<p>드라이버, 안토니오 무 노즈 세라노는 120,000 pesetas 의 벌금을 선고하고 년 동안 순환 라이선스를 금지하고 있습니다. 산트 Boi 드 Llobregat (바르셀로나)의 여정을 자신의 쿼드 운전하는 동안 안토니오 무 노즈 세라노는 작년 사고 6월 5일의 아침을 입었다. 당신의 쿼드를 사용하면 만 번호판이 필요합니다.</p>		
<p>첫 번째 줄의 팀 경기는 시간과 지역적으로 동쪽이다.</p>	<p>팀은 첫 번째 열에과 지역에서 재생됩니다.시간은 동부입니다.</p>		
<p>Raúl 이 여전히 필드로 뛰지는 않지만, 스트라이커가 경기를 할 수 있을 것이라고 모두 겨냥하고 있다.</p>	<p>라울 한편, 여전히 현장에 뛰어하지 않습니다. 그러나, 공격수가 재생됩니다 것으로 보인다.</p>		
<p>Fullana 는 스위스로 은메달이 되돌아가갈 Barbara Blatter 의 27 초 앞서 1 시간 올림픽 코스인 35,7 킬로미터를 49 분 24 초로, 이탈리아 챔피언의 33 초를 목표로 들어왔다.</p>	<p>풀라나 목표는 이탈리아 챔피언에 33 초을했다. 이탈리아 챔피언 바바라 블래터 이상 27 초의 리드와 함께 한 시간, 49 분에서 24 초 올림픽 시간에 35.7 킬로미터 실행을 덮었다. 바바라 블래터는 실버 메달과 함께 스위스로 돌아갑니다.</p>		
<p>첫 수상자 리가 Pauleta 가 첫 타이틀을 얻기 위해 그 팀이 가질 수 있는 견해를 말했다: "우리는 잘 맞고 우승을 위해 필요한 조건을 가지고 있다."</p>	<p>컴퓨터가 처음으로 리그 우승을 할 수있는 옵션에 관하여, 파울레타는 말했다 : "우리는 그것을받을 자격이 우리는 그것을 위해 필요한 조건이 있습니다.."</p>		
<p>Manuela de Madre 카탈란 지방 의회에서 선출된 여자 시장 뿐만 아니라 L'Hospitalet de Llobregat 의 시장 Celestino Corbacho 둘 다 7 월 의회까지 당을 이끌어야 하는 Manuel Chaves 에 의해 이끌어지는 정치 위원회에서 카탈란 사회주의자들을 대표하고 있다.</p>	<p>어머니 마누 외에 카탈로니아 의회에서 부시장입니다. 마누 엘라 드 마드레와 르 호스 피탈 레 데 로브 레 가트, Celestino Corbacho 의 시장이, 정치위원회의 카탈로니아 어 사회주의의 두 대표이며, 마누엘 Chave 에 의해 의장. 정치적 약속은 7 월 의회에 파티를 리드해야한다.</p>		

<p>발렌시아 선수는 따르는 과정을 결정한 Villarrubia 박사가 후에 분석했던 TAC (컴퓨터 단층 촬영)를 담당하기 위해 클럽 의료 서비스의 팀장인 Jorge Candell 과 동행하여 화요일에 바르셀로나로 여행했다.</p>	<p>발렌시아 선수는 화요일에 바르셀로나 여행, 클럽 의료 서비스의 머리와 함께, 호르헤 CANDEL 우리는 이후 박사 Villarrubia 를 분석 CT 검사 (컴퓨터 단층 촬영)를 받아야합니다. 박사 Villarrubia 가 따라야 할 프로세스를 결정했다.</p>		
<p>지난 주 월요일, 형사법원은 "Pototo"와 오늘로써 단식 농성이 59 일째 되는 그의 예전 부책임자 프랑수아 Daniel Derguy 에게 테러리스트 목적과 관련 있는 범죄에 대한 10 년의 감옥형과 최대 벌금을 선고했다.</p>	<p>월요일에, 형사 법원은 "Pototo"와 그의 전 차장, 다니엘 Derguy 프랑수어로, 테러 목적으로 범죄 단체를 들어, 감옥에 십년에 최대 벌금 그를 선고했다. 다니엘 Derguy 은 현재 기아의 파업에 59 일 계재되고 있습니다.</p>		
<p>네델란드 원고에 삽입된 최종 수정이 최근 다른 대표단을 이미 만족시켰다고 설명한 외교관은 "그 두 나라 중 어떤 나라도 마지막까지 법률안을 차단하지 않을 것이라고 본다" 라고 언급했다.</p>	<p>"나도 끝날 때까지 지시문을 차단하는 두 국가가 표시되지 않는다" 외교관은 말했다. 외교관은 초안과의 최신 개정 과목 다른 대표단, 네델란드 만난 설명했다.</p>		
<p>Microsoft 가 Windows 2000 과 Telefónica Data 그리고 Bankinter 를 착수하는 2 월 17 일 부터 열릴 "Info.com/2000" 박람회로 시작될 프로그램은 그 제품들을 소개할 것이고 그 이후로 그 회사를 위한 새로운 기술들을 퍼뜨릴 나라들 중 가장 중요한 도시들로 투어가 실현될 것이다.</p>	<p>이 프로그램은 마이크로 소프트가 Windows 2000 과 텔레포니카 데이터 및 Bankinter 자신의 제품을 선물을 해제 할 수 있는 공정 "Info.com/2000"로 시작합니다. 페어 "Info.com/2000"는 2 17 일부터 개최됩니다. 이후 그들은 국가의 20 개 도시 투어를 했다. 회사에 대한 국가의 20 개 도시 가운데 전파 신기술 .</p>		
<p>사회주의자 국회의원 라몬 하우레기(Ramón Jaúregui)는 "몬클로아에서 Ibarretxe 의 입회(출정)는 민족주의의 극단성을 요구하기 위해 사용되었다"고 고려하고 lehendakari 의 정부는 "지속불가능"하다고 고려한다.</p>	<p>사회 MP 라몬 하우 레기는 고려 "라 몽클 로아의 Ibarretxe 의 존재가 급진적 민족주의를 주장하는 데 사용되었습니다." 라몬 하우 레기 바스크 대통령 정부는 "지속 가능"하다고 생각합니다.</p>		

CEO 는 은행 대표자들이 농촌으로 가서 가게 주인 및 임차인들에게 어떻게 대출 계좌를 지불 할 것인지를 물을 것이다. “성직자에게 어떻게 행동할지”를 설명했다.	CEO 는 설명, 은행 담당자는 현장에 가서 가게 주인과 방법 계산서에게 대출하고 지불하는 세입자에게 “그것은 작동하는 방법 성직자 (제사장).”		
코르도바와 말라기 간의 고속철도 건설 프로젝트는 안달루시아 정부가 제작한 환경충격에 관한 연구를 참조했지만, 개발부는 환경부에게 새로운 성명서를 신청했다.	그는 코르도바와 말라기 간의 고속 열차는 이미위원회에 의해 제조 환경에 미치는 영향 연구가 구축 프로젝트를 설명했다. 그러나 자원부는 환경에 새로운 문을 요청했다.		
이바노브는 또한 Pérez Roque 와 “유익한” 대화 이후 “아주 만족했다”고 말했다.	이바노프는 또한 그가 페레즈 로크와 함께 “결실”대화 후 “매우 만족”라고 말했다.		
“강한 연방 러시아는 다극화 세계와 국제적 활동무대에서 중요한 역할을 한다.” “중요한 국제적 문제점”에 대해 Ivanov 와 일치한다고 강조한 쿠바의 외무부 장관이 말했다.	“강한 미국 러시아가 다극화 세계와 국제 무대에서 중요한 역할을 설정하고있다”쿠바 외교 장관은 말했다. 쿠바 외무 장관 이바노프와의 일치에 말했다, “중요한 국제 문제.”		
포르투갈을 이긴 러시아는 금메달을 위해 싸울 것입니다.	금 라이벌 러시아 될 것입니다. 러시아, 포르투갈을 이길.		
밀항자로 여행했던 바하마 국기가 있는 배 “Wuarta”는 미국을 목적지로 과야킬 (에콰도르)에서부터 지난 15 일 동안 다양한 제품의 짐들과 항해했고 지난주 콜론의 케리비안 항구에서 보급 기항했다.	밀항자로 여행 배, 바하마 플래그 “Wuarta”는 미국의 다른 제품의 부하 하루 15 과야 킬 (에콰도르)에서 항해.배는 콜론의 카리브해 포트에서 지난 주말 보충의 규모했다.		
시범게임에서 합계 총 18 로 패배는 Steinbrenner 의 참을성을 다하게 만들어 이제는 해결책을 원한다.	손실은 이제 시범 경기에서 지금까지 18 총.패배는 스타인브레너의 인내가 부족 만들었습니다. 따라서 스타인브레너 및 솔루션을 원한다.		
파리 증권거래소의 주요 지표는 CAC-40 으로 2.38 퍼센트로 강하게 떨어져 6.105 포인트로 오늘 개최한 반면 손실을 없었지만, 대부분의	한편, 파리 증권 거래소, CAC-40 의 주요 지수는 6,105 포인트 2.38 %의 급락으로 현재 세션을 열었다.추세는 세션의 다량을 통하여 계속했다. 그러나		

세션을 통해 계속되는 경향이였다.	약간 손실이 났아.		
PSC-PSOE 의 서기장 Jaime Blanco 는 비록 사회주의자들 중에서 새로운 리더로서 José Luis Rodríguez Zapatero 의 선거를 위한 결의서에 “조금 불만족”임을 덧붙였지만, 오늘 당의 25 의회의 결과가 “아주 만족스러운” 모습을 나타냈다.	PSC-PSOE, 제임스 화이트의 사무 총장은 이제 파티 XXV 의회의 결과에 “매우 만족”입니다. 그러나 제이미 블랑코 추가 : 그는 사회주의의 새로운 지도자로 호세 루이스 로드리게스 사파테로의 선거에 대해 동의 식으로 “약간 불만족”입니다.		
“늦건 이르건 만약 계속 이 길로 갈다면, 그것을 얻을 것이기 때문에 세비야는 같은 라인에 있다.”라고 언급하여 José Hurtado 는 “바르셀로나가 다양한 기회에 나타났다”고 기억했다.	호세 Hurtado 회상 : “바르셀로나는 여러 경우에 표시됩니다.” 따라서, 호세 Hurtado 는 “세비야는 같은 라인에있는 이유. : 우리는이 길을 계속하면 조만간, 내가 얻을”고 시인했다.		
노란 축구 팀에게 거대한 승리를 안겨줬던 그 경기는 우측 윙에서 Pablo Lago 가 펼쳤던 좋은 경기가 Alvaro 의 헤딩 슈트로 넓혀 졌을 때 진정한 진가로 바뀌었다.	그 장점은 노란색 집합에 큰 가치를 이미. 그 장점은 헤더가 정착 오른쪽 윙에서 파블로라고하여 좋은 움직임과 함께, 알바가 확장 보물로 변모했다.		
테헤란에서는 두번째 선거가 없을 것이다. 왜냐하면 필요 이상 보다 2 명 많은, 32 명의 후보가 이란에서 사용되는 열린 리스트 시스템에 따라 첫 번째 라운드에서 임명되기 위해 요구되는 25 퍼센트 이상의 표를 얻었기 때문이다.	테헤란에는 결선 투표가 없을 것입니다. 이유 : 필요 이상의 32 개 후보가 이란에서 사용되는 열기 목록 체계에 따라, 첫 라운드에서 선출해야하는 투표의 25 % 이상을 수상했다.		
이러한 의미에서, 에콰도르는 작은 국가이고 화폐시장의 완전히 자유로운 “조세 회피국”이 아니라고 확인한다. 왜냐하면 그런 형태의 문제점들을 피하기 위한 규범들을 채택했기 때문이다.	이런 의미에서, 그들이 말하는 : 에콰도르 작은 나라가 아니라 완전히 돈을 시장 확보 “세금 천국”입니다. 이유 : 그들은 이러한 문제를 방지하기 위해 일련의 규칙을 채택했다.		

<p>국회내부위원회에서 PSOE 대변인 Ramón Jáuregui 는 비록 그의 당이 “진보”로 나아가기 위해 정부 수상 José María Aznar 와 바스크자치정부의 수반 Juan José Ibarretxe 사이의 만남을 원한다 할지라도 그의 생각으로는 “깊은 오해”를 낳게 될 것이라고 언급했다.</p>	<p>의회 인테리어, 라몬 하우 레기의위원회에 PSOE 대변인은 말했다 : 인쇄입니다. "깊은 불일치"가있을 것입니다 그러나, 그의 당은 국무 총리, 호세 마리아 아스나르 사이의 만남을 원하고, 바스크 대통령 후안 호세 Ibarretxe 는 GET "진보".</p>		
<p>경제부 장관 Rodrigo Rato 는 오늘 정부가 “정치적 권력”과 “심지어 PSOE 의 자금조달”을 통한 합의의 시대는 지났기 때문에 “항상 공공이익을 지키 위한” BBVA 와 Telefónica 사이의 상호 참여를 실행시킬 것이라고 언급했다.</p>	<p>경제 장관 로드리고 라토, 오늘은 말했다 : "일반적인 관심의 방어에 항상"BBVA 와 텔레포니카 사이의 집행 행위 상호 참가. 이유 : 사라지다 "는 정치 권력에 대한"합의 통과하는 일이며, "심지어 PSOE 자금을 위해."</p>		
<p>아스투리아스의 농업신용은행은 단체 회장인 Román Suárez Blanco 가 지적한 것에 따라 수익성에 대한 그의 “최대 결과”를 파괴할 1999 년 대비 5 퍼센트 상승한 것인 2.237 백만의 순이익을 이번 년도에 달성할 것이라고 예견했다.</p>	<p>CAJA 농촌 데 아스투리아스이 일년에 2,237 만 페세타의 순 이익을 얻을 것으로 예상. 이는 1999 년에 5 %의 증가이다. 이것은 당신의 "레코드"성능을 이길 수, 조직, 로마 수아레스 블랑코의 회장은 말했다.</p>		
<p>며칠전, 남부도시에서 발견된 400 미터 경기를 희망한 22 살의 나이지리안 선수 Hyginus Anugo 는 올림픽 링 근처 차에 치어 상해로 인해 숨졌다.</p>	<p>몇 일전, 나이지리아 선수 Hyginus Anugo 22 올림픽 링 근처에 차에 치되었다. Hyginus Anugo 는 400m 에서 경쟁하기 위해 대기 남부 도시에 있었다. Hyginus Anugo 는 상처 사망.</p>		

Appendix D: Evaluation of the Syntactic Simplification for Spanish-Chinese Machine Translation using Google Translate

Sheet 1 (Direct MT without applying SS):

A	B	C	D
Gold Standard: Human translation	Google Translate	Grammaticality column B (1-5)	Preservation meaning A-B (1-5)
加泰罗尼亚电视 3 台台长路易斯·奥利瓦今日向媒体表示，一级方程式赛车的第五场比赛将在巴塞罗那的蒙特梅洛赛道进行。为使西班牙其他省份能在五月的第一个周末看到这场赛事，这家加泰罗尼亚电视台“已使出浑身解数”。	TV3 主任，奥利瓦 LLUIS，解释今天的加泰罗尼亚通道“具有细长的手一切已”的目标，其余的西班牙国家可以看到 5 月的第一个周末，第五测试一级方程式赛车，在巴塞罗那的加泰罗尼亚赛道举行。		
不过，录音带还是传到了马利安·克萨克萊夫斯基手中。该人是右派总统候选人，是现任总统的主要竞争对手。现任总统希望在十月八日的选举中获得连任。	然而，磁带录音被泄露玛丽安克扎克萊夫斯基，合适的候选人为总统和目前的国家元首，定于 10 月 8 日的选举中寻求连任的主要竞争对手。		
驾驶员安东尼奥·穆尼奥斯·塞拉诺被处以十二万贝赛塔的罚款，并吊销车辆行驶许可证一年。他去年六月五日上午驾驶他的四轮车经过桑博伊德约布雷加特 (Sant Boi de Llobregat) 市 (巴塞罗那) 的一段市区公路时，发生了上述车祸，这种车只须车辆行驶许可证。	驱动程序，安东尼奥·穆尼奥斯·塞拉诺，已被判处 12 万比塞塔的罚款和禁止流通的执照一年，遭遇事故的 6 月 5 日上午，去年而驾驶他四 - 其使用，只需要流通的许可 - 博伊 de Llobregat 的桑特 (巴塞罗那) 的旅程。		
第一列的团队作为地方团队进行比赛，比赛时间以东部时间为准。	球队在第一列在家里，时间是东部。		
至于劳尔，他尚未出场，但一切迹象表明他作为前锋一定会出场。	与此同时，劳尔，仍然不会跳转到球场上，但它的前锋似乎可以发挥。		

<p>富亚娜以与意大利女冠军选手 33 秒之差抵达终点。该意大利选手以 1 小时 49 分 24 秒完成了 35.7 公里的奥运比赛全程，以 27 秒领先瑞士选手芭芭拉·布拉特。芭芭拉将拿着银牌回到瑞士。</p>	<p>富利亚纳的目标是意大利冠军，覆盖 35.7 公里的奥运行程时间一小时，49 分钟和 24 秒 33 秒进入，芭芭拉·布拉特，将返回到瑞士领先 27 秒一枚银牌。</p>		
<p>在谈到自己的团队能否获得他们的首个联赛冠军时，保莱塔表示：“我们当之无愧，我们有获胜所需的一切条件。”</p>	<p>关于选择，您的计算机可能需要得到他们的第一个联赛冠军，保莱塔说，“我们应得的，我们有它的必要条件。”</p>		
<p>曼努埃拉·马德雷和赛莱斯蒂诺·科尔巴乔是加泰罗尼亚社会工人党在政治委员会中的两位代表。前者现任市长和加泰罗尼亚议会议员，后者现任奥斯皮塔莱特德布雷加特市(L'Hospitalet de Llobregat) 市长。政治委员会由曼奴埃尔·查维斯主持，七月份全国代表大会之前党的工作由他领导。</p>	<p>曼努埃拉德马德尔在加泰罗尼亚议会，这也是副市长，市长 L'Hospitalet 的 Llobregat 的，塞莱斯蒂诺 Corbacho，加泰罗尼亚社会党政治委员会的两名代表，由曼努埃尔·查韦斯，谁应该领导直到 7 月国大党。</p>		
<p>巴伦西亚足球队的这位球员已于周二在该俱乐部医疗队主任医师豪尔赫·坎德尔的陪同下抵达巴塞罗那，并进行了电子计算机轴向断层成像 (CAT) 检查。随后比亚鲁比亚医生分析了检查结果，并确定了后续疗程。</p>	<p>瓦伦西亚球员在周二前往巴塞罗那，医疗服务的俱乐部的负责人的陪同下，豪尔赫·坎德尔接受 CT 扫描 (电脑断层扫描) 随后分析的比亚鲁鲁维亚，博士确定过程中遵循。</p>		
<p>上周一，劳教法院对“波托托”及其前代理人法籍人士达涅尔·德尔吉判处十年有期徒刑，这是组织恐怖犯罪的最高量刑。达涅尔绝食抗议今日满 59 天。</p>	<p>周一，刑事法庭判处 10 年徒刑，最高刑罚为恐怖主义目的的阴谋，“pototo”和他的前副手，Daniel Derguy 的法语 - 谁是目前服务 59 天的绝食 - 。</p>		
<p>这位外交官说：“我不认为那两个国家会把这项法律封锁到底”。他解释说草案最后进行的修订已满足了另一个持保留立场的代表团的要求，这个代表团是荷兰代表团。</p>	<p>“我看不出有阻塞的指令，直到最后这两个国家，”这位外交官说，他解释说，最新的修订草案，并已达到另一个代表团少言寡语，荷兰。</p>		

<p>这个项目将以“Info.com/2000”展销会的方式启动。该展销会 2 月 17 日开幕，届时微软公司将推出 Windows 2000 操作系统，电信公司（Telefónica Data）和网上银行（Bankinter）也将展出他们的产品。随后还将在全国二十个重要城市进行巡回展出，推广一系列企业实用新技术。</p>	<p>微软将在 Windows 2000 和 Telefónica 数据和 BANKINTER 的发布将展示他们的产品，其后须为旅游的 20 个程序将开始一个公平的“Info.com/2000”从 2 月 17 日举行全国各大中城市，为公司推广新技术。</p>		
<p>社会工人党议员拉蒙·豪雷吉表示：“伊瓦雷切（Ibarretxe）蒙克洛阿宫（Moncloa）之行已被用于表达民族激进主义诉求”。他同时表示这位巴斯克自治区主席的执政是“难以持久的”。</p>	<p>拉蒙·豪雷吉社会主义 MP 认为“在 La 蒙克洛亚存在 Ibarretxe 的使用要求激进的民族主义”，并认为政府的巴斯克总统是“不可持续的”。</p>		
<p>执行董事长指出，银行代表们将下乡向店主和承租人询问贷款申请人是如何偿付贷款的，并“询问神父（牧师）人们的表现如何”。</p>	<p>CEO 解释说，去外地银行代表，并要求店主和租户如何支付账单借款和“牧师（祭司），它的行为如何。”</p>		
<p>该人解释说，关于科尔多瓦与马拉加之间高速铁路的修建计划，安达卢西亚区政府已提出一份环境影响研究报告。不过，建设部先前已请求环境保护部提供一份新的评估报告。</p>	<p>他解释说，科尔多瓦和马拉加之间的高速列车项目建设已经有一个由董事会编制的环境影响研究，但公安部工程环境要求一个新的语句。</p>		
<p>伊万诺夫表示，在与佩雷斯·罗克进行“有益的”会谈后，他感到“非常满意”。</p>	<p>伊万诺夫还表示，他“非常满意”“富有成果”的对话与佩雷斯后。</p>		
<p>古巴外交部长表示，“一个强大而统一的俄罗斯注定在这个多极世界和国际博弈中扮演重要角色。他同时强调在“诸多国际重要问题”上他与伊万诺夫均有共识。</p>	<p>“一个强大和团结的俄罗斯被设置在一个多极化的世界，并在国际舞台上发挥着重要作用，”古巴外长说，“重要的国际问题与伊万诺夫强调了他的同意。”</p>		
<p>争夺金牌的对手将是俄罗斯，它已经战胜了葡萄牙。</p>	<p>黄金的对手将是俄罗斯，谁击败了葡萄牙。</p>		
<p>他们偷渡的那条船是“瓦尔塔号”，挂着巴哈马国旗。船是 15 日从厄瓜多尔的瓜亚基尔起航的，上面装有多种运</p>	<p>旅游偷渡船，的“Wuarta”巴哈马国旗，从瓜亚基尔（厄瓜多尔）第 15 天航行了与美国不同的产品负载，周末的规模</p>		

往美国的货物，上周末在加勒比海的哥伦布港进行了补给停留。	结肠癌在加勒比海港口补给。		
表演赛已输了 18 次，这使史坦布伦纳失去了耐性。他想解决存在的问题。	的损失，现在共有 18 到目前为止，在表演赛 Steinbrenner 的耐心耗尽，让你想要的解决方案。		
同时，巴黎证券交易所股票主要指数（CAC-40 指数）今天开盘大幅下跌了 2.38%，报 6,105 点，这一走势持续了盘中大部分时间，尽管后来跌幅有所收窄。	同时，巴黎证券交易所，法国 CAC-40，主要指数今天大幅下跌 2.38% 至 6,105 点，这一趋势继续通过许多会话会议开幕虽然有点损失抹去。		
坎塔布里亚自治区社会工人党秘书长海梅·布兰科今天表示，他对第 25 届党代会的结果表示“非常满意”，唯一的“小遗憾”是大会为选举何塞·路易斯·罗德里格斯·萨帕特罗为社会工人党新领袖所确定的程序。	一般司司长的 PSC-PSOE，詹姆斯·怀特，是现在“非常满意”与党的代表大会二十五的结果，但是，他是一个“有点不满”的公式同意选萨帕特罗作为社会党的新领导人。		
何塞·乌尔塔多提醒大家“巴塞罗那已经申请了好多次”，因此他主张“塞维利亚也申请，因为如果我们这样做，塞维利亚迟早会申请到。”	何塞·乌尔塔多说：“巴塞罗那是所谓的”塞维利亚是在同一行，因为迟早，如果我们继续在这条道路上，你会得到“好几次”。		
那种优势对黄队非常有用。当阿尔瓦罗用一个顶球扩大了它的作用时，它变成了真正的法宝。巴勃罗·拉戈从右路传过球来，阿尔瓦罗用头将球顶入门中。	这种优势，这是很有价值的黄集，成为一个真正的宝藏阿尔瓦罗扩大时，用头，解决一个很好的举措从右翼由巴勃罗·拉戈。		
根据伊朗所实行的候选人名单开放制度，德黑兰将不会举行第二轮选举，因为 32 位候选人（比要求人数多出 2 位）在第一轮选举中获得的选票已超过委任所需选票的百分之二十五。	在德黑兰没有径流选举，32 名候选人，两个超过必要的，已经赢得了超过 25% 的选票，须在第一轮选举，根据在伊朗使用到开放列表系统。		
在这个意义上讲，他们认为厄瓜多尔是一个小国家，而不是一个完全开放其货币市场的“避税天堂”。因此，他们已经出台了一系列规定以防发生此类问题。	在这个意义上说，厄瓜多尔是一个小国，而不是一个“避税天堂”，完全释放的货币市场，因为他们已经采取了一系列的规则，以防止此类问题。		

<p>社会工人在国会内政委员会的发言人拉蒙·豪雷吉表示，他们党希望政府首相何塞·玛丽亚·阿斯纳尔与巴斯克自治区主席胡安·何塞·伊瓦雷切 (Juan José Ibarretxe)的会谈在有关方面取得“进展”，但他个人的感觉是会谈将出现“深刻分歧”。</p>	<p>西班牙工人社会党在国会内部，拉蒙·豪雷吉委员会发言人说，而他的党希望的会议之间首相何塞·玛丽亚·阿斯纳尔，和巴斯克总统胡安·何塞·Ibarretxe 的，获得“进步”印刷是会有一个“深刻分歧”。</p>		
<p>经济部长罗德里戈·拉托今天表示，在毕尔巴鄂比斯开银行 (BBVA) 与电信公司 (Telefónica) 之间的相互参股中，政府的作为将是“永远捍卫全局利益”，因为一切协议均通过“政治权力”甚至“通过社会工人工党资助”才能达成的时代已经成为过去。</p>	<p>经济部长罗德里戈·拉托今天说，政府将采取行动 BBVA 和西班牙电信之间的相互参与“总是在防守的整体利益”，“政治权力，因为它已经通过时间，通过协议“E”甚至 PSOE 融资“。</p>		
<p>根据阿斯图里亚斯农业储蓄银行行长罗曼·苏亚雷斯·布兰科提供的信息，该行预计在本财政年度可获得二十二亿三千七百万贝赛塔的净利润，涨幅比1999年增加百分之五，再度打破它的盈利“纪录”。</p>	<p>卡哈农村阿斯图里亚斯预计今年净利润为 2,237 万欧元，比 1999 年增长了 5%，这将击败你的“记录”的表现，称该组织总裁，罗马苏亚雷斯布兰科。</p>		
<p>几天前，22 岁的尼日利亚田径运动员希吉纳斯·阿努戈，在奥运主场管区附近被一辆小轿车撞倒，伤势严重，不治身亡。当时他住在上述南半球城市等待参加 400 米竞赛。</p>	<p>几天前，尼日利亚运动员 Anugo 的 22 团所属，谁是在南部城市等待在 400 米竞争，击中一辆汽车靠近奥林匹克环，死于创伤。</p>		

Sheet 2 (Direct SS+MT):

A	B	C	D
Gold Standard: Human translation	Google Translate	Grammaticality column B (1-5)	Preservation meaning A-B (1-5)
<p>加泰罗尼亚电视 3 台台长路易斯·奥利瓦今日向媒体表示，一级方程式赛车的第五场比赛将在巴塞罗那的蒙特梅洛赛道进行。为使西班牙其他省份能在五月的第一个周末看到这场赛事，这家加泰罗尼亚电视台“已使出浑身解数”。</p>	<p>TV3 主任，奥利瓦 LLUIS，今天解释说：加泰罗尼亚通道“具有细长的手一切已”的目标，其余的西班牙国家可以看到 5 月的第一个周末，第五测试一级方程式一级方程式第五轮将在巴塞罗那的加泰罗尼亚赛道举行。</p>		
<p>不过，录音带还是传到了马利安·克拉克莱夫斯基手中。该人是右派总统候选人，是现任总统的主要竞争对手。现任总统希望在十月八日的选举中获得连任。</p>	<p>然而，磁带录音被泄露玛丽安克拉克莱夫斯基，目前国家元首的总统和主要竞争对手的合适人选。克拉克莱夫斯基玛丽安寻求连任集 10 月 8 日。</p>		
<p>驾驶员安东尼奥·穆尼奥斯·塞拉诺被处以十二万贝塞塔的罚款，并吊销车辆行驶许可证一年。他去年六月五日上午驾驶他的四轮车经过桑博伊德约布雷加特 (Sant Boi de Llobregat) 市 (巴塞罗那) 的一段市区公路时，发生了上述车祸，这种车只须车辆行驶许可证。</p>	<p>安东尼奥·穆尼奥斯·塞拉诺的驱动程序，已被判处一年 12 万比塞塔的罚款和禁止流通的许可。安东尼奥·穆尼奥斯·塞拉诺遭受事故发生在去年 6 月 5 日上午，而驾驶他的四圣 BOI de Llobregat 的 (巴塞罗那) 的旅程。用你的四只需要移动牌照。</p>		
<p>第一列的团队作为地方团队进行比赛，比赛时间以东部时间为准。</p>	<p>在第一列和当地球队。时间是东部。</p>		
<p>至于劳尔，他尚未出场，但一切迹象表明他作为前锋一定会出场。</p>	<p>与此同时，劳尔，仍然不跳球场。然而，它似乎可以发挥的前锋。</p>		
<p>富亚娜以与意大利女冠军选手 33 秒之差抵达终点。该意大利选手以 1 小时 49 分 24 秒完成了 35.7 公里的奥运比赛全程，以 27 秒领先瑞士选手芭芭拉·布拉特。芭芭拉将拿着银牌回到瑞士。</p>	<p>富利亚纳的目标来到 33 秒到意大利的冠军。意大利的冠军涵盖奥运行程 35.7 英里 1 小时，49 分钟和 24 秒的时候，领先 27 秒芭芭拉·布拉特。芭芭拉·布拉特回到瑞士一枚银牌。</p>		

<p>在谈到自己的团队能否获得他们的首个联赛冠军时，保莱塔表示：“我们当之无愧，我们有获胜所需的一切条件。”</p>	<p>关于选项，您的计算机可能需要得到他们的第一个联赛冠军，保莱塔说：“这是我们应得的，我们有它的必要条件。”</p>		
<p>曼努埃拉·马德雷和塞莱斯蒂诺·科尔巴乔是加泰罗尼亚社会工人在政治委员会中的两位代表。前者现任市长和加泰罗尼亚议会议员，后者现任奥斯皮塔莱特德约布雷加特市(L'Hospitalet de Llobregat) 市长。政治委员会由曼奴埃尔·查维斯主持，七月份全国代表大会之前党的工作由他领导。</p>	<p>“除了母亲曼努埃拉是在加泰罗尼亚议会的副市长。曼努埃拉德马德尔和L'Hospitalet 的 Llobregat 的，塞莱斯蒂诺 Corbacho 市长，加泰罗尼亚社会主义者在政治委员会的两名代表，由曼努埃尔·查维斯主持。应带领党的政治承诺的 7 月国会。</p>		
<p>巴伦西亚足球队的这位球员已于周二在该俱乐部医疗队主任医师豪尔赫·坎德尔的陪同下抵达巴塞罗那，并进行了电子计算机轴向断层成像 (CAT) 检查。随后比亚鲁比亚医生分析了检查结果，并确定了后续疗程。</p>	<p>瓦伦西亚球员在周二前往巴塞罗那，医疗服务的俱乐部的负责人的陪同下，豪尔赫·坎德尔接受 CT 扫描 (电脑断层扫描) 随后比亚鲁维亚博士分析。博士，比亚鲁维亚确定过程中遵循。</p>		
<p>上周一，劳教法院对“波托托”及其前代理人法籍人士达涅尔·德尔吉判处十年有期徒刑，这是组织恐怖犯罪的最高量刑。达涅尔绝食抗议今日满 59 天。</p>	<p>周一，刑事法院判处十年徒刑，以恐怖主义为目标的犯罪团伙的最高刑罚，“pototo”和他的前副手，丹尼尔 Derguy 法国。丹尼尔 Derguy 目前服 59 天的绝食。</p>		
<p>这位外交官说：“我不认为那两个国家会把这项法律封锁到底”。他解释说草案最后进行的修订已满足了另一个持保留立场的代表团的要求，这个代表团是荷兰代表团。</p>	<p>“我看不出有阻塞的指令，直到最后这两个国家之一，”这位外交官说。这位外交官解释说：最新的修订草案，并已达到另一个代表团少言寡语，荷兰。</p>		
<p>这个项目将以“Info.com/2000”展销会的方式启动。该展销会 2 月 17 日开幕，届时微软公司将推出 Windows 2000 操作系统，电信公司 (Telefónica Data) 和网上银行 (Bankinter) 也将展出他们的产品。随后还将在全国二十个重要城市进行巡回展出，推广一系列企业实用新技术。</p>	<p>该计划将开始一个公平的“Info.com/2000”中，微软将发布 Windows 2000 和西班牙电信数据和 BANKINTER 的展示他们的产品。将公平“Info.com/2000”从 2 月 17 日举行，此后，他们在全国 20 个城市参观，其中在全国 20 个城市传播新技术，为公司。</p>		

<p>社会工人党议员拉蒙·豪雷吉表示：“伊瓦雷切（Ibarretxe）蒙克洛阿宫（Moncloa）之行已被用于表达民族激进主义诉求”。他同时表示这位巴斯克自治区主席的执政是“难以持久的”。</p>	<p>社会主义 MP 拉蒙·豪雷吉认为：“在 La 蒙克洛亚 Ibarretxe 的存在已被用来要求激进的民族主义”。拉蒙·豪雷吉认为：巴斯克总统的政府是“不可持续的”。</p>		
<p>执行董事长指出，银行代表们将下乡向店主和承租人询问贷款申请人是如何偿付贷款的，并“询问神父（牧师）人们的表现如何”。</p>	<p>CEO 解释说，银行的代表去到现场，并要求店主和租户如何支付账单借款和“牧师（祭司），它的行为如何。”</p>		
<p>该人解释说，关于科尔多瓦与马拉加之间高速铁路的修建计划，安达卢西亚区政府已提出一份环境影响研究报告。不过，建设部先前已请求环境保护部提供一份新的评估报告。</p>	<p>他解释说，项目建设高速列车科尔多瓦和马拉加之间已经有一个由董事会编制的环境影响研究。然而，工信部已要求发展环境的一个新的语句。</p>		
<p>伊万诺夫表示，在与佩雷斯·罗克进行“有益的”会谈后，他感到“非常满意”。</p>	<p>伊万诺夫还表示，他“非常满意”“富有成果”的对话与佩雷斯后。</p>		
<p>古巴外交部长表示，“一个强大而统一的俄罗斯注定在这个多极世界和国际博弈中扮演重要角色。他同时强调在“诸多国际重要问题”上他与伊万诺夫均有共识。</p>	<p>古巴外长说：“一个强大和团结的俄罗斯被设置在一个多极化的世界，并在国际舞台上发挥着重要的作用，”。古巴外长伊万诺夫说：“重要的国际问题上的巧合。”</p>		
<p>争夺金牌的对手将是俄罗斯，它已经战胜了葡萄牙。</p>	<p>黄金的对手将是俄罗斯。俄罗斯击败葡萄牙。</p>		
<p>他们偷渡的那条船是“瓦尔塔号”，挂着巴哈马国旗。船是 15 日从厄瓜多尔的瓜亚基尔起航的，上面装有多种运往美国的货物，上周末在加勒比海的哥伦布港进行了补给停留。</p>	<p>作为偷渡者的船前往，的“Wuarta”巴哈马国旗，从瓜亚基尔（厄瓜多尔），航行 15 天，与美国不同的产品负载。船在加勒比海港口科隆上周末补货规模。</p>		
<p>表演赛已输了 18 次，这使史坦布伦纳失去了耐性。他想解决存在的问题。</p>	<p>现在的损失共 18 到目前为止，在表演赛。击败 Steinbrenner 的耐心耗尽。因此史坦伯纳和希望的解决方案。</p>		

同时，巴黎证券交易所股票主要指数（CAC-40 指数）今天开盘大幅下跌了 2.38%，报 6,105 点，这一走势持续了盘中大部分时间，尽管后来跌幅有所收窄。	同时，巴黎证券交易所，法国 CAC-40，主要指数今天大幅下跌 2.38% 至 6,105 点宣布会议开幕。这一趋势继续下去整个的会议。然而，有些损失抹去。		
坎塔布里亚自治区社会工人党秘书长海梅·布兰科今天表示，他对第 25 届党代会的结果表示“非常满意”，唯一的“小遗憾”是大会为选举何塞·路易斯·罗德里格斯·萨帕特罗为社会工人党新领袖所确定的程序。	总书记 PSC-PSOE，詹姆斯·怀特，现在 XXV 党的代表大会的结果“非常满意”。然而，海梅·布兰科说：“他是一个”小不满“商定的公式为社会主义者的新领导人选举何塞·路易斯·罗德里格斯·萨帕特罗。		
何塞·乌尔塔多提醒大家“巴塞罗那已经申请了好多次”，因此他主张“塞维利亚也申请，因为如果我们这样做，塞维利亚迟早会申请到。”	何塞·乌尔塔多回忆说：“巴塞罗那是好几次”。因此，何塞·乌尔塔多恳求说：“塞维利亚是在同一行的原因：早晚，如果我们继续在这条道路上，我得到”。		
那种优势对黄队非常有用。当阿尔瓦罗用一个顶球扩大了它的作用时，它变成了真正的法宝。巴勃罗·拉戈从右路传过球来，阿尔瓦罗用头将球顶入门中。	这种优势已经是很有价值的黄集。这种优势转变成一个真正的宝藏阿尔瓦罗扩大时，用头，解决一个很好的举措从右翼由巴勃罗·拉戈。		
根据伊朗所实行的候选人名单开放制度，德黑兰将不会举行第二轮选举，因为 32 位候选人（比要求人数多出 2 位）在第一轮选举中获得的选票已超过委任所需选票的百分之二十五。	在德黑兰将不会有决选。原因：32 项提名，两个超过必要的，已经赢得了超过 25% 的，须在第一轮选举的投票，根据在伊朗使用到开放列表系统。		
在这个意义上讲，他们认为厄瓜多尔是一个小国家，而不是一个完全开放其货币市场的“避税天堂”。因此，他们已经出台了一系列规定以防发生此类问题。	在这个意义上说，他们说：厄瓜多尔是一个小国，不是“避税天堂”的货币市场完全释放。原因：他们已经采取了一系列的规则，以防止此类问题。		
社会工人党在国会内政委员会的发言人拉蒙·豪雷吉表示，他们党希望政府首相何塞·玛丽亚·阿斯纳尔与巴斯克自治区主席胡安·何塞·伊瓦雷切（Juan José Ibarretxe）的会谈在有关方面取得“进展”，但他个人的感觉是会谈将出	西班牙工人社会党在国会内部，拉蒙·豪雷吉，委员会的发言人说：“印刷是：会有一个”深刻的分歧。“然而，他的党之间的首相何塞·玛丽亚·阿斯纳尔，希望会议巴斯克总统胡安·何塞·Ibarretxe，“进步”。		

现“深刻分歧”。			
经济部长罗德里戈·拉托今天表示，在毕尔巴鄂比斯开银行（BBVA）与电信公司（Telefónica）之间的相互参股中，政府的作为将是“永远捍卫全局利益”，因为一切协议均通过“政治权力”甚至“通过社会工人党资助”才能达成的时代已经成为过去。	经济部长罗德里戈·拉托今天说，行政行为相互参加 BBVA 和西班牙电信之间的“普遍利益”总是在防守。原因：飘协议通过的“政治权力”和“即使融资的西班牙工人社会党”。		
根据阿斯图里亚斯农业储蓄银行行长罗曼·苏亚雷斯·布兰科提供的信息，该行预计在本财政年度可获得二十二亿三千七百万贝赛塔的净利润，涨幅比1999年增加百分之五，再度打破它的盈利“纪录”。	卡哈农村阿斯图里亚斯预计今年净利润为 2,237 万比塞塔。这是一个比 1999 年增长了 5%。这会打败你的“记录”的表现，总统说的组织，罗马苏亚雷斯布兰科。		
几天前，22 岁的尼日利亚田径运动员希吉纳斯·阿努戈，在奥运主场管区附近被一辆小轿车撞倒，伤势严重，不治身亡。当时他住在上述南半球城市等待参加 400 米竞赛。	几天前，尼日利亚选手团所属 Anugo22 奥运五环附近的一辆汽车被击中。团所属 Anugo 是在南部城市等待争夺 400 米。团所属 Anugo 去世的伤口。		

Appendix E: Reading comprehension test for foreigners learning Spanish

A1

Gracias por participar en este proyecto de investigación. Por favor, rellene los siguientes datos antes de comenzar.

Lengua materna:

Sexo:

Edad:

Nivel de Español:

Básico (A1-A2) – Intermedio (B1-B2) – Avanzado (C1-C2)

Por favor, lea el siguiente texto periodístico y conteste las preguntas a continuación:

El jefe de la Oficina de Seguridad Nacional y ministro de la Presidencia polaca, Marek Siwiec, presentó hoy su dimisión de todos los cargos después de que varias cadenas de televisión locales transmitieran anoche un vídeo con imágenes en las que se burla del papa Juan Pablo II, imitando los gestos del Sumo Pontífice cuando llega a un país en peregrinación. "Presento mi dimisión, porque como político he perdido todo el prestigio que tenía y pido perdón a todos

aquellos a los que ofendí con mi irresponsable e irrespetuoso comportamiento”, dijo el ministro Siwiec en un comunicado entregado a los medios.

El vídeo fue grabado hace tres años, durante una visita a la ciudad de Kalisz del presidente de la República Aleksander Kwasniewski, al que acompañó el ministro Siwiec. Al descender del avión Siwiec hizo el gesto de la bendición y, cuando ya pisó tierra el presidente, según se escucha en la cinta, le dijo que se había olvidado de besar la tierra, cosa que su ministro hizo inmediatamente. El suceso tuvo lugar prácticamente en privado, ya que lo presenciaron solamente los acompañantes del presidente, sus escoltas y un grupo muy reducido de personalidades de la ciudad visitada. No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado, quien busca la reelección en los comicios fijados para el 8 de octubre.

[...] La burla ha despertado la ira de los polacos más identificados con el papa Wojtyla y ha causado malestar y desagrado incluso entre los partidarios y simpatizantes de Kwasniewski. La dimisión de Siwiec y sus ruegos de perdón por su broma de mal gusto no significarán el final del asunto: las imágenes que han provocado el revuelo han sido incorporadas a un anuncio electoral del candidato Krzaklewski.

Responda a las siguientes preguntas. Marque sólo una casilla. Si no tiene una ligera idea de cuál puede ser la respuesta correcta en alguna pregunta, déjela sin contestar.

1. ¿Quién se burla del Papa?
 - a. El Sumo Pontífice
 - b. El presidente de la República
 - c. El ministro
 - d. Los escoltas del president

2. ¿Quién besó la tierra durante la visita a la ciudad de Kalisz?
 - a. El presidente de la República
 - b. El ministro de la Presidencia polaca
 - c. Un candidato a la presidencia
 - d. El Papa

3. ¿Quién dimitió a causa de estos hechos?
 - a. Kwasniewski
 - b. Wojtyla
 - c. Krzaklewski
 - d. Siwec

4. ¿Qué provocó la ira de los polacos más identificados con el Papa Wojtyla?
 - a. La dimisión del ministro
 - b. La dimisión del presidente
 - c. Que el ministro imitara al Papa durante la visita a Kalisz.
 - d. Que el presidente besara la tierra durante la visita a Kalisz.

5. ¿Quién incluyó el vídeo del suceso en un anuncio electoral?
 - a. El ministro Marek Siwec
 - b. Marian Krzaklewski
 - c. El presidente Aleksander Kwasniewski
 - d. El jefe de la Oficina de Seguridad Nacional

A2

Por favor, lea el siguiente texto periodístico y conteste las preguntas a continuación:

La dirección de la Televisión de Catalunya (TV3) ha asegurado hoy que ha puesto todos los medios para que toda la audiencia española pueda seguir el Gran Premio de España de Fórmula Uno, así como el resto de carreras de esta temporada. El director de TV3, Lluís Oliva, ha explicado hoy que el canal catalán "ha alargado la mano todo lo que ha podido", con el objetivo de que el resto del estado español pueda ver el primer fin de semana de mayo la quinta prueba de Fórmula Uno, que se celebrará en el circuito de Montmeló, en Barcelona.

"No estamos poniendo ningún impedimento para que Televisión Española (TVE) retransmita la prueba, ni que acceda a todo el mundial. El año pasado ya compartimos todas las carreras e, incluso, no solicitamos oscurecer Catalunya para que la audiencia no viera por La 2 la carrera", ha añadido Oliva, quien ha recordado que el canal nacional no quiso transmitir los premios de Silverstone ni San Marino. "Hace cuatro años TV3 hizo una apuesta por la Fórmula Uno. Si hicimos una apuesta, ahora no vamos a hacer partícipes a terceros gratuitamente. El canal vasco (ETB-1) y el valenciano (Canal 9) son los únicos que fueron sensibles con nosotros, pero no TVE ", ha aclarado Oliva. Según el director del canal catalán, TVE" incumplió parte del contrato" suscrito la temporada pasada para que el canal nacional ofreciese por La 2 la temporada de Fórmula Uno.

Responda a las siguientes preguntas. Marque sólo una casilla. Si no tiene una ligera idea de cuál puede ser la respuesta correcta en alguna pregunta, déjela sin contestar.

6. ¿Qué cadena/s de televisión retransmitió el año pasado el mundial de Fórmula Uno en Cataluña?
 - e. La 2
 - f. TV3
 - g. ETB-1 y Canal 9
 - h. La 2 y TV3

7. ¿Con qué cadena/s de televisión muestra su desacuerdo el director de TV3, Lluís Oliva?
 - e. ETB-1 y Canal 9
 - f. TVE, ETB-1 y Canal 9
 - g. TVE
 - h. ETB-1

8. ¿Cuál/es fueron el/los premio/s que La 2 no emitió el año anterior?
 - e. España
 - f. España, Silverstone y San Marino
 - g. Silverstone y San Marino
 - h. Ninguno, emitió todos

B1

Gracias por participar en este proyecto de investigación. Por favor, rellene los siguientes datos antes de comenzar.

Lengua materna:

Sexo:

Edad:

Nivel de Español:

Básico (A1-A2) – Intermedio (B1-B2) – Avanzado (C1-C2)

Por favor, lea el siguiente texto periodístico y conteste las preguntas a continuación:

Varias cadenas de televisión locales transmitieron anoche un vídeo con imágenes. En las imágenes, el jefe de la Oficina de Seguridad Nacional y ministro de la Presidencia polaca, Marek Siwiec se burla del papa Juan Pablo II, imitando los gestos del Sumo Pontífice cuando llega a un país en peregrinación. Después, el jefe de la Oficina de Seguridad Nacional y ministro de la Presidencia polaca, Marek Siwiec, presentó hoy su dimisión de todos los cargos. "Presento mi dimisión, porque como político he perdido todo el prestigio que tenía. Pido perdón a todos aquellos a los que ofendí con mi irresponsable e irrespetuoso comportamiento", dijo el ministro Siwiec en un comunicado entregado a los medios.

El vídeo fue grabado hace tres años, durante una visita a la ciudad de Kalisz del presidente de la República Aleksander Kwasniewski. El ministro Siwiec acompañó al presidente

de la República Aleksander Kwasniewski. Al descender del avión Siwiec hizo el gesto de la bendición. Cuando ya pisó tierra el presidente, según se escucha en la cinta, le dijo que se había olvidado de besar la tierra, cosa que su ministro lo hizo inmediatamente. El suceso tuvo lugar prácticamente en privado, ya que lo presenciaron solamente los acompañantes del presidente, sus escoltas y un grupo muy reducido de personalidades de la ciudad visitada. No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado. Marian Krzaklewski busca la reelección en los comicios fijados para el 8 de octubre.

[...] La burla ha despertado la ira de los polacos más identificados con el papa Wojtyla. La burla ha causado malestar y desagrado incluso entre los partidarios y simpatizantes de Kwasniewski. La dimisión de Siwiec y sus ruegos de perdón por su broma de mal gusto no significarán el final del asunto: Las imágenes que han provocado el revuelo han sido incorporadas a un anuncio electoral del candidato Krzaklewski.

Responda a las siguientes preguntas. Marque sólo una casilla. Si no tiene una ligera idea de cuál puede ser la respuesta correcta en alguna pregunta, déjela sin contestar.

1. ¿Quién se burla del Papa?
 - a. El Sumo Pontífice
 - b. El presidente de la República
 - c. El ministro
 - d. Los escoltas del presidente

2. ¿Quién besó la tierra durante la visita a la ciudad de Kalisz?
 - a. El presidente de la República
 - b. El ministro de la Presidencia polaca
 - c. Un candidato a la presidencia
 - d. El Papa

3. ¿Quién dimitió a causa de estos hechos?
 - a. Kwasniewski
 - b. Wojtyła
 - c. Krzaklewski
 - d. Siwec

4. ¿Qué provocó la ira de los polacos más identificados con el Papa Wojtyła?
 - a. La dimisión del ministro
 - b. La dimisión del presidente
 - c. Que el ministro imitara al Papa durante la visita a Kalisz.
 - d. Que el presidente besara la tierra durante la visita a Kalisz.

5. ¿Quién incluyó el vídeo del suceso en un anuncio electoral?
 - a. El ministro Marek Siwec
 - b. Marian Krzaklewski
 - c. El presidente Aleksander Kwasniewski
 - d. El jefe de la Oficina de Seguridad Nacional

A2

Lea el siguiente texto periodístico y conteste las preguntas a continuación:

La dirección de la Televisión de Catalunya (TV3) ha asegurado hoy: la dirección de la Televisión de Catalunya (TV3) ha puesto todos los medios para que toda la audiencia española pueda seguir el Gran Premio de España de Fórmula Uno, así como el resto de carreras de esta temporada. El director de TV3, Lluís Oliva, ha explicado hoy: el canal catalán "ha alargado la mano todo lo que ha podido" con el objetivo de que el resto del estado español pueda ver el primer fin de semana de mayo la quinta prueba de Fórmula Uno. La quinta prueba de Fórmula Uno se celebrará en el circuito de Montmeló, en Barcelona.

"No estamos poniendo ningún impedimento para que Televisión Española (TVE) retransmita la prueba. Tampoco estamos poniendo ningún impedimento para que Televisión Española (TVE) acceda a todo el mundial. El año pasado ya compartimos todas las carreras. Incluso, no solicitamos oscurecer Catalunya para que la audiencia no viera por La_2 la carrera", ha añadido Oliva. Oliva ha recordado: el canal nacional no quiso transmitir los premios de Silverstone ni San Marino. "Hace cuatro años TV3 hizo una apuesta por la Fórmula Uno. Si hicimos una apuesta, ahora no vamos a hacer partícipes a terceros gratuitamente. El canal vasco (ETB-1) y el valenciano (Canal_9) son los únicos que fueron sensibles con nosotros, pero no TVE ", ha aclarado Oliva. Según el director del canal catalán, TVE" incumplió parte del contrato" suscrito la temporada pasada para que el canal nacional ofreciese por La 2 la temporada de Fórmula Uno.

Responda a las siguientes preguntas. Marque sólo una casilla. Si no tiene una ligera idea de cuál puede ser la respuesta correcta en alguna pregunta, déjela sin contestar.

6. ¿Qué cadena/s de televisión retransmitió el año pasado el mundial de Fórmula Uno en Cataluña?
 - a. La 2
 - b. TV3
 - c. ETB-1 y Canal 9
 - d. La 2 y TV3

7. ¿Con qué cadena/s de televisión muestra su desacuerdo el director de TV3, Lluís Oliva?
 - a. ETB-1 y Canal 9
 - b. TVE, ETB-1 y Canal 9
 - c. TVE
 - d. ETB-1

8. ¿Cuál/es fueron el/los premio/s que La 2 no emitió el año anterior?
 - a. España
 - b. España, Silverstone y San Marino
 - c. Silverstone y San Marino
 - d. Ninguno, emitió todos

Appendix F: Evaluation test set for the SS system + Output sentences + Gold Standard

Input sentences	Output sentences	Gold Standard sentences
Al anunciar su amenaza, el dirigente sindical Thierry Durr, dijo que "está de moda".	Al anunciar su amenaza, el dirigente sindical Thierry Durr, dijo: "Está de moda".	Al anunciar su amenaza, el dirigente sindical Thierry Durr, dijo: "Está de moda".
No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado, quien busca la reelección en los comicios fijados para el 8 de octubre.	No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado. Marian Krzaklewski busca la reelección en los comicios fijados para el 8 de octubre.	No obstante, la cinta con la grabación llegó a manos de Marian Krzaklewski, candidato de la derecha a la Presidencia y principal rival del actual jefe del estado. Marian Krzaklewski busca la reelección en los comicios fijados para el 8 de octubre.
La sentencia subraya que "el índice estadístico anual de accidentes de tráfico causados por ciclomotores en el interior de núcleo urbano resulta aterrador, según las cifras del Instituto Catalán de Seguridad Viaria".	La sentencia subraya: "El índice estadístico anual de accidentes de tráfico causados por ciclomotores en el interior de núcleo urbano resulta aterrador, según las cifras del Instituto Catalán de Seguridad Viaria".	La sentencia subraya: "El índice estadístico anual de accidentes de tráfico causados por ciclomotores en el interior de núcleo urbano resulta aterrador, según las cifras del Instituto Catalán de Seguridad Viaria".
Los equipos en la primera columna juegan como locales y el horario es del Este.	Los equipos en la primera columna juegan como locales. El horario es del Este.	Los equipos en la primera columna juegan como locales. El horario es del Este.
Raúl, por su parte, sigue sin saltar al terreno de juego, pero todo apunta a que el delantero podrá jugar.	Raúl, por su parte, sigue sin saltar al terreno de juego. Sin embargo, todo apunta a que el delantero podrá jugar.	Raúl, por su parte, sigue sin saltar al terreno de juego. Sin embargo, todo apunta a que el delantero podrá jugar.
Mañana se disputará en este mismo circuito la prueba masculina de esta disciplina, en la que se darán dos vueltas más que las dadas hoy por Fullana y sus rivales.	Mañana se disputará en este mismo circuito la prueba masculina de esta disciplina. En la prueba masculina de esta disciplina se darán dos vueltas más que las dadas hoy por Fullana y sus rivales.	Mañana se disputará en este mismo circuito la prueba masculina de esta disciplina. En la prueba se darán dos vueltas más que las dadas hoy por Fullana y sus rivales.
A la ausencia por sanción de Pauleta se unen, por idéntico motivo, las de Mauro Silva y Flavio, mientras que Djalminha, Bassir y, probablemente, Naybet, serán bajas por lesión.		
Patrick Kluyvert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato y ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.	Patrick Kluyvert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato. Ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.	Patrick Kluyvert, autor del primer tanto del equipo azulgrana, aumentó a 13 su cuenta en el presente campeonato. Patrick Kluyvert ya es líder en solitario de la clasificación de goleadores del Barcelona en la Liga 1999-2000, con uno de ventaja sobre Rivaldo.
"El dolor ha sido crónico a lo largo de todo este tiempo y además le impide realizar ciertos movimientos".	"El dolor ha sido crónico a lo largo de todo este tiempo. Además le impide realizar ciertos movimientos".	"El dolor ha sido crónico a lo largo de todo este tiempo. Además, el dolor le impide realizar ciertos movimientos".
El pasado lunes, el Tribunal Correccional condenó a diez años de cárcel, pena	El pasado lunes, el Tribunal Correccional condenó a diez años de	El pasado lunes, el Tribunal Correccional condenó a diez años de

¹ Not simplified: Output sentence = Input Sentence (No simplification carried out by the system)

² Not simplified: Gold Standard sentence = Input Sentence (No simplification needed)

máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy - que hoy cumple 59 días en huelga de hambre -.	cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. El francés Daniel Derguy hoy cumple 59 días en huelga de hambre.	cárcel, pena máxima por asociación de malhechores con fines terroristas, a "Pototo" y a su antiguo lugarteniente, el francés Daniel Derguy. Daniel Derguy hoy cumple 59 días en huelga de hambre.
Los últimos borradores presentados por la presidencia francesa de la UE distinguen, por un lado, entre las operaciones financieras, inmobiliarias y económicas en general y, por otro, las misiones de representación judicial y asistencia jurídica, que quedarían protegidas, según otras fuentes comunitarias.	Los últimos borradores presentados por la presidencia francesa de la UE distinguen, por un lado, entre las operaciones financieras, inmobiliarias y económicas en general y, por otro, las misiones de representación judicial y asistencia jurídica. Las misiones de representación judicial y asistencia jurídica quedarían protegidas, según otras fuentes comunitarias.	Los últimos borradores presentados por la presidencia francesa de la UE distinguen, por un lado, entre las operaciones financieras, inmobiliarias y económicas en general y, por otro, las misiones de representación judicial y asistencia jurídica. Las misiones de representación judicial y asistencia jurídica quedarían protegidas, según otras fuentes comunitarias.
Microsoft, Telefónica Data y Bankinter, han desarrollado el programa " Conecta 2000 " que consiste en una serie de iniciativas para comunicar a las empresas las ventajas y oportunidades de Internet.	<i>Not Simplified</i>	<i>Not Simplified</i>
Reiteró, sobre el encuentro entre el presidente Aznar y el lehendakari, que " la postura de Ibarrexe ha sido decepcionante.	<i>Not Simplified</i>	Reiteró, sobre el encuentro entre el presidente Aznar y el lehendakari: "La postura de Ibarrexe ha sido decepcionante.
Los préstamos se otorgan a partir de un análisis que mide la capacidad de pago y el carácter del solicitante.	<i>Not Simplified</i>	<i>Not Simplified</i>
En materia de carreteras, Gutiérrez del Castillo trasladó al ministro la necesidad de acelerar la construcción de las autovías Ruta de la Plata (desde Sevilla al límite con la provincia de Badajoz), Córdoba-Antequera, Granada-Motril y el tramo Nerja-Adra que falta para cerrar la del Mediterráneo, así como la conexión Bahía de Cádiz-Bahía de Algeciras.	<i>Not Simplified</i>	<i>Not Simplified</i>
Ivanov también afirmó que estaba "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.	Ivanov también afirmó: estaba "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.	Ivanov también afirmó estar "muy satisfecho" tras su "provechosa" conversación con Pérez Roque.
El rival por el oro va a ser Rusia, que se impuso a Portugal.	El rival por el oro va a ser Rusia. Rusia se impuso a Portugal.	El rival por el oro va a ser Rusia. Rusia se impuso a Portugal.
Otro medio centenar que llegó a Panamá en octubre pasado, sin recursos económicos ni pasajes de regreso a Ecuador, fue repatriado nada más desembarcar en el aeropuerto internacional de Tocumen, en la capital panameña.	<i>Not Simplified</i>	<i>Not Simplified</i>
Los dobles campeones de la Serie Mundial, los Yanquis de Nueva York, no han podido dar la mejor imagen en su primera cita de la temporada, en el campo de entrenamiento de primavera, donde han estado fuera de control y con el peor récord de todos.	<i>Not Simplified</i>	<i>Not Simplified</i>
Varios centenares de personas acudieron al	Varios centenares de personas	Varios centenares de personas

palacio Brongniart, sede simbólica de la Bolsa parisiense, que amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.	acudieron al palacio Brongniart, sede simbólica de la Bolsa parisiense. La Bolsa parisiense amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.	acudieron al palacio Brongniart, sede simbólica de la Bolsa parisiense. El palacio Brongniart amaneció cubierto con los colores de Euronext, el amarillo y el azul, en forma de bandas situadas entre las catorce columnas de la fachada del edificio.
La eliminación de la candidatura española era, en opinión del Consejero, " una decisión previsible " después " de las manifestaciones de Juan Antonio Samaranch que calificaba la candidatura técnicamente muy buena, pero con muchas dificultades para pasar el corte " .	<i>Not Simplified</i>	<i>Not Simplified</i>
"Pero las ideas del lehendakari no las respetan muchos en este país, y las ideas no se pueden condenar".	"Pero las ideas del lehendakari no las respetan muchos en este país. Las ideas no se pueden condenar".	"Pero las ideas del lehendakari no las respetan muchos en este país. Las ideas no se pueden condenar".
Rafsanyani, que aspiraba a ser el próximo presidente de la Cámara, se presentó como candidato " pragmático " de consenso entre las dos facciones del régimen islámico.	Rafsanyani aspiraba a ser el próximo presidente de la Cámara. Rafsanyani se presentó como candidato " pragmático " de consenso entre las dos facciones del régimen islámico.	Rafsanyani aspiraba a ser el próximo presidente de la Cámara. Rafsanyani se presentó como candidato " pragmático " de consenso entre las dos facciones del régimen islámico.
Un documento de la Policía Antinarcoóticos difundido hoy por el diario " El Universo " señala que las mafias internacionales han usado esta decisión para blanquear dinero o introducir billetes falsos, especialmente de Colombia.	Un documento de la Policía Antinarcoóticos difundido hoy por el diario " El Universo " señala: Las mafias internacionales han usado esta decisión para blanquear dinero o introducir billetes falsos, especialmente de Colombia.	Un documento de la Policía Antinarcoóticos difundido hoy por el diario " El Universo " señala: Las mafias internacionales han usado esta decisión para blanquear dinero o introducir billetes falsos, especialmente de Colombia.
En cuanto a la entrevista entre el copresidente del BBVA, Francisco González, y el presidente del Gobierno, José María Aznar, el también vicepresidente segundo del Gobierno afirmó que es "personal y lógica".	En cuanto a la entrevista entre el copresidente del BBVA, Francisco González, y el presidente del Gobierno, José María Aznar, el también vicepresidente segundo del Gobierno afirmó: es "personal y lógica".	En cuanto a la entrevista entre el copresidente del BBVA, Francisco González, y el presidente del Gobierno, José María Aznar, el también vicepresidente segundo del Gobierno afirmó: La entrevista es "personal y lógica".
El objetivo de Caja Rural al finalizar el plan es contar con 100 oficinas, que permitirán crear 50 nuevos puestos de trabajo.	El objetivo de Caja Rural al finalizar el plan es contar con 100 oficinas. 100 oficinas permitirán crear 50 nuevos puestos de trabajo.	El objetivo de Caja Rural al finalizar el plan es contar con 100 oficinas. Las 100 oficinas permitirán crear 50 nuevos puestos de trabajo.
Dicen que honró el título, primero, antes de oficializarlo, en los terrenos de juego, y después, en la cancha de la vida, en la que dio ejemplo de deportividad.	Dicen que honró el título, primero, antes de oficializarlo, en los terrenos de juego, y después, en la cancha de la vida. En la cancha de la vida dio ejemplo de deportividad.	Dicen que honró el título, primero, antes de oficializarlo, en los terrenos de juego, y después, en la cancha de la vida. En la cancha de la vida dio ejemplo de deportividad.
El tratado de Roma, firmado en 1998, estableció la creación de la Corte Internacional, que estará encargada de juzgar a los autores de genocidio, crímenes de guerra y crímenes contra la humanidad.	El tratado de Roma, firmado en 1998, estableció la creación de la Corte Internacional. La Corte Internacional estará encargada de juzgar a los autores de genocidio, crímenes de guerra y crímenes contra la humanidad.	El tratado de Roma, firmado en 1998, estableció la creación de la Corte Internacional. La Corte Internacional estará encargada de juzgar a los autores de genocidio, crímenes de guerra y crímenes contra la humanidad.
Sobre estas "pistas" se abrirán las negociaciones de la Conferencia Intergubernamental el próximo 14 de febrero, el foro en el que los ministros de Exteriores de la UE, apoyados por un	Sobre estas "pistas" se abrirán las negociaciones de la Conferencia Intergubernamental el próximo 14 de febrero, el foro. En el foro los ministros de Exteriores de la UE,	Sobre estas "pistas" se abrirán las negociaciones de la Conferencia Intergubernamental el próximo 14 de febrero. En el foro los ministros de Exteriores de la UE, apoyados por un

"grupo preparatorio" de altos funcionarios procedentes de las quince capitales, negociarán la reforma de los tratados.	apoyados por un "grupo preparatorio" de altos funcionarios procedentes de las quince capitales, negociarán la reforma de los tratados.	"grupo preparatorio" de altos funcionarios procedentes de las quince capitales, negociarán la reforma de los tratados.
El primer número se distribuirá gratuitamente junto con el diario "A Bola", que tiene una tirada de 150.000 ejemplares.	El primer número se distribuirá gratuitamente junto con el diario "A Bola". El diario "A Bola" tiene una tirada de 150.000 ejemplares.	El primer número se distribuirá gratuitamente junto con el diario "A Bola". "A Bola" tiene una tirada de 150.000 ejemplares.
Una consulta popular ayudaría "a recuperar la confianza perdida por los ciudadanos en la UE", declaró hoy la "número dos" del Gobierno austriaco, en una entrevista al primer canal de la televisión pública alpina ORF.	<i>Not Simplified</i>	<i>Not Simplified</i>
El jefe del Estado Mayor de las Fuerzas Armadas de Sudáfrica, el general Sywo Nyanda, afirmó hoy en su primera visita oficial a Angola que las Fuerzas Armadas de este país " consiguieron destruir buena parte de las fuerzas convencionales de UNITA".	<i>Not Simplified</i>	El jefe del Estado Mayor de las Fuerzas Armadas de Sudáfrica, el general Sywo Nyanda, afirmó hoy en su primera visita oficial a Angola: las Fuerzas Armadas de este país " consiguieron destruir buena parte de las fuerzas convencionales de UNITA".
Sin embargo y contra toda previsión de la administración, en los últimos seis años el desempleo ha crecido de manera alarmante y hoy alcanza a más del 30 por ciento de la población activa.	Sin embargo y contra toda previsión de la administración, en los últimos seis años el desempleo ha crecido de manera alarmante. Hoy alcanza a más del 30 por ciento de la población activa.	Sin embargo y contra toda previsión de la administración, en los últimos seis años el desempleo ha crecido de manera alarmante. Hoy alcanza a más del 30 por ciento de la población activa.
Su primer triunfo como profesional le llegó también sobre un empinado trazado en el Alto a Penhas de Saúde, en Portugal, lo que le dio "confianza para los próximos compromisos".	Su primer triunfo como profesional le llegó también sobre un empinado trazado en el Alto a Penhas de Saúde, en Portugal. Esto le dio "confianza para los próximos compromisos".	Su primer triunfo como profesional le llegó también sobre un empinado trazado en el Alto a Penhas de Saúde, en Portugal. Esto le dio "confianza para los próximos compromisos".
En el primer semestre de 1999, el Banco Urquijo, que cuenta con unos activos de 4.000 millones de dólares, tuvo unos resultados superiores en un 20,5 por ciento a los del mismo periodo de 1998.	El Banco Urquijo cuenta con unos activos de 4.000 millones de dólares. En el primer semestre de 1999, el Banco Urquijo tuvo unos resultados superiores en un 20,5 por ciento a los del mismo periodo de 1998.	El Banco Urquijo cuenta con unos activos de 4.000 millones de dólares. En el primer semestre de 1999, el Banco Urquijo tuvo unos resultados superiores en un 20,5 por ciento a los del mismo periodo de 1998.
El crudo Brent para agosto se cotizaba a 29,40 dólares el barril hacia las 19.10 GMT en el mercado de Londres, frente a los 28,02 dólares a que estaba en la jornada anterior.	<i>Not Simplified</i>	<i>Not Simplified</i>
La presencia de más de cien mil espectadores en la ceremonia inaugural advirtió del éxito de público que podía significar estos Juegos.	<i>Not Simplified</i>	<i>Not Simplified</i>
La candidatura de Pérez Touriño fue confirmada esta tarde por la Comisión de Garantías Electorales del PSdeG, aunque deberá ser ratificada por el Comité Nacional Gallego.	La candidatura de Pérez Touriño fue confirmada esta tarde por la Comisión de Garantías Electorales del PSdeG. Sin embargo, deberá ser ratificada por el Comité Nacional Gallego.	La candidatura de Pérez Touriño fue confirmada esta tarde por la Comisión de Garantías Electorales del PSdeG. Sin embargo, deberá ser ratificada por el Comité Nacional Gallego.
El Gobierno, que suscribió en diciembre pasado un plan de ajuste con el Fondo Monetario Internacional (FMI), se comprometió a incrementar la gestión de la	El Gobierno suscribió en diciembre pasado un plan de ajuste con el Fondo Monetario Internacional (FMI). El Gobierno se comprometió a	El Gobierno suscribió en diciembre pasado un plan de ajuste con el Fondo Monetario Internacional (FMI). El Gobierno se comprometió a

Dirección de Impuestos y Aduanas Nacionales (DIAN) para enfrentar la evasión y el contrabando.	incrementar la gestión de la Dirección de Impuestos y Aduanas Nacionales (DIAN) para enfrentar la evasión y el contrabando.	incrementar la gestión de la Dirección de Impuestos y Aduanas Nacionales (DIAN) para enfrentar la evasión y el contrabando.
Actualmente el portal tiene más de 120.000 usuarios registrados y diariamente recibe solicitudes de más de 4.000 presupuestos para viajes.	Actualmente el portal tiene más de 120.000 usuarios registrados. Diariamente recibe solicitudes de más de 4.000 presupuestos para viajes.	Actualmente el portal tiene más de 120.000 usuarios registrados. Diariamente recibe solicitudes de más de 4.000 presupuestos para viajes.