Description of the Chinese-to-Spanish Rule-Based Machine Translation System Developed using a Hybrid Combination of Human Annotation and Statistical Techniques

Marta R. Costa-jussà[1] and Jordi Centelles
Institute for Infocomm Research, 1 Fusionopolis Way, 138632 Singapore
Universitat Politècnica de Catalunya, C/ Jordi Girona 1-3, 08034 Barcelona
marta.ruiz@upc.edu, jordi.centelles.sabater@alu-etsetb.upc.edu

Two of the most popular Machine Translation (MT) paradigms are rule-based (RBMT) and corpus-based, which include the statistical systems (SMT). When scarce parallel corpus is available, RBMT becomes particularly attractive. This is the case of the Chinese-Spanish language pair.

This paper presents the first RBMT system for Chinese to Spanish. We describe a hybrid method for constructing this system taking advantage of available resources such as parallel corpora that are used to extract dictionaries and lexical and structural transfer rules.

The final system is freely available on-line and open-source. Although performance lags behind standard SMT systems for an in-domain test set, the results show that the RBMT's coverage is competitive and it outperforms the SMT system in an out-of-domain test set. This RBMT system is available to the general public, it can be further enhanced and opens up the possibility of creating future hybrid MT systems.

1. INTRODUCTION

Chinese and Spanish are two of the most widely-spoken languages in the world and they are gaining importance in today's information society. For example, in the ranking of usage of content languages for websites, Spanish is in the 4th position and Chinese ranks 5th [Q-SuccessConsulting 2013].

Although communication between the two communities is challenging on both cultural and linguistic levels, there are many shared economic interests. China's trade with Latin America grew in 2011, with the top five nations involved being Brazil, Mexico, Chile, Venezuela and Argentina [IndoAsianNews 2013]. In this sense, Machine Translation (MT) between the two languages is of clear interest for companies, tourists, students and politicians. Even though the necessity is obvious, there are not many Chinese-to-Spanish MT systems available on the Internet. Systems available in

---

[1]Currently, visiting Instituto Politècnico Nacional, Mexico

Google Translate and Bing are corpus-based and these seem to produce translations through English pivoting to compensate for the lack of Chinese-Spanish parallel corpora.

When it comes to academic research, there have not been many studies on this language pair and they are mainly reviewed in [Costa-jussà et al. 2012; Bertoldi et al. 2008] and these studies also rely on the pivoting procedure. Results of such approaches do not improve the direct system (for the same quantity of data) unless there is a combination of pivot systems. The main drawback of the combination is that the final system is not efficient. Therefore, corpus-based approaches have been applied to this language pair but results are not comparable to other language pairs (such as English-Spanish) because of the limited data available. Furthermore, the linguistic differences (mainly at the morphological level) between these two languages makes the training of data-driven systems rather difficult.

1.1. Linguistic Differences between Chinese and Spanish

In fact, Chinese and Spanish are languages with multiple linguistic differences.

At the morphological level, Chinese is an analytical (or "isolating") language, which means that it shows a low ratio of morphemes to words; in fact, the correspondence is nearly one-to-one. Spanish, on the other hand, is a fusional language, meaning that morphemes (at least one independent morpheme per word) are mixed together in words with no clear limits.

At the level of syntax, Chinese and Spanish tend to follow the Subject-Verb-Object pattern and, theoretically, there are not as many long reorderings as would be expected for language pairs such as Japanese-Spanish. As stated in Cuza et al. [Cuza et al. 2013] the Chinese-Spanish contrast is particularly evident in anaphorical elements and the nominal domain. For example, the Spanish clitic/null object alternation is regulated by features of definiteness and specificity. In contrast, Chinese is a radical pro-drop, topic-oriented language that does not productively use pronouns, does not mark for definiteness and specificity in the nominal system, and null objects are identified by linking to discourse antecedents. Therefore, moving from Chinese to Spanish may depend more on the abstract meaning of the text than on the words and structure within it. Additionally, Chinese is a language with a massive number of homonyms at the lexical and morphemic level [Zhang et al. 2006], making the lexical semantic disambiguation towards Spanish even harder.

1.2. Motivation

We have shown that there is a need for translation between Chinese and Spanish (motivated by the global market). We have also mentioned that there are no reliable MT systems that can help communication between different parties and that translating from Chinese to Spanish is challenging due to the linguistic differences. Given this situation, we decided to build a Chinese-to-Spanish Rule-Based Machine Translation (RBMT) system. This type of system provides a translation based on linguistic knowledge in contrast to the existing and popular corpus-based approaches. The translation process is divided in: analysis, transfer and generation. Analysis and generation mainly cover the morphological variation of the languages, while the transfer phase is in charge of the syntactic aspects [Hutchins and Sommers 1992]. The main advantages of RBMT systems are that errors are traceable to the corresponding linguistic rules.

Our linguistic motivations for constructing a Chinese-to-Spanish RBMT are the following ones:

(1) The proposed system can directly tackle the differences in morphology between Chinese and Spanish by following the analysis and generation steps.

(2) Reordering from Chinese and Spanish can benefit from the use of manually post-edited transfer rules.

(3) The RBMT approach is able to exploit the use of linguistic tools that are separately available for Chinese and Spanish.

The main drawbacks of a RBMT system are that it requires a lot of human dedication and years of development [Costa-Jussà et al. 2012] and that this type of system exhibits weaknesses in lexical selection, something which is quite relevant for this language pair. To overcome these drawbacks, we used the Apertium platform [Forcada et al. 2011], which makes the process of constructing an RBMT system easier. Moreover, when applying the proposed RBMT approach, we used automatic techniques at different stages of analysis, generation, and lexical and structural transfer to feed the system from parallel corpora.

1.3. Paper Contribution and Organization

This paper describes how the Chinese-to-Spanish RBMT system has been developed under the Apertium open-source platform. The integration was undertaken using the General Public License (GPU) tools available for the analysis and generation steps of the languages at hand. Monolingual and bilingual dictionaries have been created combining human effort together with statistical lexical extraction from parallel corpora. For the transfer phase, rules have been automatically generated from a parallel corpus using recently developed algorithms [Sánchez-Martínez and Forcada 2009a]. These rules have been experimentally validated. This new freely available on-line language pair is the first RBMT approach developed for Chinese to Spanish.

The rest of the paper is organized as follows. Section 2 presents a detailed description of the Chinese-to-Spanish RBMT architecture including the procedure of compiling monolingual and bilingual dictionary procedures as well as the inference of lexical and structural transfer rules. Section 3 provides a complete evaluation of the system including a comparison with a standard Statistical Machine Translation (SMT) system. Finally, Section 4 discusses the results and draws the final conclusions.

2. RULE-BASED MACHINE TRANSLATION ARCHITECTURE

The general architecture of a RBMT system has been defined previously [Hutchins and Sommers 1992; Forcada et al. 2011]. However, in most cases, RBMTs have been developed by private companies. Therefore, the inner workings, strategies and functioning of RBMTs have not been revealed because the private companies commercialize licenses to closed-source systems.

In this section, we describe in detail how we have developed our RBMT system using the Apertium platform [Forcada et al. 2011] and following similar procedures to those used in previous studies [Cortés et al. 2012]. Novel aspects of our work include: (1) a particularly challenging language pair with few bilingual speakers capable of developing the resources required to compile the target system; and (2) combining statistical techniques with human annotation.

Figure 1 shows the representative block diagram modules of the RBMT system.

Human annotation was supported by two bilingual English-Spanish annotators and one trilingual Chinese-English-Spanish annotator, who was in charge of checking each step.

2.1. System Architecture

The system is based on the Apertium platform [Forcada et al. 2011] which is a free/open-source platform for shallow transfer MT. As well as the platform, the linguistic data for the MT systems are also available under the terms of the GNU-GPL.
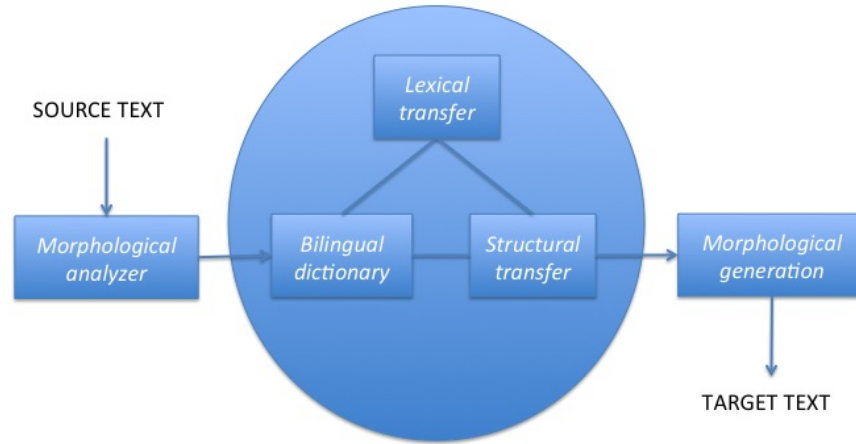
Fig. 1. Block diagram of the RBMT modules. Inside the cercle, the ones that have been built.

The platform was originally designed for the Romance languages of Spain, so the targeted language pair deviates from the original objectives. However, in practice, statistical techniques were used to build our system, and we have not changed its architecture.

Development to date has consisted of:

— feeding monolingual and bilingual dictionaries, to extend coverage, with statistical methods,
— feeding monolingual and bilingual dictionaries, to extend coverage, with human annotation,
— filtering and cleaning monolingual and bilingual dictionaries to make them consistent,
— computing lexical selection rules by automatically extracting them from parallel corpora,
— computing structural transfer rules by combining manual and automatic procedures.

Coverage (i.e. the percentage of translated words versus the total number of words used) is defined as the primary measure to evaluate the system.

This first development track took place in over the course of just six months, in contrast to the length of time required to develop classical RBMT systems. Most of the effort (around 80%) revolved around feeding the monolingual and bilingual dictionaries with human annotation. The key point here is that our system has been developed using a hybrid approach. Although the system achieves competitive results in terms of

coverage, it is still under construction. The latest version of the system is available for downloading at the Apertium development site[2].

2.2. Bilingual Dictionary

The bilingual dictionary was computed and principally constructed following two different methods.

The first involves the Yellow Bridge resource[3]. This website, as mentioned by the authors, is the premier guide to Chinese language and culture for English speakers and it provides comprehensive tools for learning the Chinese language. In particular, it gives a list of Chinese words classified following grammatical categories (i.e. adjectives, adverbs, conjunctions, interjections, measure words, nouns, numerals, onomatopoeia, particles, prefixes, prepositions, pronouns, question words, suffixes, time words and different types of verbs). For each category, the words have corresponding translations into English. Our work consisted of translating English into Spanish and this was done by an English-Spanish expert. To double-check the translations provided, each word was translated using another on-line dictionary[4] and Google Translate. After this, the bilingual Chinese-Spanish entries were added to the bilingual dictionary. This procedure allowed several hundreds of numerals, conjunctions, adverbs, pronouns, determinants, adjectives, 3,000 nouns and 2,000 verbs to be added.

The second procedure is statistical. The parallel corpus of the United Nations General Assembly Resolutions (UN) [Rafalovitch and Dale 2009] was aligned at the word level by using the standard GIZA++ [Och and Ney 2003] software. Alignment was performed from source to target and target to source. Symmetrization was done using intersection because this is expected to provide the most reliable links. Then we extracted phrases of length one meaning that we extracted translation from word-to-word. The dictionary was manually filtered to eliminate incorrect entries. This procedure allowed around 3,500 words to be added to the dictionaries. As a result of the two procedures, the bilingual dictionary now contains approximately 9,000 words.

2.3. Chinese Monolingual Dictionary

The Chinese monolingual dictionary was extracted from the source part of the bilingual dictionary. Additionally, it was filtered with regular expressions to avoid repeated entries.

From a morphological point of view, as mentioned earlier, Chinese is an isolating language. In practice, this means that words (or symbols) cannot be segmented in submorphemes. In this sense, no morphological analysis is required. However, the main challenge of Chinese at this level is that most of the time symbols appear concatenated and sentences are not segmented into words as is typical in other languages. Therefore, Chinese must be segmented. We considered using popular tools, such as the Stanford Segmenter [Chang et al. 2008], but this is written in Java, so it was difficult to integrate into Apertium. For this reason, we used the ZhSeg [Dyer 2013] programmed in C++. We evaluated the performance of this segmenter in comparison to the left-to-right longest-match (LRLM) strategy, which is the parsing strategy used by Apertium in analysis mode. This procedure reads tokens from left to right, matching the longest sequence that is in the dictionary (like "greedy" matching of regular expressions). Both ZhSeg and LRLM were compared using a manually constructed segmented test set of 456 words as a reference. The Word Error Rate (WER) measured [McCowan et al. 2004] for the ZhSeg was 16.56% and 16.89% for LRLM. Given that results were com-

---

[2]http://sourceforge.net/projects/
[3]http://www.yellowbridge.com/chinese/chinese-parts-of-speech.php
[4]http://www.chinese-tools.com/

parable, we decided to use the Apertium morphological analyzer, which applies the LRLM strategy. All finite-state tools in Apertium use the LRLM strategy.

It is mandatory that the monolingual and bilingual dictionaries are coherent, meaning that they should have the same entries. Both dictionaries were cleaned up with different regular expressions. It is therefore necessary to ensure that there are no situations where a word appears in the monolingual dictionary which is not in the bilingual dictionary, and vice versa. In order to check this out, we used a testvoc. As mentioned in the Apertium documentation[5], a testvoc is literally a test of vocabulary. At the most basic level, this just expands the monolingual dictionary, and runs each possibly analyzed lexical form through all the translation stages to see that for each possible input, a sensible translation is generated in the target language.

2.4. Spanish Generation

The generation of Spanish was taken directly from the Apertium repository. Given that this has been developed over several years, examples of earlier publications explaining Spanish generation can be found in [Armentano-Oller et al. 2006; Corbí-Bellot et al. 2005]. Basically, it consists of two Apertium modules: the morphological generator and the post-generator. The morphological generator delivers a surface (inflected) form for each transferred word and for each lemma and part-of-speech tag it is able to generate the final form. The post-generator performs orthographic operations such as contractions.

2.5. Lexical Selection

Lexical selection is very important in any type of MT system. If the lexical module is not implemented, Apertium deals with translation ambiguity by using multi-word expressions encoded in the system's dictionaries as described in previous studies [Brandt et al. 2011]. In most systems, for any given source word, the most frequent, or most general translation is presented. Obviously, this is not a desirable situation as it poses a translation problem: in most cases the correct translation depends on the context.

In this work, we used the procedure proposed in [Tyers 2013] which tries to learn lexical selection rules directly from a parallel corpus. The procedure is as follows: the source and target corpus are morphologically analyzed. Then, the parallel corpus is aligned at the word level. The source corpus is run through the lexical-transfer stage, which generates three outputs: the source and target tagged corpus and the possible translations of the source words. At the same time, sentence pairs for which at least one lexically ambiguous source word is aligned to a word in the target (also found in the bilingual dictionary) are extracted from the parallel corpus. For each of these extracted sentence pairs, n-grams of context around the ambiguous source word(s), belonging to the categories of adjective, noun and verb, are extracted. From these n-grams we extract relative frequencies by counting. If a possible given translation appears aligned to a word in a given context more frequently than other possible translations, then a rule is extracted. The rule basically selects the most frequently aligned translation in that same context over other translations. This method enabled 3,400 lexical rules to be extracted.

2.6. Transfer Rules

Structural transfer rules were extracted by combining two procedures: manual and statistical.

The manual procedure consisted of translating a source text and contrasting the output translation, the source and the reference. From this observation, manual pat-

---

[5]http://wiki.apertium.org/wiki/Testvoc

terns were extracted in order to design a rule that could cover any necessary modifications. Following this procedure, 28 intrasyntagms and 34 intersyntagms rules were extracted. The main difference is that the former rules modify within a syntagm and the latter rules modify between different syntagms. An example of a handwritten rule for the first level is as follows:

```
< rule comment = "RULE : adj nom" >
      < pattern >
          < pattern − item n = "adj" / >
          < pattern − item n = "nom" / >
      < /pattern >
      < action >
      < call − macro n = "f − concord2" >
          < with − param pos = "2" / >
          < with − param pos = "1" / >
      < /call − macro >
      < out >
          < chunk name = "j_n" case = "caseFirstWord" >
              < tags >
                  < tag >< lit − tag v = "SN" / >< /tag >
                  < tag >< clip pos = "2" side = "tl" part = "gen" / >< /tag >
                  < tag >< clip pos = "2" side = "tl" part = "nbr" / >< /tag >
                  < tag >< lit − tag v = "p3" / >< /tag >
              < /tags >
          < lu >
              < clip pos = "2" side = "tl" part = "whole" / >
          < /lu >
          < b pos = "1" / >
              < lu >
                  < clip pos = "1" side = "tl" part = "lem" / >
                  < clip pos = "1" side = "tl" part = "a_adj" / >
                  < clip pos = "1" side = "tl" part = "gen" / >
                  < clip pos = "1" side = "tl" part = "nbr" / >
              < /lu >
          < /chunk >
      < /out >
      < /action >
< /rule >
```

This rule reorders *adjective + noun* into *noun + adjective*. Moreover, it ensures that the number and gender of the noun and the adjective agree. In order to do the latter, we call the *f_concord2* macro function. This allows forced agreement in terms of gender and number. Notice that these are first level rules (t1x), meaning that they work within syntagms.

For a statistical procedure, we can use the example of [Sánchez-Martínez and Forcada 2009b]. This consists of aligning at word level, parallel corpora aligned at sentence level; extracting alignment templates and applying a set of restrictions derived from the bilingual dictionary of the RBMT system. Word alignment is done using standard statistical methods, in particular GIZA++ in the source to target and target to source direction and symmetrizing using the refined intersection method [Och and Ney 2003]. Alignment templates are extracted as detailed in [Och and Ney 2004]. Finally, restrictions are obtained from the RBMT system's bilingual dictionary where the

inferred transfer rules are integrated. These restrictions explicitly ensure that the inflection information of the translation that is changed from one language to the other is coded.

A more recent statistical procedure is the extension of the first procedure mentioned above [Sánchez-Cartagena 2014]. In this case, the study allows more general rules and exploits the information contained in the bilingual phrases to decide, in a context-dependent manner, the level of generalization, i.e. the morphological attributes to be removed from the word classes and the lexical words. This approach is the one we have used in this study and which allowed us to extract 100 rules.

Finally, manual and automatic rules are combined together by giving preference to the manual rules. In case no manual rule is applicable, then the system uses an automatic rule (when available).

3. EVALUATION FRAMEWORK

This section explains the evaluation framework for analyzing the quality of the described Chinese-to-Spanish RBMT system. We present the data used both for training statistical techniques and testing. Then, we detail how the baseline SMT system was built. Finally, we report figures for coverage and other relevant MT quality measures. Note that coverage was computed using the available Apertium script[6].

3.1. Data

As far as we know, and as discussed in [Costa-jussà et al. 2012], four parallel corpora are available for the Chinese–Spanish language pair: BTEC (Basic Travel Expression Corpus), the *Holy Bible*, the KDE (K Desktop Environment, which is available from the OPUS project[7]) and the UN. The first was used in the 2008 IWSLT and complete experiments of pivot strategies are reported in works such as [Bertoldi et al. 2008]. The *Holy Bible* was used for similar purposes in [Henríquez Q. et al. 2010]. In this study, we decided to use the UN corpus because it is freely accessible, it uses a more general vocabulary and it is the largest one.

Table I shows the main statistics for all the corpora used once divided for experimentation. Additionally, we had another development corpus called NEWS (extracted from the web[8,9]) for checking the evolution of the RBMT system. This corpus was monolingual, which is enough when checking the coverage. Finally, we utilised a small corpus developed in-house for the transportation and hospitality domains.

Table I. Corpus Statistics used for this research (all figures are given in thousands)

Dataset	Domain	Language	Sentences	Words	Vocabulary
Train	UN	Chinese	58	1,700	17
		Spanish	58	2,300	20
Dev	UN	Chinese	1	33.0	3
		Spanish	1	43.4	5
	NEWS	Chinese	-	1.6	0.7
		Spanish	-	-	-
Test	UN	Chinese	0.5	14.6	2.5
		Spanish	0.5	17.9	3.1
	IN-HOUSE	Chinese	0.5	3.2	1.9
		Spanish	0.5	3.2	1.1

---

[6]http://wiki.apertium.org/wiki/Calculating_coverage

[7]http://opus.lingfil.uu.se/

[8]http://politics.people.com.cn/n/2013/0709/c1001-22134594.html

[9]http://finance.people.com.cn/n/2013/0722/c1004-22275982.html

3.2. Statistical Machine Translation System

In order to compare our system, we built a standard phrase-based SMT system. The basic idea was to segment the given source sentences s into segments of one or more words, then translate each source segment using a bilingual phrase obtained from the training corpus and, finally, compose the target sentence from these phrase translations. Further description of the theory behind the baseline system can be found in previous works [Koehn et al. 2003].

Our system was built using revision 4075 of the MOSES system [Koehn et al. 2007]. We used the default MOSES parameters, which include the grow-diagonal-final-and word alignment symmetrization, the lexicalized reordering, relative frequencies, lexical weights and the phrase bonus for the translation model (with phrases up to length 10), a 5-gram language model using Kneser-Ney smoothing and a word penalty model. Therefore, 14 different features are combined. The language model was built using SRILM [Stolcke 2002]. Optimization was carried out using MERT [Och 2003]. For word alignment we used GIZA++, as mentioned earlier in this article.

3.3. Automatic Results

MT systems (and particularly RBMT ones) may be evaluated in terms of coverage. We used texts from different domains to perform the coverage evaluation. Table II shows the coverage results of the RBMT system for the development and test data sets described in Section 3.1. The coverage of the SMT system is also shown. The coverage of the SMT system is higher in the UN corpus because this test corpus is from the same domain (in-domain) as the training corpus. For the other corpora (NEWS and IN-HOUSE), which are out-of-domain, the RBMT system doubles the coverage of the SMT system.

Table II. Coverage results.

Dataset	Domain	Words	Translations	Coverage (%)	SMT
DEV	NEWS	1,651	1,465	**88.7**	42.1
TEST	UN	14,608	12,080	82.7	**92.1**
	IN-HOUSE	3,637	2,982	**82.0**	47.3

In addition to evaluating the coverage, we also wanted to show the quality of the translation in terms of BLEU [Papineni et al. 2002], WER and PER [McCowan et al. 2004], ROUGE [Lin and Och 2004] and METEOR [Lavie and Agarwal 2007], which are standard measures used to evaluate SMT. Evaluation was undertaken using the ASIYA toolkit [Gonzàlez et al. 2012].

Table III shows the results for the different systems. RBMT is the baseline system, it does not include lexical rules, and only uses manually written structural rules. The next system (+LexicalTransfer) adds statistically extracted lexical rules, as explained in Section 2.5. The system in the third row adds the automatic structural rules extracted to the manual structural rules, as explained in Section 2.6. Finally, the results for the SMT system are also shown.

We see that both adding lexical and structural rules extracted statistically from parallel corpora increases the performance of the RBMT system in all measures in the UN test set, which makes sense since automatic rules where extracted with the UN training corpus from Table I.

When testing the improvements in the out-of-domain test sets, we see that the automatically extracted lexical and structural rules improve the performance of the translation system in terms of PER, METEOR and ROUGE. Although the translation quality is quite low, in this case, the rule-based method is capable of outperforming the statistical system.

Table III. MT results. In bold the best results.

Dataset	System	BLEU	WER	PER	METEOR	ROUGE
UN	RBMT	0.0627	0.8427	0.7151	0.1233	0.3079
	+LexicalTransfer	0.0705	0.8401	0.6676	0.1351	0.3079
	+StructuralTransfer	**0.0796**	**0.8472**	**0.6266**	**0.1539**	**0.3359**
	SMT	**0.3521**	**0.6685**	**0.3227**	**0.3228**	**0.5466**
IN-HOUSE	RBMT	**0.0175**	**0.8322**	0.6920	0.0920	0.1863
	+LexicalTransfer	0.0164	0.8525	0.6811	0.0962	0.1996
	+StructuralTransfer	0.0166	0.8516	**0.6750**	**0.0989**	**0.2012**
	SMT	0.0149	0.9156	0.7864	0.0607	0.1103

Table IV shows some examples of the translation outputs when using statistically extracted lexical transfer rules, and without these rules.

Table IV. Translation examples with statistically extracted lexical transfer rules or without for the UN test set.

RBMT	régimen
+LexicalTransfer	examen **mecanismo**
Reference	mecanismo
RBMT	por sexo y raza
+LexicalTransfer	**género** y raza
Reference	por género y etnia
RBMT	rectora ejecutiva
+LexicalTransfer	**dirección** ejecutiva
Reference	dirección ejecutiva

Table V presents some examples of the translation outputs when using statistically extracted structural transfer rules and without them.

Table V. Translation examples with o rwithout statistically extracted structural transfer rules for the UN test set.

RBMT	convención de viabilidad
+StructuralTransfer	**viabilidad de los convenios**
Reference	viabilidad de los convenios
RBMT	régimen de expertos
+StructuralTransfer	los **expertos del régimen**
Reference	los expertos del régimen
RBMT	fruta persona diálogo
+StructuralTransfer	**diálogo de la fruta**
Reference	diálogo sobre la fruta

3.4. Human Linguistic Analysis

Finally, a linguistic annotator evaluated 100 random sentences from the in-house test set using a human linguistic analysis that classified all the translation errors into one of the five following linguistic levels: orthographic, morphological, lexical, semantic, and syntactic. Linguistic guidelines for the target language are those presented in [Farrús et al. 2012] and for each linguistic level involved in the classification, a list of error subtypes was provided. For the RBMT system we used the best automatic output from Table III.

Table VI. Linguistic evaluation results.

System	ort.	mor.	lex.	sem.	syn.
RBMT	**0**	**43**	**80**	**52**	32
SMT	1	45	181	60	**25**

Table VI shows the results of the linguistic analysis and the RBMT system outperforms the SMT system at all linguistic levels with the exception of syntax. However, the differences are not very large apart from at the lexical level. The coverage of both systems is very different in the out-of-domain test set. Table VII shows some examples of the outputs.

Table VII. Translation examples comparing the RBMT and SMT outputs of the IN-HOUSE test set.

RBMT	yo sólo cincuenta dólar
SMT	cincuenta dólares de los estados unidos . sólo
Reference	tengo sólo cincuenta dólares encima .
RBMT	tiene qué cerveza ?
SMT	la que pueden
Reference	qué cerveza toma usted ?
RBMT	yo fundamentalmente hace no trabajo .
SMT	que no fundamental en toda labor .
Reference	no estoy haciendo ningún trabajo .

4. CONCLUSIONS AND FURTHER WORK

The novel aspects covered in this article include: (1) a detailed description of the first Chinese-to-Spanish open-source RBMT system; and (2) a RBMT system built using hybrid techniques, combining human knowledge and statistical techniques.

In particular, human expertise has been used to create exhaustive monolingual and bilingual dictionaries as well as define structural transfer rules. Additional statistical knowledge complements all the steps mentioned. Moreover, statistical information has been the only source for the lexical transfer rules. The improvement in the RBMT system provided by the use of statistical knowledge in the RBMT system has been evaluated and it has been shown to enhance translation output. In this sense, we present effective techniques for constructing a RBMT system using hybrid techniques. Moreover, the RBMT system outperforms the statistical system in the out-of-domain test.

The new RBMT system and the methods for constructing it have been evaluated both automatically and through a linguistic human analysis. Moreover, the output of the lastest version of the RBMT system has been contrasted with a statistical state-of-the-art system for the out-of-domain test set. The RBMT system outperforms the SMT system at all linguistic levels except at the syntax level. Clear outperformance is shown in terms of lexical coverage.

Future work includes improving the RBMT system with new dictionary entries and more complex transfer rules. Both enhancements can be made combining human and statistical knowledge. Also, the fact of having two systems based on different principles in such a challenging language pair may be useful for research in the active field of hybrid MT. Finally, another RBMT toolkit, such as Matxin [Mayor et al. 2011], could be applied to solve reordering problems.

REFERENCES

C. Armentano-Oller, R. C. Carrasco, A. M. Corb-Bellot, M. L. Forcada, M. Ginestí-Rosell, S. Ortiz-Rojas, J. A. Pérez-Ortiz, G. Ramírez-Sánchez, F. Sánchez-Martínez, and M. A. Scalco. 2006. Open-source Portuguese-Spanish machine translation. In *Computational Processing of the Portuguese Language, Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese, PROPOR 2006*, R. Vieira, P. Quaresma, M.d.G.V. Nunes, N.J. Mamede, C. Oliveira, and M.C. Dias (Eds.). Lecture Notes in Computer Science, Vol. 3960. Springer-Verlag, 50–59.

N. Bertoldi, R. Cattoni, M. Federico, and M. Barbaiani. 2008. FBK @ IWSLT-2008. In *Proc. of the International Workshop on Spoken Language Translation*. Hawaii, USA, 34–38.

M. D Brandt, H. Loftsson, H. Sigurrsson, and F. M. Tyers. 2011. Apertium-IceNLP: A rule-based Icelandic to English machine translation system . In *Proceedings of the 16th Annual Conference of the European Association of Machine Translation*.

P.C. Chang, M. Galley, and C. D. Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the Third Workshop on Statistical Machine Translation (StatMT '08)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 224–232. http://dl.acm.org/citation.cfm?id=1626394.1626430

A. M. Corbí-Bellot, M. L. Forcada, S. Ortiz-Rojas, J. A. Pérez-Ortiz, G. Ramírez-Sánchez, F. Sánchez-Martínez, I. Alegria, A. Mayor, and K. Sarasola. 2005. An open-source shallow-transfer machine translation engine for the romance languages of Spain. In *Proceedings of the Tenth Conference of the European Association for Machine Translation*. 79–86.

J. P. Martínez Cortés, J. O'Regan, and F. M. Tyers. 2012. Free/Open Source Shallow-Transfer Based Machine Translation for Spanish and Aragonese. In *LREC*. 2153–2157.

M. R. Costa-Jussà, M. Farrús, J. B. Mariño, and J. A. R. Fonollosa. 2012. Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems. *Computing and Informatics* 31, 2 (2012), 245–270.

M. R. Costa-jussà, C. A. Henríquez Q, and R. E. Banchs. 2012. Evaluating indirect strategies for Chinese-Spanish statistical machine translation. *Journal of Artificial Intelligence Research* 45, 1 (Sept. 2012), 761–780. http://dl.acm.org/citation.cfm?id=2444851.2444870

A. Cuza, M. Perez-Leroux, and L. Sánchez. 2013. The role of semantic transfer in clitic-drop among Chinese L1-Spanish L2 bilinguals. *Studies in Second Language Acquisition* 35, 1 (2013), 93–125.

C. Dyer. 2013. http://code.google.com/p/zhseg/. (2013).

M. Farrús, M. R. Costa-jussà, and M. Popović. 2012. Study and correlation analysis of linguistic, perceptual, and automatic machine translation evaluations. *Journal of the Association for Information Sciences and Technolgy (JASIST)* 63, 1 (Jan. 2012), 174–184. DOI:http://dx.doi.org/10.1002/asi.21674

M. L. Forcada, M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez, and F. M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation* 25, 2 (2011), 127–144.

M. Gonzàlez, J. Giménez, and Ll. Màrquez. 2012. A Graphical Interface for MT Evaluation and Error Analysis. In *The 50th Annual Meeting of the Association for Computational Linguistics*.

C. A. Henríquez Q., R. E. Banchs, and J. B. Mariño. 2010. Learning Reordering Models for Statistical Machine Translation with a Pivot Language. (2010). Internal Report TALP-UPC.

W. J. Hutchins and L. Sommers. 1992. An introduction to machine translation. *Academic Press* 362 (1992).

IndoAsianNews. 2013. China's trade with Latin America grew in 2011. (Accessed on-line September 2013). http://in.news.yahoo.com/chinas-trade-latin-america-grew-2011-050334275.html

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*. Prague, Czech Republic, 177–180.

P. Koehn, F.J. Och, and D. Marcu. 2003. Statistical Phrase-Based Translation. In *Proc. of the 41th Annual Meeting of the Association for Computational Linguistics*.

A. Lavie and A. Agarwal. 2007. Meteor: an automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation (StatMT '07)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 228–231. http://dl.acm.org/citation.cfm?id=1626355.1626389

C.-Y. Lin and F. J. Och. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04)*. Association for Computational Linguistics, Stroudsburg, PA, USA, Article 605. DOI:http://dx.doi.org/10.3115/1218955.1219032

A. Mayor, I. Alegria, A. Díaz de Ilarrraza, G. Labaka, M. Lersundi, and K. Sarasola. 2011. Matxin, an open-source rule-based machine translation system for Basque. *Machine Translation* 25 (2011), 53–82.

I. A. McCowan, D. Moore, J. Dines, D. Gatica-Perez, M. Flynn, P. Wellner, and H. Bourlard. 2004. *On the Use of Information Retrieval Measures for Speech Recognition Evaluation*. Idiap-RR Idiap-RR-73-2004. IDIAP, Martigny, Switzerland.

F.J. Och. 2003. Minimum Error Rate Training In Statistical Machine Translation. In *Proc. of the 41th Annual Meeting of the Association for Computational Linguistics*. 160–167.

F.J. Och and H. Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29, 1 (March 2003), 19–51.

F.J. Och and H. Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics* 30, 4 (December 2004), 417–449.

K Papineni, S. Roukos, T. Ward, and W.-J. Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 311–318. DOI:http://dx.doi.org/10.3115/1073083.1073135

Q-SuccessConsulting. 2013. Web Technology Surveys. (Accessed on-line September 2013). http://w3techs.com/technologies/overview/content_language/all

A. Rafalovitch and R. Dale. 2009. United Nations General Assembly Resolutions: A Six-Language Parallel Corpus. In *Proc. of the MT Summit XII*. Ottawa, 292–299. http://www.uncorpora.org/

V. Sánchez-Cartagena. 2014. *To be published*. Ph.D. Dissertation. PhD thesis 2014. Departament de Llenguatges i Sistemes Infomtics, Universitat d'Alacant, Spain.

F. Sánchez-Martínez and M. L. Forcada. 2009a. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research* 34 (2009), 605–635.

F. Sánchez-Martínez and M. L. Forcada. 2009b. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research* 34 (2009), 605–635.

A. Stolcke. 2002. SRILM: an extensible language modeling toolkit.. In *Proc. of the Int. Conf. on Spoken Language Processing*. Denver, CO, 901–904.

F. M. Tyers. 2013. *Feasible lexical selection for rule-based machine translation*. Ph.D. Dissertation. PhD thesis 2013. Departament de Llenguatges i Sistemes Infomtics, Universitat d'Alacant, Spain.

Y. Zhang, N. Wu, and M. Yip. 2006. Lexical ambiguity resolution in Chinese sentence processing. *Handbook of East Asian Psycholinguistics* 1 (2006), 268–278.