A Feasibility Study for Chinese-Spanish Statistical Machine Translation

Rafael E. Banchs, Josep M. Crego, Patrik Lambert, José B. Mariño

Universitat Politècnica de Catalunya
Department of Signal Theory and Communications
08034 Barcelona, Spain

Abstract. This article presents and describes an experimental prototype system for performing Chinese-to-Spanish and Spanish-to-Chinese machine translation. The system is based on the statistical machine translation (SMT) framework and, more specifically, it implements the bilingual n-gram SMT approach. Since, as far as we know, no large Chinese-Spanish parallel corpus is currently available for training purposes, an alternative experimental method for building a training corpus was used. This method is compared, in terms of translation quality, to the simpler approach of using English as a bridge language for performing Chinese-to-Spanish and Spanish-to-Chinese translations.

1 Introduction

The origins of statistical machine translation (SMT) can be associated to the appearance of digital computers, immediately after World War II, and closely related to the ideas from which information theory arose. According to this view, MT was conceived as the problem of finding a sentence by decoding a given "encrypted" version of it [17], and although the idea seemed very feasible, enthusiasm faded out shortly afterwards because of the computational limitations of the time. It was finally during the decade of the nineties, when SMT became an actual and practical issue. Two factors made it possible: the significant increase in computational and storage capacity, and the availability of large volumes of bilingual data.

As an extension of the MT problem, additional advances in the fields of automatic speech recognition (ASR) and text to speech synthesis (TTS) made it possible to envision the new challenge of spoken language translation (SLT) [8]. According to this, SMT has also been approached from a finite-state point of view as the most natural way of integrating ASR and SMT [11, 16, 9, 1]. This kind of models, different from the phrase-based translation models, rely on probabilities among sequences of bilingual units which are defined by the transitions of a transducer. Following this idea, the SMT system used in this work implements a bilingual n-gram translation model. This system is described in detail by Mariño *et al* in [10].

Nevertheless the large efforts invested and the recent advances in MT technologies, the field continues to be one of the most important challenges of artificial intelligence (AI) and natural language processing (NLP); and, accordingly, research and development activities related to MT have significantly increased all over the globe during the last decade. As a simple example of this some ongoing projects can be mentioned, such as, *MANOS*[1] in the Popular Republic of China, *TC-STAR*[2] in the European Union, and *GALE*[3] in the United States of America.

In spite of all this global effort, it is interesting to notice that most of it is currently concentrated in some specific translation pairs such as French-English, Spanish-English, Chinese-English, Arabic-English, Japanese-English and Chinese-French among others; while some other language pairs such as Chinese-Spanish are, as far as we know, somehow unattended from both research and commercial perspectives. A very simple exercise of searching through the web for information dealing with machine translation reveals that outcomes related to Chinese-English and Spanish-English are about ten and twenty times larger, respectively, than outcomes related to Chinese-Spanish. In addition to this, it can be verified that most of the Chinese-Spanish translation systems that are currently available do not actually perform a direct translation between both languages, but use a third language (generally English) as a bridge.

In response to what has been exposed, the main objective of this work is to attempt building and evaluating a direct Chinese-Spanish SMT system. Currently, the main drawback for building such a SMT system is the inexistence, at least as a publicly available resource, of a bilingual Chinese-Spanish parallel corpus large enough to perform an appropriate training of the bilingual translation model. In this way, an alternative experimental method for building the required training corpus, starting from two independent bilingual parallel corpora: Chinese-English and Spanish-English, is proposed and evaluated. This method is further compared, in terms of translation quality, to the simpler approach of using English as a bridge language for performing Chinese-to-Spanish and Spanish-to-Chinese translations.

This document is structured as follows. The following section presents a brief overview of the SMT system used for the experimental procedures within this work. Next, section 3, describes in detail the alternative experimental method proposed for developing the Chinese-Spanish parallel training corpus. Then, section 4, describes and discusses the experiments conducted for comparing translation performances between the Chinese-Spanish SMT system developed and a system that uses English as a bridge language. Finally, section 5, presents our conclusions, as well as the further work to be performed in the near future in order to promote and improve the development of Chinese-to-Spanish and Spanish-to-Chinese SMT systems.

---

[1] Multilingual Application Network for Olympic Services. `http://nlpr-web.ia.ac.cn/english/cip/english/project.htm`

[2] Technology and Corpora for Speech-to-Speech Translation. `http://www.tc-star.org/`

[3] Global Autonomous Language Environments. `http://ciir.cs.umass.edu/research/nightingale.html`

2 The Bilingual N-gram SMT Approach

This section presents a brief description of the SMT system used in this work. For a more detailed description see [10]. The system implements a translation model that has been derived from the finite-state perspective; more specifically, from the work of Casacuberta presented in [3] and [4]. However, different from it, where the translation model is implemented by using a finite-state transducer, the SMT approach used here considers a translation model which is based on 3-grams. In this way, the bilingual n-gram translation model actually constitutes a language model of a sort of "bi-language" composed of bilingual units which are referred to as tuples [7]. This model is described by the following equation:

$$p(T,S) \approx \prod_{k=1}^{K} p((t,s)_k|(t,s)_{k-1},(t,s)_{k-2}) \quad (1)$$

where t refers to target, s to source and $(t,s)_k$ to the k^{th} tuple of a given bilingual sentence pair.

Tuples are extracted from Viterbi alignments; more specifically, from the union set of source-to-target and target-to-source alignments, which are automatically computed by using GIZA++ [12]. Tuples are extracted from alignments according to the following two constraints [5]: first, tuple extraction should produce a monotonic segmentation of bilingual sentence pairs, and second, no smaller tuples can be extracted without violating the previous constraint. Figure 1 illustrates the resulting segmentation for a given aligned pair of sentences.

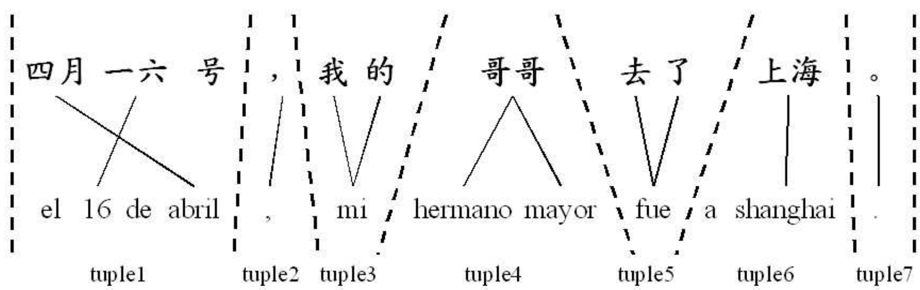

Fig. 1. Example of tuple extraction.

In addition to the tuple 3-gram translation model, the considered SMT system implements four additional feature functions which are log-linearly combined with the translation model for decoding purposes. These feature functions are the following:

– *Target language model.* This feature provides information about the target language structure and fluency. It favors those partial-translation hypotheses which are more likely to constitute correctly structured target sentences over those which are not. The model is implemented by using a word 4-gram model of the target language, which is computed according to the following expression:

$$h_{TL}(T) = log \prod_{k=1}^{K} p(w_k | w_{k-1}, w_{k-2}, w_{k-3}) \tag{2}$$

where w_k refers to k^{th} word in the considered partial-translation hypothesis.

– *Word bonus model.* This feature introduces a bonus which depends on the partial-translation hypothesis length. This is done in order to compensate the system preference for short translations over large ones. The model is implemented through a bonus factor which directly depends on the total number of words contained in the partial-translation hypothesis, and it is computed as follows:

$$h_{WP}(T) = M \tag{3}$$

where M is the number of words contained in the partial-translation hypothesis.

– *Source-to-target lexicon model* This feature actually constitutes a complementary translation model. This model provides, for a given tuple, a translation probability estimate between the source and target sides of it. This feature is implemented by using the IBM-1 lexical parameters [2, 13]. According to this, the source-to-target lexicon probability is computed for each tuple according to the following equation:

$$h_{LF}(T, S) = log \frac{1}{(J+1)^I} \prod_{i=1}^{I} \sum_{j=0}^{J} q(t_i^n | s_j^n) \tag{4}$$

where s_j^n and t_i^n are the j^{th} and i^{th} words in the source and target sides of tuple $(t, s)_n$, being J and I the corresponding total number of words in each side of it. In the equation, $q(.)$ refers to IBM-1 lexical parameters which are estimated from alignments computed in the source-to-target direction.

– *Target-to-source lexicon model.* Similar to the previous feature, this feature function constitutes a complementary translation model too. It is computed exactly in the same way the previous model is, with the only difference that IBM-1 lexical parameters are estimated from alignments computed in the target-to-source direction instead.

Finally, a search engine, which implements a beam-search strategy based on dynamic programming and allows for threshold pruning and hypothesis recombination, was specifically developed for this SMT approach. It was developed by Crego *et al.*, and a detailed description can be found in [6]. Although this search engine allows for non-monotonic search, all experiments performed in this work are performed by using monotonic search[4]. Additionally, an optimization tool, based on a downhill simplex method [15], was also developed. This algorithm adjusts the log-linear weights for each feature function so that the translation *BLEU* [14] is maximized over a development data set for each translation direction under consideration.

3 Chinese-Spanish Parallel Corpus Development

As already mentioned, the main drawback for building a Chinese-Spanish SMT system is the inexistence, at least as a publicly available resource, of a bilingual parallel corpus for performing the translation model training. Although much information in Chinese and Spanish is available through the LDC consortium[5], the intersection of databases containing either Spanish or Chinese is null, so the extraction of a Chinese-Spanish parallel corpus from the existing databases is not possible at all. For this reason, an alternative method for building the required training corpus is proposed and evaluated. The method is depicted in detail within this section.

The proposed method relies on using English-to-Chinese and English-to-Spanish SMT systems for constructing a Chinese-Spanish parallel corpus starting from two independent bilingual parallel corpora. In this way, either the English side of a Chinese-English parallel corpus is to be translated into Spanish or, similarly, the English side of a Spanish-English parallel corpus is to be translated into Chinese. However, as will be seen later, a better translation quality is currently achieved by our SMT system for English-to-Spanish than for English-to-Chinese. According to this, in this work, we restricted ourselves to constructing a Chinese-Spanish parallel corpus by translating into Spanish the English side of a Chinese-English parallel corpus.

The Chinese-English parallel corpus used in this work corresponds to official transcriptions of speeches at the United Nations, for which a large database is available through the LDC consortium. More specifically, a subset of the *UN Chinese English Parallel Text* (LDC2004E12) was used as Chinese-English data. On the other hand, in the case of the English-Spanish parallel corpus, the EPPS (European Parliament Plenary Sessions) data available through the TC-STAR consortium[6] was considered. More specifically, the training data made publicly available by the consortium during its second evaluation campaign was used. Table 1 presents basic statistics for the Chinese-English and Spanish-English parallel data sets.

---

[4] We are conscious that word reordering plays a very important role in the translation tasks under consideration, however as a first approximation to the Chinese-Spanish translation problem, and in order to maintain computational time manageable, we have opted for monotonic decoding.

[5] Linguistic Data Consortium. `http://www.ldc.upenn.edu/`

[6] Technology and Corpora for Speech-to-Speech Translation. `http://www.tc-star.org/`

Table 1. Number of sentences, running words, vocabulary size, and minimum, maximum and average sentence length for the original Chinese-English and Spanish-English corpora. (*K* and *M* stand for thousands and millions, respectively.)

Corpus	Language	Sentences	Words	Vocab.	Min.SL	Max.SL	Aver.SL
ZH−EN	Chinese	1.6 M	48.7 M	189.2 K	1	100	29.2
	English	1.6 M	52.9 M	231.4 K	2	100	31.7
ES−EN	Spanish	1.2 M	36.5 M	152.1 K	1	102	28.5
	English	1.2 M	34.9 M	106.4 K	1	100	27.2

From each parallel data set presented in table 1 a 100K-sentence subset was extracted. These subsets constitute the actual training data sets to be used for the experiments within this work. Such a selection was done with a twofold objective in mind: first, to guarantee a common English vocabulary for both corpora, and second, to maintain computational time manageable for experimentation and optimization purposes. Statistics for these training data sets are presented in table 3.

3.1 Generation of the Parallel Corpus

Once the Chinese-English and Spanish-English training data sets were defined, the generation of a Chinese-Spanish training data set was attempted by translating into Spanish the English side of the Chinese-English training set. In order to do this, two translation systems (English-to-Chinese and an English-to-Spanish) were to be trained and optimized by using the bilingual n-gram approach described in section 2.

For optimization purposes, a development data set had to be defined for each translation language pair. In order to achieve this, a trilingual parallel corpus was created by manually translating into Spanish the English side of a 330-sentence development set extracted from the complete Chinese-English parallel corpus. This development set was selected such that no overlap occurred with the 100K-sentence training set. Such a trilingual parallel corpus allows for the availability of a common development corpus for each of the three language pairs to be considered in the experiments: Chinese-English, English-Spanish and Chinese-Spanish. Table 2 presents the basic statistics for the constructed development data set.

In this way, the English-to-Chinese and English-to-Spanish translation systems were trained by using the corresponding 100K-sentence training sets presented in table 3 and optimized by maximizing the translation *BLEU* over the corresponding development sets presented in table 2. From these two optimization procedures, maximum *BLEU* scores of 0.0525 and 0.4067 were reached for the English-to-Chinese and English-to-Spanish translation systems, respectively. As already mentioned, and according to this huge difference in translation accuracy, we restricted ourselves to constructing a Chinese-Spanish parallel corpus

Table 2. Number of sentences, running words, vocabulary size, and minimum, maximum and average sentence length for the trilingual development data set. (K stands for thousands.)

Corpus	Language	Sentences	Words	Vocab.	Min.SL	Max.SL	Aver.SL
DEV	Chinese	330	6.0 K	1.6 K	10	30	18.6
	English	330	6.7 K	1.9 K	10	30	20.5
	Spanish	330	6.8 K	2.0 K	9	36	20.7

by translating into Spanish the English side of a Chinese-English parallel corpus. The basic statistics for the resulting Chinese-Spanish corpus, along with the original Chinese-English and Spanish-English corpora, are depicted in table 3.

Table 3. Number of sentences, running words, vocabulary size, and minimum, maximum and average sentence length for the constructed Chinese-Spanish training corpus, and the original Chinese-English and Spanish-English training sets. (K and M stand for thousands and millions, respectively.)

Corpus	Language	Sentences	Words	Vocab.	Min.SL	Max.SL	Aver.SL
ZH−EN	Chinese	105 K	1.9 M	29.5 K	4	86	17.9
	English	105 K	2.1 M	34.8 K	10	30	20.5
ES−EN	Spanish	105 K	2.0 M	40.0 K	5	53	19.7
	English	105 K	2.0 M	27.0 K	10	30	19.5
ZH−ES	Chinese	105 K	1.9 M	29.5 K	4	86	17.9
	Spanish	105 K	2.0 M	34.8 K	3	43	19.6

3.2 Chinese-Spanish Corpus Filtering

As will be demonstrated in section 4, the procedure of artificially constructing a Chinese-Spanish bilingual corpus by the automatic translation of the English side of a Chinese-English parallel corpus does not conduce to any improvement in translation accuracy by itself. According to this result, it seems that the only way to actually exploit a procedure of this kind is by being able to retain the most useful sentence pairs of the artificially generated corpus. In this way, the proposed methodology is complemented with a corpus preprocessing stage in which the constructed Chinese-Spanish bilingual corpus is filtered in order to eliminate possible noisy data resulting from the translation errors implied in its generation.

To this end, a language-model-based filtering strategy is proposed and evaluated. This filtering strategy consists on using a Spanish language model for selecting those best Spanish sentences in the Chinese-Spanish parallel corpus. Notice that this filtering is conducted only in the Spanish side of the corpus because it corresponds to the one which was artificially generated by translating the English side of the original Chinese-English parallel corpus. So the noise expected to occur in the Chinese-Spanish corpus should be related to the translation errors produced by the English-to-Spanish translation system. A 3-gram language model, trained from the Spanish side of the 100K-sentence English-Spanish training corpus presented in table 3, was used. Since language model probabilities are affected by sentence length, this filtering is performed independently for each subset of Spanish sentences of equal length.

Translation accuracy is automatically evaluated in terms of translation *BLEU* and other evaluation metrics for the proposed filtering strategy. Notice that although a training data size reduction has a negative impact on translation accuracy, it is expected that the noisy data reduction provided by the filtering process prevails over the data reduction effect so the overall system performance is incremented. These experimental results are presented in the second part of next section.

4 Experimental Procedure and Results

This section presents and discusses the experimental procedures considered in this work and their corresponding results. Experiments has been divided in two groups. The first set of experiments attempts to compare direct Chinese-Spanish translation with indirect Chinese-Spanish translation by using English as a bridge. These first set of experiments is presented in subsection 4.1. A second set of experiments was designed to evaluate the possibility of improving translation accuracy for the direct Chinese-Spanish translation system by using the filtering technique described in the previous section. These experiments are described and presented in subsection 4.2.

4.1 Direct vs. Indirect Chinese-Spanish Translations

The comparison between the direct and indirect Chinese-Spanish translation strategies under consideration was performed in both directions Chinese-to-Spanish and Spanish-to-Chinese. According to this, a total of six different translation systems had to be independently trained and optimized. More specifically, these systems were: Chinese-to-English, English-to-Spanish, Chinese-to-Spanish, and their corresponding opposite direction systems. After all system optimizations were carried out, translation accuracy, in terms of *BLEU* and *NIST*, and error rates, in terms of *PER* and *WER*, were computed over the development set. Results corresponding to all considered systems in both translation directions are summarized in table 4.

Table 4. Translation accuracy, in terms of *BLEU* and *NIST*, and error rates, in terms of *PER* and *WER*, over the development data set for the direct and the indirect translation strategies.

Strategy	Direction	BLEU	NIST	WER	PER
Direct	ZH→ES	0.1087	4.157	83.81	62.14
Indirect	ZH→EN	0.1666	5.218	75.97	53.10
	EN→ES	0.1145	4.413	78.04	58.21
Direct	ES→ZH	0.0391	3.946	76.16	58.54
Indirect	ES→EN	0.3597	7.544	43.04	33.28
	EN→ZH	0.0397	3.378	75.62	59.22

As can be seen from table 4, performing a direct translation does not help to increase translation quality at all, in any of both translation directions. Indeed, although in the case of Spanish-to-Chinese all evaluation metrics are basically the same for both direct and indirect translation strategies, in the case of Chinese-to-Spanish the indirect translation approach seems to perform slightly better than the direct one. So, contrarily to what we were expecting, these results suggest that translation errors occurring during the generation of the direct system's training data are not resolved by the alignment and training stages of the direct translation system, but are increased and reinforced instead.

Notice also from the table, that translation *BLEU* and *NIST* obtained for the Spanish-to-Chinese direction is much more smaller than the one obtained for the Chinese-to-Spanish direction. This might be indicating either a possible problem in our Chinese language model implementation, or that translating-into-Chinese actually constitutes a more challenging task than translating-from-Chinese. However, it is interesting to notice that in terms of the error rates both translation directions seem to be performing fairly similar. Understanding this strange inconsistency between accuracy measures and the error rates requires further study and analysis.

4.2 Direct Chinese-Spanish Translation by using a Filtered Corpus

The experiments within this section attempts to evaluate the possibility of improving translation accuracy for the direct Chinese-Spanish translation system by using the filtering technique previously proposed. As already mentioned, notice that such a filtering implies a reduction in the available amount of training data, which it is well known to have a negative impact on translation accuracy, and error rates. However, it is also expected that the elimination of noisy data provided by the filtering process prevails over the data reduction effect so that, at the end, the overall system performance is incremented.

In addition to evaluate the effects of filtering in translation quality the experiments presented in this section also look for determining which would be the best trade off between the amount of filtered data and translation quality. In this way, the artificially constructed Chinese-Spanish training set was filtered by using the language model criterion described in section 3.2 by considering different threshold values in order to generate training subsets of some predefined different sizes. More specifically, training subsets of 90K-, 70K- and 50K-sentences were generated by filtering the original 100K-sentence training set.

In order to evaluate the effect of filtering independently from the effect of reducing the size of the training data set, an additional control experiment was performed for each of the three filtering experiments under consideration. Such control experiments consisted in training and optimizing a translation system by using a randomly generated equal-size training subset. In this way each filtering experiment can be compared with a control translation system that was trained with the same amount of data; but different from the filtered one, the control training data was selected at random.

In summary, a total amount of six training subsets were generated. Three of them by using the proposed filtering strategy, and the other three at random. Basic statistics for these training subsets are presented in table 5. Accordingly, six Chinese-to-Spanish translation systems were independently trained and optimized. The corresponding translation results are presented in table 6.

Table 5. Number of sentences, running words, vocabulary size, and minimum, maximum and average sentence length for the filtered and control training subsets. (*K* and *M* stand for thousands and millions, respectively.)

Size	Corpus	Language	Words	Vocab.	Min.SL	Max.SL	Aver.SL
90K	Filtered	Chinese	1.7 M	26.5 K	4	86	17.95
		Spanish	1.9 M	32.9 K	3	43	19.89
	Control	Chinese	1.7 M	28.3 K	4	86	17.89
		Spanish	1.9 M	33.4 K	3	43	19.70
70K	Filtered	Chinese	1.3 M	21.8 K	4	82	17.57
		Spanish	1.4 M	27.9 K	3	43	19.85
	Control	Chinese	1.3 M	25.2 K	4	86	17.88
		Spanish	1.5 M	30.1 K	3	43	19.69
50K	Filtered	Chinese	0.9 M	17.1 K	4	78	17.05
		Spanish	1.0 M	22.4 K	4	43	19.66
	Control	Chinese	0.9 M	21.6 K	4	86	17.89
		Spanish	1.0 M	26.1 K	4	43	19.69

As seen from table 6, it is evident that, in the cases of 90K- and 70K-sentence corpora, filtering actually helps improving translation accuracy and error rates with respect to a system which has been trained with the same amount of non-filtered data. In the case of the 50K-sentence corpus and smaller-sized corpora

Table 6. Translation accuracy, in terms of *BLEU* and *NIST*, and error rates, in terms of *PER* and *WER*, over the development data set, for all six Chinese-to-Spanish translation systems corresponding to training data presented in table 5.

Size	Corpus	BLEU	NIST	WER	PER
100K	Original	0.1087	4.157	83.81	62.14
90K	Filtered	0.1097	4.141	83.88	62.84
	Control	0.1032	4.107	85.04	63.18
70K	Filtered	0.1065	4.081	84.28	63.68
	Control	0.0992	4.008	85.21	63.87
50K	Filtered	0.1024	3.933	86.34	65.52
	Control	0.0963	3.933	85.67	65.12

(for which results are not presented) this improvement tends to fade out progressively.

Nevertheless, notice from the table that no actual translation quality improvement has been achieved by any of the filtered systems with respect to the original 100K- system. This is clearly suggesting that the negative effect resulting from data training reduction is more relevant for the overall system performance than the positive effect resulting from filtering. According to these results different filtering strategies should be designed and studied.

5 Conclusions and Further Work

This work presented a preliminary feasibility study for performing Chinese-Spanish Statistical Machine Translation. As already mentioned, the main drawback for building a Chinese-Spanish SMT system is the inexistence, at least as a publicly available resource, of a bilingual parallel corpus for performing the translation model training. According to this, a method for artificially constructing and filtering a Spanish-Chinese parallel corpus by automatically translating into Spanish the English side of an English-Spanish parallel corpus was proposed and evaluated. Experimental results have shown that although filtering the artificially constructed training corpus does actually improve translation quality, the negative effect resulting from data reduction is more relevant for the overall system performance than the positive effect resulting from filtering. As a consequence, the proposed technique does not provide better translations than the simpler approach of performing indirect Chinese-Spanish translations by using English as a bridge language.

For further research we are planning to work in two main directions. First, we will attempt improving the Chinese-Spanish parallel corpus construction technique. In this sense, different alternatives for filtering the artificially constructed data set should be designed and evaluated, such as the use of dictionaries and morpho-syntactic information. The second main direction of work should be related to improvements in the translation system, by including additional features and allowing for non-monotonic search in the translation tasks under consideration.

References

1. Bangalore, S., Riccardi, G.: Stochastic finite-state models for spoken language machine translation. *Proc. Workshop on Embedded Machine Translation Systems* (2000) 52–59
2. Brown, P., Della Pietra, S., Della Pietra, V., Mercer, R.: The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics* **19(2)** (1993) 263–311
3. Casacuberta, F.: Finite-state transducers for speech input translation. *Proc. IEEE ASRU* (2001) Madonna di Campiglio, Italy.
4. Casacuberta, F., Vidal, E.: Machine translation with inferred stochastic finite-state transducers. *Computational Linguistics* **30(2)** (2004) 205–225
5. Crego, J.M., Mariño, J.B., de Gispert, A.: Finite-state-based and phrase-based statistical machine translation. *Proc. of the 8th Int. Conf. on Spoken Language Processing* (2004) 37–40 October.
6. Crego, J.M., Mariño, J.B., de Gispert, A.: A Ngram-based statistical machine translation decoder. *INTERSPEECH 2005* (2005) Lisbon, September.
7. de Gispert, A., Mariño, J.B.: Using X-grams for speech-to-speech translation. *Proc. of the 7th Int. Conf. on Spoken Language Processing* (2002)
8. Kay, M., Gawron, J., Norvig, P.: *Verbmobil: a translation system for face-to-face dialog.* (1992) CSLI.
9. Knight, K., Al-Onaizan, Y.: Translation with finite-state devices. *AI Lecture Notes in Artificial Intelligence* **1529** 421–437 Springer-Verlag.
10. Mariño, J.B., Banchs, R., Crego, J.M., de Gispert, A., Lambert, P., Fonollosa, J.A.R., Ruiz, M.: Bilingual N-gram Statistical Machine Translation. *Proc. of the tenth Machine Translation Summit* (2005) 275–282 Phuket, Thailand.
11. Riccardi, G., Pieraccini, R., Bocchieri, E.: Stochastic automata for language modeling. *Computer Speech and Language* **10(4)** (1996) 265–293
12. Och, F.J.: Internal report, RWTH Aachen University. Giza++ software (2003) `http://www.i6.informatik. rwth-aachen.de/`.
13. Och, F.J., Gildea, D., Khudanpur, S., Sarkar, A., Yamada, K., Fraser, A., Kumar, S., Shen, L., Smith, D., Eng, K., Jain, V., Jin, Z., Radev., D.: A smorgasbord of features for statistical machine translation. *Proc. of the Human Language Technology Conf. NAACL* (2004) 161–168 Boston, MA, May.
14. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. *Proc. of the 40th Ann. Conf. of the ACL* (2002) Philadelphia, PA, July.
15. Press, W.H., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes in C++: the Art of Scientific Computing* (2002) Cambridge University Press.
16. Vidal, E.: 1997. Finite-state speech-to-speech translation. *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (1997) 111–114 Munich, Germany.
17. Weaver, W.: Translation. in Locke W. and Booth D. (ed.) *Machine translation of languages: fourteen essays.* (1955) 15–23, John Wiley & Sons, New York.