# Statistical phrase-based models for interactive computer-assisted translation

**Jesús Tomás** and **Francisco Casacuberta**
Instituto Tecnológico de Informática
Universidad Politécnica de Valencia
46071 Valencia, Spain
{jtomas,fcn}@upv.es

## Abstract

Obtaining high-quality machine translations is still a long way off. A post-editing phase is required to improve the output of a machine translation system. An alternative is the so called computer-assisted translation. In this framework, a human translator interacts with the system in order to obtain high-quality translations. A statistical phrase-based approach to computer-assisted translation is described in this article. A new decoder algorithm for interactive search is also presented, that combines monotone and non-monotone search. The system has been assessed in the TransType-2 project for the translation of several printer manuals, from (to) English to (from) Spanish, German and French.

## 1 Introduction

Computers have become an important tool to increase the translator's productivity. In a more extended framework, a *machine translation* (MT) system can be used to obtain initial versions of the translations. Unfortunately, the state of the art in MT is far from being perfect, and a human translator must edit this output in order to achieve high-quality translations.

Another possibility is *computer-assisted translation* (CAT). In this framework, a human translator interacts with the system in order to obtain high-quality translations. This work follows the approach of *interactive CAT* initially suggested by (Foster et al., 1996) and developed in the TransType2 project (SchlumbergerSema S.A. et al., 2001; Barrachina et al., 2006). In this framework, the system suggests a possible translation of a given source sentence. The human translator can accept either the whole suggestion or accept it only up to a certain point (that is, a character prefix of this suggestion). In the latter case, he/she can type one character after the selected prefix in order to direct the system to the correct translation. The accepted prefix and the new corrected character can be used by the system to propose a new suggestion to complete the prefix. The process is repeated until the user completely accepts the suggestion proposed by the system. Figure 1 shows an example of a possible CAT system interaction.

*Statistical machine translation* (SMT) is an adequate framework for CAT since the MT models used can be learnt automatically from a training bilingual corpus and the search procedures developed for SMT can be adapted efficiently to this new interactive framework (Och et al., 2003). *Phrase-based* models have proved to be very adequate statistical models for MT (Tomás et al., 2005). In this work, the use of these models has been extended to interactive CAT.

The organization of the paper is as follows. The following section introduces the statistical approach to MT and section 3 introduces the statistical approach to CAT. In section 4, we review the phrase-based translation model. In section 5, we describe the decoding algorithm used in MT, and how it can be adapted to CAT. Finally, we will present some experimental results and conclusions.

## 2 Statistical machine translation

The goal of SMT is to translate a given source language sentence $s_1^J = s_1...s_J$ to a target sentence $t_1^I = t_1...t_I$. The methodology used (Brown et al., 1993) is based on the definition of a function $Pr(t_1^I|s_1^J)$ that returns the probability that $t_1^I$ is a

| | |
|---:|:---|
| source | `Transferir documentos explorados a otro directorio` |
| interaction-0 | *Move documents scanned to other directory* |
| interaction-1 | Move $\boxed{\text{s}}$*canned documents to other directory* |
| interaction-2 | Move scanned documents to $\boxed{\text{a}}$*nother directory* |
| interaction-3 | Move scanned documents to another $\boxed{\text{f}}$*older* |
| acceptance | Move scanned documents to another folder |

Figure 1: Example of CAT system interactions to translate the Spanish source sentence into English. In interaction-0, the system suggests a translation. In interaction-1, the user accepts the first five characters "Move " and presses the key $\boxed{\text{s}}$, then the system suggests completing the sentence with "*canned documents to other directory*". Interactions 2 and 3 are similar. In the final interaction, the user completely accepts the present suggestion.

translation of a given $s_1^J$. Once this function is estimated, the problem can be reduced to search a sentence $\hat{t}_1^{\hat{I}}$ that maximizes this probability for a given $s_1^J$.

$$\hat{t}_1^{\hat{I}} = \underset{I,t_1^I}{\operatorname{argmax}} \, Pr(t_1^I|s_1^J) = \underset{I,t_1^I}{\operatorname{argmax}} \, Pr(t_1^I)Pr(s_1^J|t_1^I)$$
(1)

Equation 1 summarizes the following three matters to be solved: First, an *output language model* is needed to distinguish valid sentences from invalid sentences in the target language, $Pr(t_1^I)$. Second, a *translation model*, $Pr(s_1^J|t_1^I)$. Finally, the design of an algorithm to *search* for the sentence $\hat{t}_1^{\hat{I}}$ that maximizes this product.

## 3 Statistical computer-assisted translation

In a CAT scenario, the source sentence $s_1^J$ and a given prefix of the target sentence $t_1^i$ are given. This prefix has been validated by the user (using a previous suggestion by the system plus some corrected words). Now, we are looking for the most probable words that complete this prefix.

$$
\begin{aligned}
\hat{t}_{i+1}^{\hat{I}} &= \underset{I,t_{i+1}^I}{\operatorname{argmax}} \, Pr(t_{i+1}^I|s_1^J, t_1^i) \\
&= \underset{I,t_{i+1}^I}{\operatorname{argmax}} \, Pr(t_1^I)Pr(s_1^J|t_1^I) \quad (2)
\end{aligned}
$$

This formulation is very similar to the previous case, but in this one, the search is constrained to the set of possible suffixes $t_{i+1}^I$ instead of the whole target sentences $t_1^I$. Therefore, the same techniques (translation models, decoder algorithm, etc.) which have been developed for SMT can be used in CAT.

Note that the statistical models are defined at word level. However, the CAT interface described in the first section works at character level. This is not a problem: the transformation can be performed in an easy way.

Another important issue is the computational time required by the system to produce a new suggestion. In the CAT framework, real-time is required.

## 4 Phrase-based models

The usual statistical translation models can be classified as single-word based alignment models. Models of this kind assume that an input word is generated by only one output word (Brown et al., 1993). This assumption does not correspond to the characteristics of natural language; in some cases, we need to know a word group in order to obtain a correct translation.

One initiative for overcoming the above-mentioned restriction of single-word models is known as the template-based approach (Och, 2002). In this approach, an entire group of adjacent words in the source sentence may be aligned with an entire group of adjacent target words. As a result, the context of words has a greater influence and the changes in word order from source to target language can be learned explicitly. A template establishes the reordering between two sequences of word classes. However, the lexical model continues to be based on word-to-word correspondence.

A simple alternative to these models has been proposed, the phrase-based (PB) approach (Tomás and Casacuberta, 2001; Marcu and Wong, 2002; Zens et al., 2002). The principal innovation of the phrase-based alignment model is that it attempts to calculate the translation probabilities of word sequences (phrases) rather than of only single words. These methods explicitly learn the probability of a

sequence of words in a source sentence ($\tilde{s}$) being translated as another sequence of words in the target sentence ($\tilde{t}$).

To define the PB model, we segment the source sentence $s_1^J$ into $K$ phrases ($\tilde{s}_1^K$) and the target sentence $t_1^I$ into $K$ phrases ($\tilde{t}_1^K$). A uniform probability distribution over all possible segmentations is assumed. If we assume a monotone alignment, that is, the target phrase in position $k$ is produced only by the source phrase in the same position (Tomás and Casacuberta, 2001) we get:

$$Pr(s_1^J|t_1^I) \propto \sum_{K,\tilde{t}_1^K,\tilde{s}_1^K} \prod_{k=1}^{K} p(\tilde{s}_k|\tilde{t}_k) \qquad (3)$$

where the parameter $p(\tilde{s}|\tilde{t})$ estimates the probability of translating the phrase $\tilde{t}$ into the phrase $\tilde{s}$. A phrase can be comprised of a single word (but empty phrases are not allowed). Thus, the conventional word to word statistical dictionary is included.

If we permit the reordering of the target phrases, a hidden phrase level alignment variable, $\alpha_1^K$, is introduced. In this case, we assume that the target phrase in position $k$ is produced only by the source phrase in position $\alpha_k$.

$$Pr(s_1^J|t_1^I) \propto \sum_{K,\tilde{t}_1^K,\tilde{s}_1^K,\alpha_1^K} \prod_{k=1}^{K} p(\alpha_k|\alpha_{k-1}) \cdot p(\tilde{s}_k|\tilde{t}_{\alpha_k})$$

$$(4)$$

where the distortion model $p(\alpha_k|\alpha_{k-1})$ (the probability of aligning the target segment $k$ with the source segment $\alpha_k$) depends only on the previous alignment $\alpha_{k-1}$ (first order model). For the distortion model, it is also assumed that an alignment depends only on the distance of the two phrases (Och and Ney, 2000):

$$p(\alpha_k|\alpha_{k-1}) = p_0^{|\gamma_{\alpha_k} - \gamma_{\alpha_{k-1}}|} . \qquad (5)$$

There are different approaches to the parameter estimation. The first one corresponds to a direct learning of the parameters of equations 3 or 4 from a sentence-aligned corpus using a maximum likelihood approach (Tomás and Casacuberta, 2001; Marcu and Wong, 2002). The second one is heuristic and tries to use a word-aligned corpus (Zens et al., 2002; Koehn et al., 2003). These alignments can be obtained from single-word models (Brown et al., 1993) using the available public software GIZA++ (Och and Ney, 2003). The latter approach is used in this research.

## 5 Decoding in interactive machine translation

The search algorithm is a crucial part of a CAT system. Its performance directly affects the quality and efficiency of translation. For CAT search we propose using the same algorithm as in MT. Thus, we first describe the search in MT.

### 5.1 Search for MT

The aim of the search in MT is to look for a target sentence $t_1^I$ that maximizes the product $P(t_1^I) \cdot P(s_1^J|t_1^I)$. In practice, the search is performed to maximise a log-linear model of $Pr(t_1^I)$ and $Pr(t_1^I|s_1^J)^\lambda$ that allows a simplification of the search process and better empirical results in many translation tasks (Tomás et al., 2005). Parameter $\lambda$ is introduced in order to adjust the importance of both models. In this section, we describe two search algorithms which are based on multi-stack-decoding (Berger et al., 1996) for the monotone and for the non-monotone model.

The most common statistical decoder algorithms use the concept of partial translation hypothesis to perform the search (Berger et al., 1996). In a partial hypothesis, some of the source words have been used to generate a target prefix. Each hypothesis is scored according to the translation and language model. In our implementation for the monotone model, we define a hypothesis search as the triple $(J', t_1^{I'}, g)$, where $J'$ is the length of the source prefix we are translating (i.e. $s_1^{J'}$); the sequence of $I'$ words, $t_1^{I'}$, is the target prefix that has been generated and $g$ is the score of the hypothesis ($g = Pr(t_1^{I'}) \cdot Pr(t_1^{I'}|s_1^{J'})^\lambda$).

The translation procedure can be described as follows. The system maintains a large set of hypotheses, each of which has a corresponding translation score. This set starts with an initial empty hypothesis. Each hypothesis is stored in a different stack, according to the source words that have been considered in the hypothesis ($J'$). The algorithm consists of an iterative process. In each iteration, the system selects the best scored partial hypothesis to extend in each stack. The extension consists in selecting one (or more) untranslated word(s) in the source and selecting one (or more) target word(s) that are attached to the existing output prefix. The process continues several times or until there are no more hypotheses to extend. The final hypothesis with the highest score and with no untranslated source words is the out-

put of the search.

The search can be extended to allow for non-monotone translation. In this extension, several reorderings in the target sequence of phrases are scored with a corresponding probability. We define a hypothesis search as the triple $(w, t_1^{I'}, g)$, where $w = \{1..J\}$ is the coverage set that defines which positions of source words have been translated. For a better comparison of hypotheses, the store of each hypothesis in different stacks according to their value of $w$ is proposed in (Berger et al., 1996). The number of possible stacks can be very high ($2^J$); thus, the stacks are created on demand. The translation procedure is similar to the previous one: In each iteration, the system selects the best scored partial hypothesis to extend in each created stack and extends it.

### 5.2 Search algorithms for iterative MT.

The above search algorithm can be adapted to the iterative MT introduced in the first section, i.e. given a source sentence $s_1^J$ and a prefix of the target sentence $t_1^i$, the aim of the search in iterative MT is to look for a suffix of the target sentence $\hat{t}_{i+1}^{\hat{I}}$ that maximises the product $Pr(t_1^I) \cdot Pr(s_1^J | t_1^I)$ (or the log-linear model: $\Pr(t_1^{I'}) \cdot \Pr(t_1^{I'} | s_1^{J'})^\lambda$). A simple modification of the search algorithm is necessary. When a hypothesis is extended, if the new hypothesis is not compatible with the fixed target prefix, $t_1^i$, then this hypothesis is not considered. Note that this prefix is a character sequence and a hypothesis is a word sequence. Thus, the hypothesis is converted to a character sequence before the comparison.

In the CAT scenario, speed is a critical aspect. In the PB approach monotone search is more efficient than non-monotone search and obtains similar translation results for the tasks described in this article (Tomás and Casacuberta, 2004). However, the use of monotone search in the CAT scenario presents a problem: If a user introduces a prefix that cannot be obtained in a monotone way from the source, the search algorithm is not able to complete this prefix. In order to solve this problem, but without losing too much efficiency, we use the following approach: Non-monotone search is used while the target prefix is generated by the algorithm. Monotone search is used while new words are generated.

Note that searching for a prefix that we already know may seem useless. The real utility of this phase is marking the words in the target sentence that have been used in the translation of the given prefix.

A desirable feature of the iterative machine translation system is the possibility of producing a list of target suffixes, instead of only one (Civera et al., 2004). This feature can be easily obtained by keeping the $N$-best hypotheses in the last stack. In practice these $N$-best hypotheses are too similar. They differ only in one or two words at the end of the sentence. In order to solve this problem, the following procedure is performed: First, generate a hypotheses list using the $N$-best hypotheses of a regular search. Second, add to this list, new hypotheses formed by a single translation-word from a non-translated source word. Third, add to this list, new hypotheses formed by a single word with a high probability according to the target language model. Finally, sort the list maximising the diversity at the beginning of the suffixes and select the first $N$ hypotheses.

## 6 Experimental results

### 6.1 Evaluation criteria

Four different measures have been used in the experiments reported in this paper. These measures are based on the comparison of the system output with a single reference.

- *Word Error Rate* (WER): Edit distance in terms of words between the target sentence provided by the system and the reference translation (Och and Ney, 2003).

- *Character Error Rate* (CER): Edit distance in terms of characters between the target sentence provided by the system and the reference translation (Civera et al., 2004).

- *Word-Stroke Ratio* (WSR): Percentage of words which, in the CAT scenario, must be changed in order to achieve the reference.

- *Key-Stroke Ratio* (KSR): Number of key-strokes that are necessary to achieve the reference translation divided by the number of running characters (Och et al., 2003) [1].

---

[1] In others works, an extra keystroke is added in the last iteration when the user accepts the sentence. We do not add this extra keystroke. Thus, the KSR obtained in the interaction example of Figure 1, is $3/40$.

| time (ms) | WSR | KSR |
|---:|---|---|
| 10 | 33.9 | 11.2 |
| 40 | 30.9 | 9.8 |
| 100 | 30.0 | 9.3 |
| 500 | 27.8 | 8.5 |
| 13000 | 27.5 | 8.3 |

Table 2: Translation results obtained for several average response time in the Spanish/English "XRCE" task.

| | monotone | | non-monotone | |
|---|---|---|---|---|
| | WSR | KSR | WSR | KSR |
| English/Spanish | 36.1 | 11.2 | 28.7 | 8.9 |
| Spanish/English | 32.2 | 10.4 | 30.0 | 9.3 |
| English/French | 66.0 | 24.9 | 60.7 | 22.6 |
| French/English | 64.5 | 23.6 | 61.6 | 22.2 |
| English/German | 71.0 | 27.1 | 67.6 | 25.6 |
| German/English | 66.4 | 23.6 | 62.0 | 21.9 |

Table 3: Comparison of monotone and non-monotone search in "XRCE" corpora.

| | 1-best | | 5-best | |
|---|---|---|---|---|
| | WSR | KSR | WSR | KSR |
| English/Spanish | 28.7 | 8.9 | 28.4 | 7.3 |
| Spanish/English | 30.0 | 9.3 | 29.7 | 7.6 |
| English/French | 60.7 | 22.6 | 59.8 | 18.8 |
| French/English | 61.6 | 22.2 | 60.7 | 17.6 |
| English/German | 67.6 | 25.6 | 67.1 | 20.9 |
| German/English | 62.0 | 21.9 | 61.6 | 16.5 |

Table 4: CAT results for the "XRCE" task for 1-best hypothesis and 5-best hypothesis.

WER and CER measure the post-editing effort to achieve the reference in an MT scenario. On the other hand, WSR and KSR measure the interactive-editing effort to achieve the reference in a CAT scenario. WER and CER measures have been obtained using the first suggestion of the CAT system, when the validated prefix is void.

### 6.2 Task description

In order to validate the approach described in this paper a series of experiments were carried out using the XRCE corpus. They involve the translation of technical Xerox manuals from English to Spanish, French and German and from Spanish, French and German to English. In this research, we use the *raw* version of the corpus. Table 1 shows some statistics of training and test corpus.

### 6.3 Results

Table 2 shows the WSR and KSR obtained for several average response times, for Spanish/English translations. We can control the response time changing the number of iterations in the search algorithm. Note that real-time restrictions cause a significant degradation of the performance. However, in a real CAT scenario long iteration times can render the system useless. In order to guarantee a fast human interaction, in the remaining experiments of the paper, the mean iteration time is constrained to about 80 ms.

Table 3 shows the results using monotone search and combining monotone and non-monotone search. Using non-monotone search while the given prefix is translated improves the results significantly.

Table 4 compares the results when the system proposes only one translation (1-best) and when the system proposes five alternative translations (5-best). Results are better for 5-best. However, in this configuration the user must read five different alternatives before choosing. It is still to be shown if this extra time is compensated by the fewer key strokes needed.

Finally, in table 5 we compare the post-editing effort in an MT scenario (WER and CER) and the interactive-editing effort in a CAT scenario (WSR and KSR). These results show how the number of characters to be changed, needed to achieve the reference, is reduced by more than 50%. The reduction at word level is slight or none. Note that results from English/Spanish are much better than from English/French and English/German. This is because a large part of the English/Spanish test corpus has been obtained from the index of the technical manual, and this kind of text is easier to translate.

It is not clear how these theoretical gains translate to practical gains, when the system is used by real translators (Macklovitch, 2004).

## 7 Related work

Several CAT systems have been proposed in the TransType projects (SchlumbergerSema S.A. et al., 2001):

In (Foster et al., 2002) a maximum entropy version of IBM2 model is used as translation model. It is a very simple model in order to achieve rea-

|       |                  | English/Spanish | English/German | English/French |
|-------|------------------|-----------------|----------------|----------------|
| Train | Sent. pairs (K)  | 56              | 49             | 53             |
|       | Run. words (M)   | 0.6/0.7         | 0.6/0.5        | 0.6/0.7        |
|       | Vocabulary (K)   | 26/30           | 25/27          | 25/37          |
| Test  | Sent. pairs (K)  | 1.1             | 1.0            | 1.0            |
|       | Run. words (K)   | 8/9             | 9/10           | 11/10          |
|       | Perplexity       | 107/60          | 93/169         | 193/135        |

Table 1: Statistics of the "XRCE" corpora English to/from Spanish, German and French. Trigram models were used to compute the test perplexity.

|                 | WER  | CER  | WSR  | KSR  |
|-----------------|------|------|------|------|
| English/Spanish | 31.1 | 21.7 | 28.7 | 8.9  |
| Spanish/English | 34.9 | 24.7 | 30.0 | 9.3  |
| English/French  | 61.6 | 49.2 | 60.7 | 22.6 |
| French/English  | 58.0 | 48.2 | 61.6 | 22.2 |
| English/German  | 68.0 | 56.9 | 67.6 | 25.6 |
| German/English  | 59.5 | 50.6 | 62.0 | 21.9 |

Table 5: Comparison of post-editing effort in MT scenario (WER/CER) and the interactive-editing effort in CAT scenario (WSR/KSR). Non-monotone search and 1-best hypothesis is used.

sonable interaction times. In this approach, the length of the proposed extension is variable in function of the expected benefit of the human translator.

In (Och et al., 2003) the Alignment-Templates translation model is used. To achieve fast response time, it proposes to use a word hypothesis graph as an efficient search space representation. This word graph is precalculated before the user interactions.

In (Civera et al., 2004) finite state transducers are presented as a candidate technology in the CAT paradigm. These transducers are inferred using the GIATI technique (Casacuberta and Vidal, 2004). To solve the real-time constraints a word hypothesis graph is used. The $N$-best configuration is proposed.

In (Bender et al., 2005) the use of a word hypothesis graph is compared with the direct use of the translation model. The combination of two strategies is also proposed.

## 8   Conclusions

Phrase-based models have been used for interactive CAT in this work. We show how SMT can be used, with slight adaptations, in a CAT system. A prototype has been developed in the framework of

the TransType2 project (SchlumbergerSema S.A. et al., 2001).

The experimental results have proved that the systems based on such models achieve a good performance, possibly, allowing a saving of human effort with respect to the classical post-editing operation. However, this fact must be checked by actual users.

The main critical aspect of the interactive CAT system is the response time. To deal with this issue, other proposals are based on the construction of a word graphs. This method can reduce the generation capability of the fully fledged translation model (Och et al., 2003; Bender et al., 2005). The main contribution of the present proposal is a new decoding algorithm, that combines monotone and non-monotone search. It runs fast enough and the construction of word graph is not necessary.

## Acknowledgments

## References

S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda, H. Net, J. Tomás, E.Vidal, and J.M. Vilar. 2006. Statistical approaches to computer-assisted translation. *In preparation.*

O. Bender, S. Hasan, D. Vilar, R. Zens, and H. Ney. 2005. Comparison of generation strategies for interactive machine translation. In *Proceedings of EAMT 2005 (10th Annual Conference of the European Association for Machine Translation)*, pages 30–40, Budapest, Hungary, May.

A. L. Berger, P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, J. R. Gillett, A. S. Kehler, and R. L. Mercer. 1996. Language translation apparatus and method of using context-based translation models. United States Patent, No. 5510981, April.

P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

F. Casacuberta and E. Vidal. 2004. Machine translation with inferred stochastic finite-state transducers. *Computational Linguistics*, 30(2):205–225.

J. Civera, J. M. Vilar, E. Cubel, A. L. Lagarda, S. Barrachina, E. Vidal, F. Casacuberta, D. Picó, and J. González. 2004. From machine translation to computer assisted translation using finite-state models. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP04)*, Barcelona, Spain.

G. Foster, P. Isabelle, and P. Plamondon. 1996. Word completion: A first step toward target-text mediated IMT. In *COLING '96: The 16th Int. Conf. on Computational Linguistics*, pages 394–399, Copenhagen, Denmark, August.

G. Foster, P. Langlais, and G. Lapalme. 2002. User-friendly text prediction for translators. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP02)*, pages 148–155, Philadelphia, USA, July.

P. Koehn, F. J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL)*, pages 48–54, Edmonton, Canada, June.

E Macklovitch. 2004. The contribution of end-users to the transtype2 project. volume 3265 of *Lecture Notes in Computer Science*, pages 197–207. Springer-Verlag.

D. Marcu and W. Wong. 2002. A phrase-based joint probability model for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Philadelphia, USA, July.

F. J. Och and H. Ney. 2000. Improved statistical alignment models. In *Proc. of the 38th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 440–447, Hong Kong, October.

F. J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, March.

F. J. Och, R. Zens, and H. Ney. 2003. Efficient search for interactive statistical machine translation. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 387.–393, Budapest, Hungary, April.

F. J. Och. 2002. *Statistical Machine Translation: From Single-Word Models to Alignment Templates*. Ph.D. thesis, Computer Science Department, RWTH Aachen, Germany, October.

SchlumbergerSema S.A., Intituto Tecnológico de Informática, Rheinisch Westfälische Technische Hochschule Aachen Lehrstul für Informatik VI, Recherche Appliquée en Linguistique Informatique Laboratory University of Montreal, Celer Soluciones, Société Gamma, and Xerox Research Centre Europe. 2001. TT2. TransType2 - computer assisted translation. Project technical annex.

J. Tomás and F. Casacuberta. 2001. Monotone statistical translation using word groups. In *Procs. of the Machine Translation Summit VIII*, pages 357–361, Santiago de Compostela, Spain.

J. Tomás and F. Casacuberta. 2004. Statistical machine translation decoding using target word reordering. In *Structural, Syntactic, and Statistical Pattern Recongnition*, volume 3138 of *Lecture Notes in Computer Science*, pages 734–743. Springer-Verlag.

J. Tomás, J. Lloret, and F. Casacuberta. 2005. Phrase-based alignment models for statistical machine translation. In *Pattern Recognition and Image Analysis*, volume 3523 of *Lecture Notes in Computer Science*, pages 605–613. Springer-Verlag.

R. Zens, F. J. Och, and H. Ney. 2002. Phrase-based statistical machine translation. *Advances in Artificial Inteligence*, LNAI 2479(25):18–32, September.