SURVEY PAPER

State-of-the-Art Word Reordering Approaches in Statistical Machine Translation: A Survey

Marta R. COSTA-JUSSÀ[†a)], *Nonmember and* **José A. R. FONOLLOSA**[†b)], *Member*

SUMMARY This paper surveys several state-of-the-art reordering techniques employed in Statistical Machine Translation systems. Reordering is understood as the word-order redistribution of the translated words. In original SMT systems, this different order is only modeled within the limits of translation units. Relying only in the reordering provided by translation units may not be good enough in most language pairs, which might require longer reorderings. Therefore, additional techniques may be deployed to face the reordering challenge. The Statistical Machine Translation community has been very active recently in developing reordering techniques. This paper gives a brief survey and classification of several well-known reordering approaches.

key words: *Word Reordering, Statistical Machine Translation*

1. Introduction

Reordering in Statistical Machine Translation is understood as the word-order redistribution of the translated words as shown in Fig. 1. That is why many extended approaches face the challenge of statistical machine translation as a concatenation of two sub-tasks: predicting the collection of words in a translation and deciding the order of the predicted words [1].

This survey is organized as follows. Section 2 briefly describes the statistical machine translation approach. Section 3 focuses on the classification of several reordering approaches. The following sections describe in more detail each of the different main types of word reordering algorithms: straight approaches, source-based, syntax-based and rescoring-based. Finally, Sect. 8 summarizes the survey and provides some remarks about the future research in word reordering.

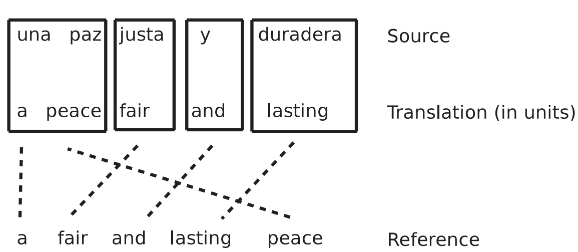

Fig. 1 Source, translation (in units) and reference example. Translation and reference differ in word order.

Manuscript received March 24, 2009.
Manuscript revised June 2, 2009.

[†]The authors are working at Fundació Barcelona Media and Universitat Politècnica de Catalunya, respectively, Spain.
a) E-mail: marta.ruiz@barcelonamedia.org
b) E-mail: adrian@gps.tsc.upc.edu

DOI: 10.1587/transinf.E92.D.2179

2. Statistical Machine Translation Approach

The main goal of SMT is the translation of a text given in some source language into a target language. We are given a source string $s_1^J = s_1 \dots s_j \dots s_J$, which is to be translated into a target string $t_1^I = t_1 \dots t_i \dots t_I$. Among all possible target strings, we will choose the string with the highest probability:

$$\tilde{t}_1^I = \underset{t_1^I}{argmax}\, P(t_1^I|s_1^J) \tag{1}$$

where I and J are the number of words of the target and source sentence, respectively.

In recent systems, a general maximum entropy approach is used in which a log-linear combination of multiple feature functions is implemented [2]. This approach leads to maximising a linear combination of feature functions:

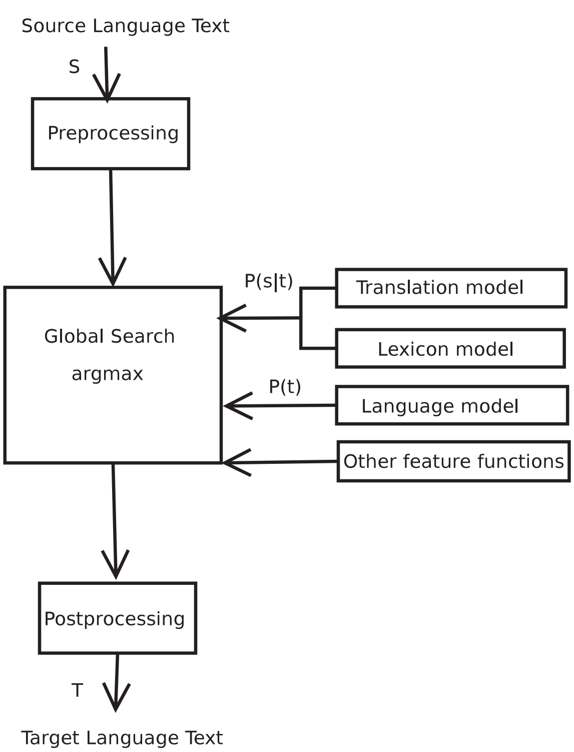

Fig. 2 Architecture of the SMT translation approach based on the log-linear framework approximation: different feature functions are combined in the global search.

$$\tilde{t} = \underset{t}{argmax} \left\{ \sum_{m=1}^{M} \lambda_m h_m(t, s) \right\} \quad (2)$$

The overall architecture of this statistical machine translation (SMT) approach is summarized in Fig. 2. The main model in this structure is the translation model which contains the translation units. Translation units are bilingual sequences of words such as: *a house#una casa*.

By default, the different word order is only modeled within the limits of translation units. However, a good quality translation requires reordering generalization and achieving longer reorderings than the ones modeled within the limits of translation units. Therefore, additional techniques may be deployed to face the reordering challenge.

3. Reordering Approach Classification

Reordering between two languages is a widely studied challenge in SMT since different languages have different word-order requirements.

In recent years, several alternatives to tackle the word ordering in translation have been proposed. These alternatives may be classified into the following groups:

- **Straight heuristic reordering search constraints**, which are founded on the application of distance-based restrictions to the search space.
- **Source reordering approaches**, where reordering rules are defined in the source language. The idea is to reorder the source language in a way that better matches the target language.
- **Reordering based on syntax structures**, which is not carried out using standard phrases. In fact, it solves translation following hierarchical structures.
- **Reordering in rescoring**, typically the rescoring methods have generally provided small accuracy gains given the restriction of being applied to an N-best list.

In the following sections, several reordering approaches are described following the above classification criteria. It is worth noticing that this classification is clearly subjective because there are no clear boundaries among categories. For the sake of simplicity a reordering approach is only included in one category.

4. Straight Heuristic Reordering Search Constraints

The straight reordering approach is to introduce in the search space multiple permutations of the input sentence, aiming at acquiring the right word order of the resulting target sentence. However, systems are forced to restrict their distortion abilities because of the high cost in decoding time that permutations imply. The first SMT decoders introducing reordering capabilities were founded on the brute force of computers, aiming at finding the best hypothesis through traversing a fully reordered graph (the whole permutations of source-side words are allowed in the search). This approach is computationally expensive, even for short input

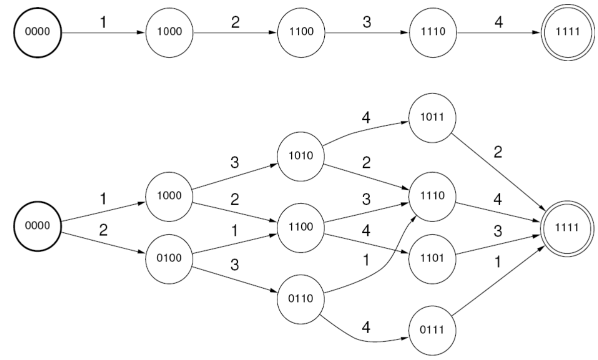

Fig. 3 Permutation graph: monotonic (top) and reordered using IBM constraints (bottom) search.

sentences. In [3], decoding with arbitrary word reorderings is shown to be NP-hard.

Therefore, different distance-based constraints are commonly used to make the search feasible. The use of these constraints generates simple reordering models which imply a necessary balance between translation accuracy and efficiency. One simple model is a *'weak'* distance-based distortion model that was initially used to penalize the longest reorderings, only allowed if sufficiently promoted by the rest of models [4], [5]. Other reordering constraints are the ones based on: IBM [6], [7]; inversion transduction grammars (ITG) [8]; maxjumps [9] and lexical parameters [10].

IBM constraints are allowed to deviate from the monotonic order by postponing translations up to a limited number of words, *i.e.* at each state, translations can be performed on the first l word positions not yet covered. At each state, the covered words are shown in the form of a bit vector. Figure 3 shows the permutations graph computed for a monotonic (top) and a reordered (bottom) search of an input sentence of $J = 4$ words. The reordered graph shows the valid permutations computed following IBM constraints, defined in [11], for a value of $l = 2$.

Another simple reordering model is the flat model: ITG [8], which is not content dependent either. The constituents of the input sentence are interpreted as blocks. Two consecutive blocks are either kept in a monotone order or inverted in order. Therefore, the ITG allows all permutations defined by all binary branching structures where any constituent may be swapped in order (see an example in Fig. 4). The flat model assigns constant probabilities for monotone and non-monotone order. The two probabilities can be set to prefer monotone or non-monotone orientations, depending on the language pairs. [11] shows a comparison between IBM and ITG constraints. [12] proposes a novel solution for unit reordering. Under the ITG constraint they need to consider just two kinds of reorderings, straight and inverted between two consecutive blocks. In this way, reordering can be modeled as a problem of classification with only two labels. The main drawback is that not all reorderings can be captured [13].

The maxjumps approach [9] limits the number of re-

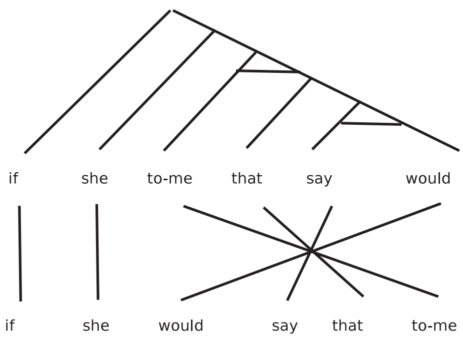

Fig. 4 Example of reordering using ITG. The alignments are added below the tree, and the horizontal bars in the tree indicate the swap.

orderings for a search path (whole translation) to a given number using two strategies:

- A distortion limit (m). A source word (phrase or tuple) is only allowed to be reordered if it does not exceed a distortion limit, measured in words.
- A reordering limit (j). Any translation path is only allowed to perform j reordering jumps.

In view of content-independence of the distortion and flat reordering models, [10], [14] propose the lexicalized reordering model. This reordering model learns local orientations with probabilities for each bilingual phrase from training data. Therefore, for each bilingual unit, the basic idea is to learn the likelihood that the particular unit directly follows a previous bilingual unit (monotone). This particular unit is then swapped with a previous bilingual unit (swap) or it is determined to be disconnected from the previous unit (discontinuous). The model learns local orientations (monotone, non-monotone) with probabilities for each bilingual phrase from the training material. While decoding, the model attempts to find a Viterbi local orientation sequence. The approach reports interesting performance gains. However, since reorderings are related to concrete phrases, researchers have to design their systems carefully in order not to cause other problems, e.g. the data sparseness problem. This lexicalized reordering approach is implemented in the open source Moses toolkit†.

Additionally to the reordering constraints, a reordering approach is defined by a distortion model. The distortion models are allowed to assign a score to each reordering constraint or permutation explored by the search. A standard distance-based distortion model is defined as follows [9]:

$$P(u_1^K) = \exp\left(-\sum_{k=1}^{K} d_k\right)$$

where d_k is the distance between the first word of the k^{th} translation unit (u), and the last word+1 of the $(k-1)^{th}$ translation unit. To sum up, the main idea of distance-based reorderings is that if N words are skipped, a penalty of N will be paid regardless of which words are reordered. This model takes the risk of penalizing long distance jumps which are common between two languages with very different orders.

5. Source Reordering Approaches

Machine translation is the automation of a human task. The big question here is how the human translation process work. One extended theory is that: *The translator will undo the syntactic structure of the source text and then formulate the corresponding message in the target language*††. Following the human model, the source reordering approaches try to undo the structure/order of the source text to better match the target text. This approach sometimes is refered to as word harmonization.

The reordering alternative detailed in this section aims at applying a set of permutations to the words of the input sentence to help the system build the translation hypothesis in the right word order. The main difference with the previous straight heuristic reordering search constraints is that these approaches attempt to specifically learn the source reordering that better matches the target language. The reordering rules and/or constraints in the source language may be defined following different criteria such as:

- Deterministic reordering. Word-order harmonization was first proposed in [15], where morpho-syntactic information was used to account for the reorderings needed between German and English. In this work reordering was done by prepending German verb prefixes and by treating interrogative sentences using syntactic information. In [16] and [17], the source corpus is reordered following a set of rules. In the former, the rules are lexicalized, in the latter, these rules have been automatically learned using lexical and/or morphological information, i.e. *Part of Speech* (POS). The decoder search is monotonic. In [18], the main novelty is that the rules are learned statistically from the training set and new rules can be inferred without requiring additional knowledge. Figure 5 (top) shows how reordering and decoding are decoupled under this approach in two main blocks. One of the main drawbacks of this approach is that it takes reordering decisions in a preprocessing step, thus discarding much of the information available in the global search that could play an important role if it was taken into account. Thus far, the reordering challenge is only tackled in preprocessing, so the errors introduced in this step may remain in the final translation output.
- Clause restructuring. Here, the methods use syntactic information to reorder source words in SMT as a preprocessing step. [19] proposes a set of automatically learned reordering rules (using morpho-syntactic information in the form of POS tags) which are then applied to a French-English translation task. In [20] a German parse tree is used for moving German verbs towards

†http://www.statmt.org/moses/
††http://accurapid.com/journal/05theory.htm

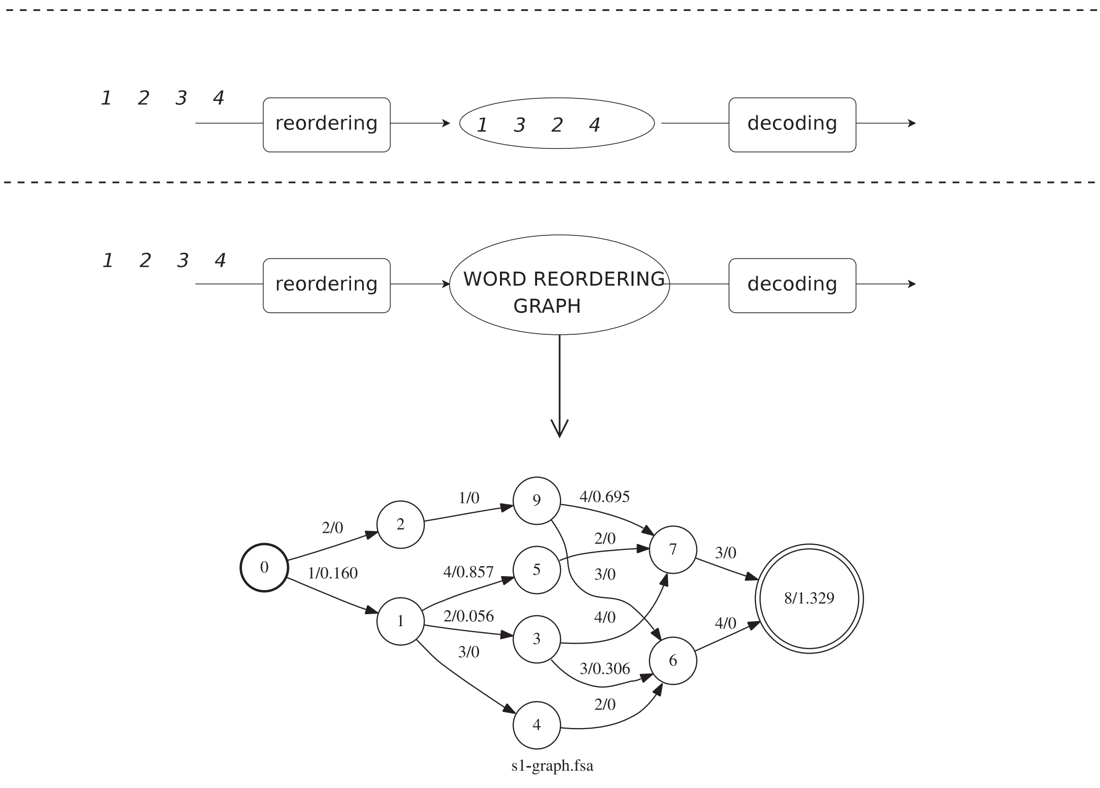

Fig. 5 Source reordering approaches: deterministic reordering (up) and input reordering graph (down).

the beginning of the clause. [21] uses manually written Chinese rules for moving Chinese structure towards the English one. [22] employs dependency trees to capture reordering needs of an Arabic-to-English translation system. In these works, the source reordering is complemented with a local reordering in search.

- Input reordering graph. A natural evolution of coupling the deterministic source reordering and the translation steps is shown in Fig. 5 (bottom). It consists in using a word graph, containing the *N*-best reordering decisions, instead of the single-best used in the above strategy. The reordering problem is equally approached by alleviating the difficulty of needing highly accurate reordering decisions in preprocessing. The final decision is delayed, to be subsequently in the global search, where all the information is then available. Inspired by [23], the following approaches permute the source sentence and provide a source input graph that extends the search graph. In [24], [25], reordering is addressed through a source input graph. In this case, the reordering hypotheses are defined from a set of linguistically motivated rules (either using *Part of Speech*; chunks; or parse trees). Following the same rewriting idea and making use of a permutation graph to couple reordering and decoding, [26] confronts the reordering challenge using the powerful statistical machine translation techniques. It employs either a set of automatically learnt word classes [27] or linguistic POS tag classes showing slightly better accuracy results for an Spanish-English task than those shown in [24].

6. Reordering Based on Syntax Structures

In spite of the great success of the phrase-based systems, a key limitation of these systems is that they make little or no direct use of syntactic information. However, it appears likely that syntactic information can be of great help in order to accurately model many systematic differences [28] between the word order of different languages.

This section is dedicated to describe the reordering approaches which incorporate syntactic information. Therefore, we briefly report several translation approaches that use this syntactic information and, more particularly, focus on the way reorderings are handled in each approach. The most popular approaches are: syntax-based, hierarchical phrase-based and syntactic phrase-based.

In recent years, syntax-based statistical machine translation has begun to emerge, aiming at applying statistical models to structured data. Advances in natural language parsing, especially the broad-coverage parsers trained from tree banks, for example [29], have made possible the utilization of structural analysis of different languages. A syntax-based translator consists of two components, a source language parser and a recursive converter, which is usually modeled as a top-down tree-to-string transducer. This approach models the two languages using tree-based transduction rules or a synchronous grammar, possibly probabilistic. Machine translation is done either as a stochastic tree-to-tree transduction or synchronous parsing process. Some methods linguistically syntax-based are [30], [31]. The former makes use of bitext grammars to parse both the source and target languages. The latter makes use of syntactic information only in the target language. The main advantage of these approaches is that they can theoretically cover long distance reorderings in case reorderings respect linguistic syntax in translation. Empirical evidence only partially supports this. In fact, reordering tends to respect the boundaries of syntactic phrases, but as shown by [13] there are some systematic exceptions.

The above syntax-based models share a key weakness: they enable only word-to-word translation, which requires a reordering decision for each word. Given that the phrase-based models solve this, the so-called hierarchical phrase-based model tries to benefit from the insights behind both syntax-based models and phrase-based models. The hierarchical phrase-based model uses constituent trees [32], [33] in which phrases are reorganized into hierarchical by reducing subphrases to variables. Context-free rules are inferred from string-to-string pairs (here, no parsing is required). The grammar can be automatically learned from a bitext. This template-based scheme is able to capture and generalize the reorderings of phrases.

Finally, much active research aims to combine the advantages of the above reordering approaches using dependency treelets. In [34]–[36], a dependency tree-based reordering model is inferred from aligned string-tree pairs. Therefore, reordering and translation operations are defined on tree fragments rather than strings.

Translation and reordering models presented in this section have structures and parametrization that are radically different from the reordering models implemented in the phrase-based SMT systems. Therefore, new decoders have emerged that aim at dealing with pairs of languages with very different syntactic structures for which the word context introduced in phrase-based decoders is not sufficient to cope with long reorderings. They have gained much popularity because of the significant improvements made by exploiting the power of synchronous rewriting systems. Syntax-directed systems have been typically attacked with the argument of showing a main weakness in their poor efficiency results. However, this argument has been recently overridden through the development of those new decoders, which show significant improvements when handling syntactically divergent language pairs under large-scale data translation tasks. An example of such a system can be found in [37], which has obtained state-of-the-art results in Arabic-to-English and Chinese-to-English large-sized data tasks.

7. Reordering in Rescoring

Rescoring techniques have also been proposed as a method for using syntactic information to identify translation hypotheses expressed in the right target word order [38]–[40]. In these approaches, a baseline system is used to generate N-best translation hypotheses. Syntactic features are then used in a second model that reranks the N-best lists, in an attempt to improve over the baseline approach. [38] apply a reranking approach to the sub-task of noun-phrase translation.

[41] introduces super-tag information or '*almost parsing*' [42] into a standard phrase-based SMT system in the reranking process. It is shown how syntactic constraints can improve translation quality for an Arabic-to-English translation task. Later, in [43], the same researchers introduce the super-tag information into the overall search in the form of an additional log-linearly combined model.

Another kind of approach is to use syntactic information in rescoring methods. [44] apply a reranking approach to the sub-task of noun-phrase translation. [40] describes the use of syntactic features in reranking the output of a full translation system, but the syntactic features give very small gains.

8. Summary and Future Research Lines

This paper has presented a brief classification of several reordering approaches in Statistical Machine Translation without aiming at its completeness.

Regarding reordering approaches refered to as straight heuristic reordering search constraints, the main drawback is that long distance reorderings are not completely solved. Additionally, the generalization of reorderings is not clearly achieved. Most methods included in the source reordering approaches are limited by their sparseness. Some reordering based on syntax structures are theoretically able to cope with sparseness and long distance reorderings. However, in practice, their performance is not far from the above techniques due to the several reasons (i.e. parsing errors). Finally, the reordering in rescoring, in most cases, seems to have an structural problem given the restriction of being applied to an N-best list.

To sum up, it can be said that long distance reorderings are not completely solved. The main drawback of most reordering approaches is that although they can generalize reorderings, long-distance reorderings are not usually achieved. Therefore, translation pairs as Chinese-Spanish still present problems in word reordering.

In future research, the Statistical Machine Translation community has to focus on this challenge (long distance reorderings) which may be accurately solved by further exploiting syntactic information and introducing newer methods from neighbouring fields such as Machine Learning and Information Retrieval. Clearly, introducing methodologies from other fields could open new frontiers in the near future.

Acknowledgements

The authors would especially like to thank the anonymous reviewers of this paper, who helped improve the quality of the presentation with their insightful suggestions. This work has been partially founded by the Spanish Department of Education and Science through the *Juan de la Cierva* fellowship program.

References

[1] P.C. Chang and K. Toutanova, "A discriminative syntactic word order model for machine translation," Proc. 45th Annual Meeting of the Association for Computational Linguistics, pp.9–16, Prague, Czech Republic, June 2007.
[2] F. Och, "Minimum error rate training in statistical machine translation," Proc. 41th Annual Meeting of the Association for Computational Linguistics, pp.160–167, Sapporo, July 2003.
[3] K. Knight, "Decoding complexity in word-replacement translation models," Computational Linguistics, vol.25, no.4, pp.607–615, Dec. 1999.
[4] F. Och and H. Ney, "The alignment template approach to statistical machine translation," Computational Linguistics, vol.30, no.4, pp.417–449, Dec. 2004.
[5] P. Koehn, F. Och, and D. Marcu, "Statistical phrase-based translation," Proc. Human Language Technology Conf., HLT-NAACL'03, pp.48–54, Edmonton, Canada, May 2003.
[6] A. Berger, S.D. Pietra, and V.D. Pietra, "A maximum entropy approach to natural language processing," Computational Linguistics, vol.22, no.1, pp.39–72, March 1996.
[7] C. Tillmann and H. Ney, "Word reordering and a dynamic programming beam search algorithm for statistical machine translation," Computational Linguistics, vol.29, no.1, pp.97–133, March 2003.
[8] D. Wu, "A polynomial-time algorithm for statistical machine translation," Annual Meeting of the Association for Computational Linguistics, Santa Cruz, June 1996.
[9] J. Crego, J. Mariño, and A. de Gispert, "An ngram-based statistical machine translation decoder," Proc. 9th Int. Conf. on Spoken Language Processing, ICSLP'05, pp.3185–3188, Lisboa, April 2005.
[10] C. Tillmann, "A unigram orientation model for statistical machine translation," Proc. Human Language Technology Conf., HLT-NAACL'04, pp.101–104, Boston, May 2004.
[11] R. Zens and H. Ney, "A comparative study on reordering constraints in statistical machine translation," Annual Meeting of the Association for Computational Linguistics, pp.144–151, Sapporo, July 2003.
[12] D. Xiong, Q. Liu, and S. Lin, "Maximum entropy based phrase reordering model for statistical machine translation," Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp.521–528, Sydney, Australia, July 2006.
[13] B. Wellington, S. Waxmonsky, and I.D. Melamed, "Empirical lower bounds on the complexity of translational equivalence," Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp.977–984, Sydney, Australia, July 2006.
[14] P. Koehn, A. Amittai, A. Birch, C. Callison-Burch, M. Osborne, D. Talbot, and M. White, "Edinburgh system description for the 2005 iwslt speech translation evaluation," Proc. International Workshop on Spoken Languages Translation, Pittsburgh, Oct. 2005.
[15] S. Nießen and H. Ney, "Morpho-syntactic analysis for reordering in statistical machine translation," Proc. MT-Summit VII, pp.247–252, Sept. 2001.
[16] Y.S. Lee, Y. Al-Onaizan, K. Papineni, and S. Roukos, "IBM spoken language translation system," TC-STAR Workshop on Speech-to-Speech Translation, pp.13–18, Barcelona, Spain, June 2006.
[17] M. Popovic and H. Ney, "Pos-based word reorderings for statistical machine translation," 5th Int. Conf. on Language Resources and Evaluation, LREC'06, pp.1278–1283, Genoa, May 2006.
[18] M. Costa-jussà and J. Fonollosa, "Using reordering in statistical machine translation based on alignment block classification," 6th Int. Conf. on Language Resources and Evaluation, LREC'08, Marrakech, 2008.
[19] F. Xia and M. McCord, "Improving a statistical mt system with automatically learned rewrite patterns," Proc. 20th International Conference on Computational Linguistics, p.508, Morristown, 2004.
[20] M. Collins, P. Koehn, and I. Kucerová, "Clause restructuring for statistical machine translation," Proc. 43rd Annual Meeting of the Association for Computational Linguistics, pp.531–540, Michigan, 2005.
[21] C. Wang, M. Collins, and P. Koehn, "Chinese syntactic reordering for statistical machine translation," Proc. 2007 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp.737–745, Prague, June 2007.
[22] N. Habash, "Syntactic preprocessing for statistical machine translation," Proc. MT-Summit XI, Sept. 2007.
[23] K. Knight and Y. Al-Onaizan, "Translation with finite-state devices," Proc. 4th Conf. of the Association for Machine Translation in the Americas, AMTA'02, pp.421–437, Langhorne, Dec. 1998.
[24] J. Crego and J. Mariño, "Improving SMT by coupling reordering and decoding," Machine Translation, vol.20, no.3, pp.199–215, 2007.
[25] Y. Zhang, R. Zens, and H. Ney, "Chunk-level reordering of source language sentences with automatically learned rules for statistical machine translation," Proc. Human Language Technology Conf. (HLT-NAACL'06): Proc. Workshop on Syntax and Structure in Statistical Translation (SSST), pp.1–8, Rochester, April 2007.
[26] M. Costa-jussà and J. Fonollosa, "An ngram-based reordering model," Computer Speech and Language, vol.23, pp.362–375, 2009.
[27] F. Och, "An efficient method for determining bilingual word classes," Proc. 9th Conf. of the European Chapter of the Association for Computational Linguistics, pp.71–76, June 1999.
[28] B. Bonnie, "Machine translation divergences: A formal description and proposed solution," Computational Linguistics, vol.20, no.4, pp.597–633, 1994.
[29] M. Collins, Head-driven Statistical Models for Natural Language Parsing, PhD Thesis, University of Pennsylvania, 1999.
[30] K. Yamada and K. Knight, "A decoder for syntax-based statistical MT," Proc. 40th Annual Meeting of the Association for Computational Linguistics, pp.303–310, July 2002.
[31] D. Wu, "Stochastic inversion transduction grammars and bilingual parsing of parallel corpora," Computational Linguistics, vol.23, no.3, pp.377–403, Sept. 1997.
[32] D. Chiang, "A hierarchical phrase-based model for statistical machine translation," Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), Ann Arbor, pp.263–270, Michigan, June 2005.
[33] T. Watanabe, H. Tsukada, and H. Isozaki, "Left-to-right target generation for hierarchical phrase-based translation," Proc. COLING-ACL'06, pp.777–784, Sydney, Australia, July 2006.
[34] C. Quirk, A. Menezes, and C. Cherry, "Dependency treelet translation: Syntactically informed phrasal SMT," Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pp.271–279, Ann Arbor, Michigan, June 2005.
[35] P. Langlais and F. Gotti, "Phrase-based SMT with shallow treephrases," Proc. Workshop on Statistical Machine Translation, pp.39–46, June 2006.
[36] M. Galley and M. Hopkins, "What's in a translation rule?," Proc. Human Language Technology Conf., HLT-NAACL'04, pp.273–280, Boston, MA, May 2004.
[37] D. Marcu, W. W, A. Echihabi, and K. Knight, "SPMT: Statistical machine translation with syntactified target language phrases," Proc. Conf. on Empirical Methods in Natural Language Processing, EMNLP'06, pp.44–52, Sydney, Australia, July 2006.
[38] K. Koehn and K. Knight, "Empirical methods for compound splitting," Proc. 10th Conf. of the European Chapter of the Association

for Computational Linguistics, pp.347–354, April 2003.
[39] F.J. Och, D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smith, K. Eng, V. Jain, Z. Jin, and D. Radev, "A smorgasbord of features for statistical machine translation," Proc. Human Language Technology Conf., HLT-NAACL'04, pp.161–168, 2004.
[40] L. Shen, A. Sarkar, and F. Och, "Discriminative reranking for machine translation," Proc. Human Language Technology Conf., HLT-NAACL'04, ed. D.M. Susan Dumais and S. Roukos, pp.177–184, Boston, Massachusetts, USA, May 2004.
[41] H. Hassan, M. Hearne, A. Way, and K. Sima'an, "Syntactic phrase-based statistical machine translation," 1st IEEE/ACL Workshop on Spoken Language Technology, Dec. 2006.
[42] S. Bangalore and A. Joshi, "Supertagging: An approach to almost parsing," Computational Linguistics, vol.25, no.2, pp.237–265, 1999.
[43] H. Hassan, K. Sima'an, and A. Way, "Supertagged phrase-based statistical machine translation," Proc. 45th Annual Meeting of the Association for Computational Linguistics, pp.288–295, Prague, Czech Republic, June 2007.
[44] P. Koehn and K. Knight, "Feature-rich statistical translation of noun phrases," Proc. 41st Annual Meeting of the Association for Computational Linguistics, pp.311–318, 2003.

Marta R. Costa-jussà received her PhD degree in Electrical Engineering from the Universitat Politècnica de Catalunya, UPC, Spain, in 2008. She has been twice at LIMSI-CNRS (Paris) as a research visitor working both in speech recognition and in statistical machine translation. In April 2009 she started as a research fellow in the Fundació Barcelona Media under a *Juan de la Cierva* fellowship program given by the Spanish Department of Education and Science. She combines her research work together with the University teaching and consultancy in several translation companies. Recently, she has participated in spoken language translation projects (local and EU funded) as TC-STAR, AVIVAVOZ and TECNOPARLA.

José A. R. Fonollosa received his PhD degree in Electrical Engineering from the Universitat Politècnica de Catalunya (UPC), Spain, in 1989. Since 1986, he is with the Department of Signal Theory and Communications of the UPC, where he is now full professor. From August 1991 to July 1992 he was visiting the Signal and Image Processing Institute, University of Southern California. He received the 1992 Marconi Young Scientist Award. Since 2006, he is director of the Center for Language and Speech Technologies and Applications (TALP Research Center). Prof. Fonollosa has published more than 100 scientific papers in statistical signal analysis and their application in communication systems, speech processing and statistical machine translation. During the last 5 years, he has concentrated his efforts on (spoken) language translation participating in local as well as EU funded projects as FAME, LC-STAR and TC-STAR.