

Using Alignment Templates to Infer Shallow-Transfer Machine Translation Rules

Felipe Sánchez-Martínez¹ and Hermann Ney

Lehrstuhl für Informatik VI – Computer Science Department
RWTH Aachen University, D-52056 Aachen, Germany
`fsanchez@dlsi.ua.es`
`ney@informatik.rwth-aachen.de`

Abstract. When building rule-based machine translation systems, a considerable human effort is needed to code the transfer rules that are able to translate source-language sentences into grammatically correct target-language sentences. In this paper we describe how to adapt the alignment templates used in statistical machine translation to the rule-based machine translation framework. The alignment templates are converted into structural transfer rules that are used by a shallow-transfer machine translation engine to produce grammatically correct translations. As the experimental results show there is a considerable improvement in the translation quality as compared to word-for-word translation (when no transfer rules are used), and the translation quality is close to that achieved when hand-coded transfer rules are used. The method presented is entirely unsupervised, and needs only a parallel corpus, two morphological analysers, and two part-of-speech taggers, such as those used by the machine translation system in which the inferred transfer rules are integrated.

1 Introduction

When building rule-based machine translation (MT) systems, a considerable human effort is needed to code transfer rules. Transfer rules are needed when translating source language (SL) into target language (TL) to perform some syntactic and lexical changes. In this paper we explore the use of alignment templates (ATs) [1,2,3], already used in the statistical machine translation framework, as structural transfer rules within a shallow-transfer MT system. An alignment template (AT) can be defined as a generalization performed over aligned phrase pairs (or *translation units*) using word classes instead of the words themselves. Our approach uses some linguistic information to automatically learn from a parallel corpus a set of ATs that are then used as transfer rules. The method is entirely unsupervised and only needs a parallel corpus, two morphological

¹ Permanent address: Transducens Group, Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, E-03071, Alacant, Spain.

analysers, and two part-of-speech taggers, more likely the two morphological analysers, and the two part-of-speech taggers used by the MT system in which the learned rules are then integrated.

To adapt the ATs to a shallow-transfer MT system the part-of-speech, or lexical category, of each word is used as the word class when extracting the ATs (see section 2). Moreover, the method needs to be provided with some linguistic information, such as the set of *closed lexical categories*² on both languages. After extracting the ATs two different criteria have been tested to choose the AT to apply when more than one can be applied. Those criteria are: (1) to choose always the longest AT that can be applied, and (2) to choose always the most frequent AT that can be applied. Nevertheless, before applying either criterion, infrequent ATs are discarded (see section 3.3). The method we present has been tested using an existing shallow-transfer MT system, and the experimental results show that the use of ATs within a shallow-transfer MT system drastically improves the translation quality as compared to word-for-word translation, i. e. when no transfer rules are used.

There have been other attempts to learn automatically or semi-automatically those structural transformations needed to produce correct translations into the TL. Those approaches can be classified according to the translation framework to which the learned rules are applied. On the one hand, some approaches learn transfer rules to be used in rule-based MT [4,5]. In [4,5] transfer rules for MT involving “minor” languages (e. g. Quechua) are learned with very limited resources. To this end, a small parallel corpus (of a few thousand sentences) is built with the help of a small set of bilingual speakers of the two languages. The parallel corpus is obtained by translating a controlled corpus from a “major” language (English or Spanish) to a “minor” language by means of an elicitation tool. This tool is also used to graphically annotate the word alignments between the two sentences. Finally, some hierarchical syntactic rules, that can be seen as a context-free transfer grammar, are inferred from the aligned parallel corpus.

On the other hand, in the example-based machine translation (EBMT) framework some research works deal with the problem of inferring some kind of translations rules called *translation templates* [6,7,8]. A translation template can be defined as a bilingual pair of sentences in which corresponding units (words or phrases) are coupled and replaced by variables. In [9] there is an interesting review of the different research works dealing with translation templates. In [7] the author uses a parallel corpus and some linguistic knowledge in the form of equivalence classes (both syntactic and semantic) to perform a generalization over the bilingual examples collected. The method works by replacing each word by its corresponding equivalence class and then using a set of grammar rules to replace patterns of words and tokens by more general tokens. In [8] the authors formulate the acquisition of translation templates as a machine learning problem. In that work the translation templates are learned from the differences and similarities observed in a set of different translation examples, using no morpho-

² Closed lexical categories are those categories that cannot easily grow by adding new words to the dictionaries (articles, pronouns, conjunctions, etc.).

logical information at all. In [6] a bilingual dictionary and a syntactic parser are used to determine the correspondences between translation units while learning the translation templates. In any case, the translation templates used in EBMT differ from the approach presented in this paper, firstly because our approach is largely based on part-of-speech information and the inferred translation rules are flatter, less structured and non-hierarchical (because of this, they are suitable for shallow-transfer MT translation); and secondly, because the way in which the transformations to apply are chosen differs from those used in the EBMT framework.

The rest of the paper is organized as follows: the next section overviews the alignment templates approach; section 3 explains how to adapt the ATs in order to use them as transfer rules within a shallow-transfer MT system. Section 4 overviews the shallow-transfer MT system used to evaluate the presented approach, and describes the experiments conducted and the results achieved. Finally, in section 5 we draw some conclusions and outline future work.

2 The alignment templates approach

Alignment templates (ATs) [1,2,3] were introduced in the statistical machine translation framework as a feature function to be used in the log-linear maximum entropy model [10]. An AT represents a generalization performed over aligned phrase pairs³ using word classes. The ATs are learned following a three-stage procedure: First, word alignments are computed, then aligned phrase pairs are extracted; and finally, a generalization over the extracted aligned phrase pairs is performed using word classes instead of the words themselves. In [2] the word classes used to perform such generalization were automatically obtained using the method described in [11]. However, using non-automatic classes such as part-of-speech or semantic categories is feasible as suggested in [2]. The use of word classes allows for generalization, considering word reordering, preposition changes and other dissimilarities between SL and TL.

Formally, an AT is a tuple $z = (S_n, T_m, A)$ that describes the alignment A between a source sequence S_n of n SL word classes and a target sequence T_m of m TL word classes.

3 Alignment templates for shallow-transfer machine translation

For a correct extraction, filtering, and application of the alignment templates (ATs) as transfer rules some linguistic knowledge needs to be used. This section explains how an indirect rule-based MT systems works and introduces the linguistic knowledge used. Then it explains in detail how to extract, filter and apply the ATs.

³ Linguists would not agree with our use of the word “*phrase*”. In this paper phrase means any sequence of consecutive words, not necessarily whole syntactic constituents.

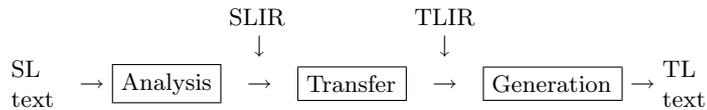


Fig. 1. Main structure of an (indirect) rule-based transfer MT system. Source language (SL) text is analyzed and converted into an intermediate representation (SLIR), then transformations are applied giving as a result a target language intermediate representation (TLIR), finally the TLIR is used to generate the output translation into the target language (TL).

3.1 Indirect rule-based machine translation

Shallow-transfer MT is a special case of the (indirect) rule-based transfer MT framework. Shallow transfer rules simply detect patterns of lexical forms and perform some lexical and syntactic changes. Figure 1 summarizes how an indirect rule-based MT system works. First, the SL text is analyzed and converted into a source-language intermediate representation (SLIR); then, transformations are applied, transforming the SLIR into a target language intermediate representation (TLIR); finally, the TLIR is used to generate the output translation.

Usually the transformations to apply consist in using a bilingual dictionary to translate each word (lexical transfer) and applying some rules to ensure the grammatical correctness of the translation output (structural transfer). The work reported in this paper focused on automatically learning the structural transfer rules needed to produce correct translations; to this end, the ATs approach already introduced in section 2 is used.

In order to apply the ATs within a shallow-transfer MT system the parallel corpus must be preprocessed. The SL part must be in the format in which the input will be presented to the transfer module, that is the SLIR; analogously, the TL part must be in the format in which the transfer module will produce the output, that is, the TLIR. Notice that for the SL this preprocessing is exactly the same done by the MT system when translating an input text, and that the preprocessing of the TL part is equivalent to that for the SL.

Indirect rule-based MT systems, such as the shallow-transfer MT system used in the experiments (see section 4.1), are usually based on morphological and bilingual dictionaries. In shallow-transfer MT systems the source language intermediate representation (SLIR) and the target language intermediate representation (TLIR) are usually based on lexical forms containing the lemma, the part-of-speech, and the inflection information for each word. For example, an intermediate representation for the English sentence *The green houses* would be *the-(art) green-(adj) house-(noun,plural)*.

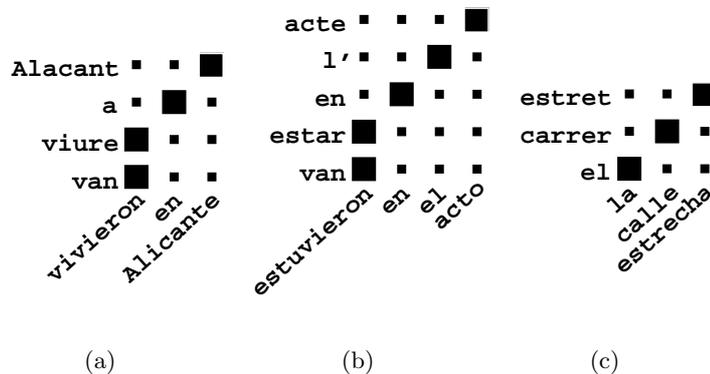


Fig. 2. Examples of the kind of alignments that can be found in a Spanish–Catalan parallel corpus.

3.2 Extraction and filtering of the alignment templates

As the transformations to apply are mainly based on the part-of-speech of SL and TL words, the method to adapt the ATs to a shallow-transfer MT system needs to be provided with the following linguistic information:

- The set of *closed lexical categories* in both source and target languages. Closed lexical categories are those categories that cannot easily grow by adding new words to the dictionaries: articles, auxiliary verbs, pronouns, etc. From now on we will refer as *closed words* to those words whose part-of-speech is in the set of closed lexical categories. Analogously, we will refer as *open words* to those words whose part-of-speech is not in the set of closed lexical categories.
- The set of *dominant categories* in the target language. A dominant category is a lexical category which usually propagates its inflection information (such as gender or number) to neighboring lexical categories. Usually the only dominant category is the noun, which propagates its gender and number to articles and adjectives. From now on we will refer as *dominant words* to those words whose part-of-speech is in the set of dominant categories.

To extract the ATs, the part-of-speech (including all the inflection information such as gender, number or verb tense) is used to assign a word class to each open word. For closed words, the lemma is also used to define the word class, therefore each closed word is in its own single class. For example, the English nouns *book* and *house* would be in the same word class, but the prepositions *to* and *for* would be in different classes even if they have the same part-of-speech. In this way the method is allowed to learn some changes such as preposition changes or auxiliary verbs usages in the target language.

Figure 2 shows examples of the kind of alignments that can be found in a Spanish–Catalan parallel corpus. In figure 3 the ATs extracted from the align-

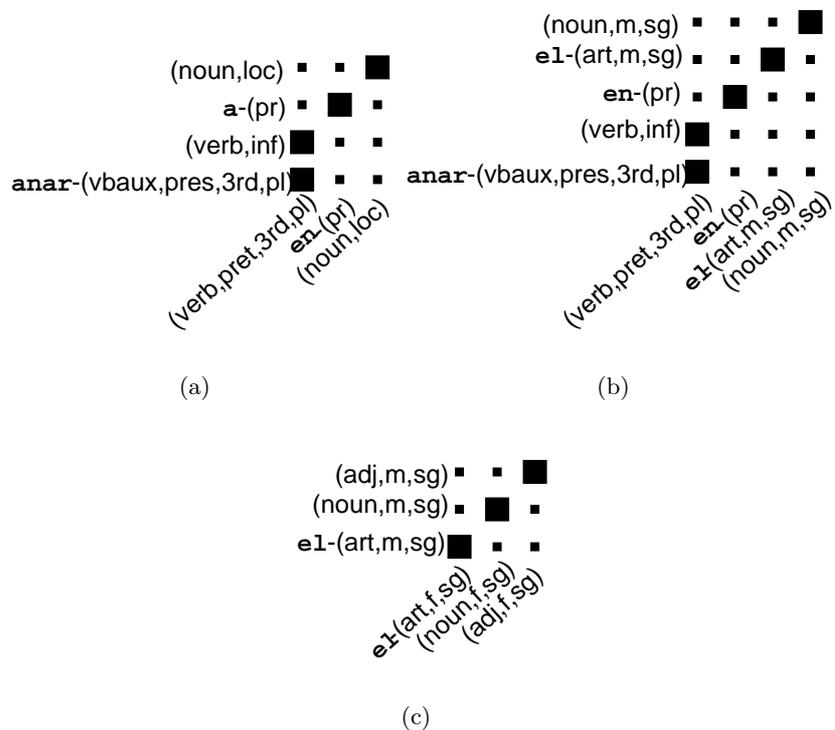


Fig. 3. Alignment templates (ATs) obtained from the alignments found in figure 2. These ATs have been extracted using the part-of-speech as word classes, but putting each closed word in its own single word class (see section 3.2) .

ments found in figure 2 are shown. To extract these ATs the part-of-speech of each word has been used as word class; remember however, that closed words have their own single class (in the example reported, the prepositions *en* and *a*, the article *el*, and the Catalan verb *anar* that works as an auxiliary verb). As can be seen the ATs represent a generalization of the transformations to apply when translating from Spanish to Catalan and vice versa. The AT 3(c) generalizes the rule to apply in order to propagate the gender from the noun (a dominant category) to the article and the adjective. The AT 3(a) generalizes, on the one hand, the use of the auxiliary Catalan verb *anar* to express the past (preterite) tense and, on the other hand, the preposition change when it refers to a place name, such as the name of a city or a country. The AT 3(b) also generalizes the use of the auxiliary Catalan verb *anar*, but it does not specify any preposition change because the noun does not refer to a location name.

Finally it must be noticed that those ATs that have a different number of *open* lexical categories on both sides (source and target) of the AT cannot be

applied. This is because it makes no sense to delete or introduce an open word (i. e. a noun or an adjective) when translating from SL into TL.

3.3 Application of the alignment templates

When translating an input SL text into the TL, the ATs are applied from left to right. Once an AT has been applied, the search for a new AT to apply starts from the next SL word that was not covered by the previous applied AT. If there is not any AT to apply the word from which the search was started is translated in isolation (by looking it up in the bilingual dictionary) and a new search is started from the next SL word.

In our approach we have used two different criteria to select the AT to apply. The first criterion uses the number of times each AT has been seen in the training corpus to select the most frequent AT that matches a SL text segment; the second one chooses the longest AT.⁴ In both cases, the transfer module is provided with a frequency threshold that is used to discard infrequent ATs according to their counts.

Matching of alignment templates. To apply an AT its SL part must match exactly the SL text segment being translated, and it must be *applicable*. An AT is *applicable* if the TL inflection information provided by the bilingual dictionary for the dominant words being translated (see section 3.2) is not modified by the TL part of the AT. The last must hold because it usually makes no sense to change the TL gender or number provided by the bilingual dictionary for a dominant category (a noun for example).

Application of alignment templates. The application of an AT is done by translating each open word by looking it up in a bilingual dictionary, and replacing the part-of-speech information provided by the bilingual dictionary by the part-of-speech information provided by the TL part of the AT. The alignment information is used to put each word in their correct place in the TL. Closed words are not translated, they are taken from the TL part of the AT; in this way the method can perform transformations such as preposition or verb tense changes when translating.

The next example illustrates how an AT is applied. Suppose that we are translating to Catalan the Spanish text segment *permanecieron en Alemania*⁵ with the following source language intermediate representation (SLIR): *permanecer-(verb,pret,3rd,pl) en-(pr) Alemania-(noun,loc)*. The AT shown in figure 3(a) matches the given Spanish text segment and is applicable. To apply this AT, first all open words are translated into TL (Catalan) by looking them up in a bilingual dictionary: *permanecer-(verb,pret,3rd,pl)* is translated as *romandre-(verb,pret,3rd,pl)* and *Alemania-(noun,loc)* is translated as *Alemanya-(noun,loc)*. After that, the output of the transfer module

⁴ Note that in this case the ATs can be applied in a left-to-right longest-match (LRLM) way, as in OpenTrad Apertium (<http://apertium.sourceforge.net> [12,13]).

⁵ Translated into English as *They remained in Germany*.

is constructed taking into account the inflection information provided by the TL part of the AT for the open words and copying closed words to the output as they appear in the TL part of the AT. The alignment information is used to put each word in the correct place in the TL. For the running example we have, after applying the AT, the following target language intermediate representation (TLIR): *anar*-(**vbaux,pres,3rd,pl**) *romandre*-(**verb,inf**) *a*-(**pr**) *Alemanya*-(**noun,loc**), which the generation module transforms into the Catalan phrase *van romandre a Alemanya*.

4 Experiments

In this section we overview the shallow-transfer MT system used to test the approach presented in this paper, then we describe the experiments conducted and the results achieved. The performance of the presented approach is compared to that of a word-for-word MT system and that of a MT system using hand-coded transfer rules.

4.1 Shallow-transfer machine translation engine

For the experiments we used the Spanish–Catalan MT system interNOSTRUM [14],⁶ which basically follows a (shallow) transfer architecture with the following pipelined modules:

- A *morphological analyzer* that divides the text in surface forms and delivers, for each surface form, one or more lexical forms consisting of *lemma*, *lexical category* and morphological inflection information.
- A *part-of-speech tagger* (categorial disambiguator) that chooses, using a first-order hidden Markov model, one of the lexical forms corresponding to each ambiguous surface form.
- A *lexical transfer* module that reads each SL lexical form and delivers the corresponding TL lexical form.
- A *structural transfer* module that (parallel to the lexical transfer) detects patterns of lexical forms, like “article–noun–adjective”, which need to be processed for word reordering, agreement, etc. This is the module we are trying to automatically learn from bilingual corpora.
- A *morphological generator* delivers a TL surface form for each TL lexical form, by suitably inflecting it; in addition, it performs some inter-word orthographical operations such as contractions and apostrophations.

⁶ A complete rewriting of this MT engine [12,13] (together with data for several language pairs) was released in July 2005 under an open source license (<http://apertium.sourceforge.net>).

Language	Sentences	Running words	Vocabulary size
Spanish (training)	400 000	7 480 909	157 841
Catalan (training)	400 000	7 285 133	155 446
Spanish (AT extraction)	15 000	288 084	31 409
Catalan (AT extraction)	15 000	296 409	30 228
Spanish (test)	1 498	32 092	7 473
Catalan (test)	1 498	31 468	7 088

Table 1. Data about the training corpus used to compute the word alignments, the part of the corpus used to extract the alignment templates, and the disjoint corpora used for evaluation (test).

4.2 Results

We have done the experiments using the MT system presented above when translating in both directions, from Spanish to Catalan and from Catalan to Spanish. To train the word alignments and to extract the alignment templates (ATs) we have used a Spanish–Catalan parallel corpus from *El Periódico de Catalunya*, a daily newspaper published both in Catalan and Spanish.

Before training the word alignments, the parallel corpus was preprocessed so as to have it in the intermediate representations used by the shallow-transfer MT system. The preprocessing consisted on analyzing both sides of the parallel corpus by means of the morphological analysers and part-of-speech taggers used by the MT system when translating.

Once the parallel corpus was preprocessed, the word alignments were trained using the open-source GIZA++ toolkit.⁷ The training of the word alignments consisted in training the IBM model 1 [15] for 4 iterations, the hidden Markov model (HMM) alignment model [16] for 5 iterations, and the IBM model 4 [15] for 8 iterations. After training the word alignment a symmetrization that applies an heuristic postprocessing is performed to combine the alignments on both translation directions; in this way, a source word is allowed to be aligned with more than one target word. For a deeper description of the symmetrization method see [17].

After training the word alignments the alignment templates (ATs) were extracted using a small part of the training corpus because this is a very resource-consuming task. Therefore, it must be said that the experiments did not exploit the full strength of the statistical approach; much better results must be expected for the alignment templates approach when the full training corpus is used to extract the ATs.

Once the ATs are extracted, they are filtered according to the guidelines explained in section 3.2 to discard those ATs that cannot be applied. Table 1 summarizes basic data about the training corpus, the part of the training corpus used to extract the ATs, and the disjoint corpora used for evaluation.

⁷ <http://www.fjoch.com/GIZA++.html>

Translation direction	MT setup	WER	PER	NIST	BLEU
Spanish→Catalan	word-for-word	29.41	26.99	10.07	53.07
	longest AT	24.63	22.86	10.75	59.41
	most frequent AT	24.50	22.70	10.77	59.75
	hand-coded rules	22.94	21.05	10.88	62.50
Catalan→Spanish	word-for-word	30.01	27.46	9.76	52.59
	longest AT	25.32	23.25	10.51	57.69
	most frequent AT	25.90	23.78	10.44	56.66
	hand-coded rules	23.77	22.19	10.53	60.23

Table 2. Results for the two translation directions and the different MT setups used in the experiment. The error measures reported are, from left to right, word error rate, position independent error rate, the NIST score, and the BLEU score. The results reported are for word-for-word translation (baseline), hand-coded transfer rules, and the two different approaches tested to choose the automatically-extracted ATs to apply.

In the experiments we have tested two different criteria to select the AT to apply when processing the SL text from left to right (see section 3.3). Remember that the first criterion chooses the most frequent AT that can be applied, and in case of equal frequency the AT that covers the longest SL word sequence, i. e. the longest AT. The second criterion chooses the AT that covers the longest SL word sequence, and in case of equal length, the most frequent AT.

Table 2 shows the results achieved when using the ATs automatically extracted from bilingual corpora. For comparison purposes the results of a word-for-word translation (that is, when no structural transformations are applied and all words are translated in isolation by looking them up in a bilingual dictionary), and the results achieved when using hand-coded transfer rules are reported; in both cases the same MT system was used. The errors reported in table 2 were calculated on a test corpus extracted from the newspaper *El Periódico de Catalunya* with only one reference translation (see table 1).

As can be seen in table 2 the room for improvement between word-for-word and hand-coded rules is about 9.4 BLEU points for the Spanish→Catalan translation, and about 7.6 BLEU points for the Catalan→Spanish translation. As can be seen the improvement in the translation quality is around 6 BLEU points in the Spanish→Catalan translation, and about 4.5 BLEU points in the Catalan→Spanish. Moreover, both selecting criteria give comparable results, but slightly better (around 1 BLEU point) when the translation is from Catalan into Spanish and the longest AT is selected for application.

5 Discussion

In this paper the introduction of statistically-inferred alignment templates (ATs) as transfer rules within a shallow-transfer MT system has been tested. To this end, some linguistic information has been used in order to learn the transformations to apply when translating SL into TL. In any case, the linguistic information used can be easily provided to the alignment templates extraction

algorithm, and is a commonly used information in indirect rule-based transfer MT systems, which rely on monolingual and bilingual dictionaries.

The approach presented has been tested using an existing shallow-transfer MT system. The performance of the system when using the automatically extracted ATs has been compared to that of word-for-word translation (when no structural transformations are applied) and that of hand-coded rules application using the same MT engine. In both translation directions there is a significant improvement in the translation qualities compared to word-for-word translation. Furthermore, the translation quality is very close to that achieved when using hand-coded transfer rules. Moreover, it must be noticed that the relative improvement in both translation directions, if the best translation quality that can be achieved is assumed to be that of hand-coded rules, is about 70% for the Spanish→Catalan translation, and around 60% for the Catalan→Spanish translation.

Two different selection criteria has been tested to choose the AT to apply (longest or most frequent) when more than one can be applied to the same SL text segment. The performance for both selecting criteria is more or less the same when translating Spanish into Catalan. However, when translating Catalan into Spanish, choosing the longest AT gives better results (around 1 BLEU point) than choosing the most frequent AT. As future work we plan to study the reason why choosing the longest AT gives better results and why the improvement in the translation quality is lower in the Catalan→Spanish translation.

Finally we plan to merge both selecting criteria into a single one by means of a log-linear combination, despite the fact that due to the comparable translation results for both criteria we do not expect a great improvement.

Acknowledgements

Work funded by the Spanish Comisión Interministerial de Ciencia y Tecnología through project TIC2003-08681-C02-01 and by the Spanish Ministerio de Educación y Ciencia and the European Social Fund through grant BES-2004-4711. We thank Mikel L. Forcada (Universitat d'Alacant) and Richard Zens (RWTH Aachen University) for useful comments on this work.

References

1. Och, F.J.: Statistical Machine Translation: From Single-Word Models to Alignment Templates. PhD thesis, RWTH Aachen University, Aachen, Germany (2002)
2. Och, F.J., Ney, H.: The alignment template approach to statistical machine translation. *Computational Linguistics* **30**(4) (2004) 417–449
3. Bender, O., Zens, R., Matusov, E., Ney, H.: Alignment templates: the RWTH SMT system. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Kyoto, Japan (2004) 79–84
4. Probst, K., Levin, L., Peterson, E., Lavie, A., Carbonell, J.: MT for minority languages using elicitation-based learning of syntactic transfer rules. *Machine Translation* **17**(4) (2002) 245–270

5. Lavie, A., Probst, K., Peterson, E., Vogel, S., Levin, L., Font-Llitjós, A., Carbonell, J.: A trainable transfer-based machine translation approach for languages with limited resources. In: Proceedings of Workshop of the European Association for Machine Translation (EAMT-2004), Valletta, Malta (2004)
6. Kaji, H., Kida, Y., Morimoto, Y.: Learning translation templates from bilingual text. In: Proceedings of the 14th Conference on Computational Linguistics, Morristown, NJ, USA, Association for Computational Linguistics (1992) 672–678
7. Brown, R.D.: Adding linguistic knowledge to a lexical example-based translation system. In: Proceedings of the Eighth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99), Chester, England (1999) 22–32
8. Cicekli, I., Güvenir, H.A.: Learning translation templates from bilingual translation examples. *Applied Intelligence* **15**(1) (2001) 57–76
9. Liu, Y., Zong, C.: The technical analysis on translation templates. In: Proceedings of the IEEE International Conference on Systems, Man & Cybernetics (SMC), The Hague, Netherlands, IEEE (2004) 4799–4803
10. Och, F.J., Ney, H.: Discriminative training and maximum entropy models for statistical machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, PA (2002) 295–302
11. Och, F.J.: An efficient method for determining bilingual word classes. In: EACL'99: Ninth Conference of the European Chapter of the Association for Computational Linguistics, Bergen, Norway (1999) 71–76
12. Corbí-Bellot, A.M., Forcada, M.L., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Ramírez-Sánchez, G., Sánchez-Martínez, F., Alegria, I., Mayor, A., Sarasola, K.: An open-source shallow-transfer machine translation engine for the Romance languages of Spain. In: Proceedings of the 10th European Association for Machine Translation Conference, Budapest, Hungary (2005) 79–86
13. Armentano-Oller, C., Carrasco, R.C., Corbí-Bellot, A.M., Forcada, M.L., Ginestí-Rosell, M., Ortiz-Rojas, S., Pérez-Ortiz, J.A., Ramírez-Sánchez, G., Sánchez-Martínez, F., Scalco, M.A.: Open-source Portuguese-Spanish machine translation. In: Computational Processing of the Portuguese Language, Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese, PROPOR 2006. Volume 3960 of Lecture Notes in Computer Science. Springer-Verlag (2006) 50–59
14. Canals-Marote, R., Esteve-Guillén, A., Garrido-Alenda, A., Guardiola-Savall, M., Iturraspe-Bellver, A., Montserrat-Buendia, S., Ortiz-Rojas, S., Pastor-Pina, H., Pérez-Antón, P., Forcada, M.: The Spanish-Catalan machine translation system interNOSTRUM. In: Proceedings of MT Summit VIII: Machine Translation in the Information Age, Santiago de Compostela, Spain (2001) 73–76
15. Brown, P.F., Pietra, S.A.D., Pietra, V.J.D., Mercer, R.L.: The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics* **19**(2) (1993) 263–311
16. Vogel, S., Ney, H., Tillmann, C.: HMM-based word alignment in statistical translation. In: COLING '96: The 16th International Conference on Computational Linguistics, Copenhagen (1996) 836–841
17. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. *Computational Linguistics* **29**(1) (2003) 19–51