The UA-Prompsit hybrid machine translation system for the 2014 Workshop on Statistical Machine Translation

Víctor M. Sánchez-Cartagena,[*][‡] **Juan Antonio Pérez-Ortiz,**[*] **Felipe Sánchez-Martínez**[*]

[*]Dep. de Llenguatges i Sistemes Informàtics,
Universitat d'Alacant, E-03071, Alacant, Spain

[‡]Prompsit Language Engineering,
Av. Universitat, s/n. Edifici Quorum III, E-03202, Elx, Spain

{vmsanchez, japerez, fsanchez}@dlsi.ua.es

Abstract

This paper describes the system jointly developed by members of the Departament de Llenguatges i Sistemes Informàtics at Universitat d'Alacant and the Prompsit Language Engineering company for the shared translation task of the 2014 Workshop on Statistical Machine Translation. We present a phrase-based statistical machine translation system whose phrase table is enriched with information obtained from dictionaries and shallow-transfer rules like those used in rule-based machine translation. The novelty of our approach lies in the fact that the transfer rules used were not written by humans, but automatically inferred from a parallel corpus.

1 Introduction

This paper describes the system jointly submitted by the Departament de Llenguatges i Sistemes Informàtics at Universitat d'Alacant and the Prompsit Language Engineering company to the shared translation task of the ACL 2014 Ninth Workshop on Statistical Machine Translation (WMT 2014). We participated in the English–French translation task with a hybrid system that combines, in a phrase-based statistical machine translation (PB-SMT) system, bilingual phrases obtained from parallel corpora in the usual way (Koehn, 2010, ch. 5), and also bilingual phrases obtained from the existing dictionaries in the Apertium rule-based machine translation (RBMT) platform (Forcada et al., 2011) and a number of shallow-transfer machine translation rules automatically inferred from a small subset of the training corpus.

Among the different approaches for adding linguistic information to SMT systems (Costa-Jussà and Farrús, 2014), we followed the path we started with our submission to the Spanish–English WMT 2011 shared translation task (Sánchez-Cartagena et al., 2011b) which consisted of enriching the phrase table of a PBSMT system with phrase pairs generated using the dictionaries and rules in the Apertium (Forcada et al., 2011) Spanish–English RBMT system; our approach was one of the winners[1] (together with two online SMT systems that were not submitted for the task but were included in the evaluation by the organisers and a system by Systran) in the pairwise manual evaluation of the English–Spanish translation task (Callison-Burch et al., 2011). In this submission, however, we only borrow the dictionaries from the Apertium English–French RBMT system and use them to automatically infer the rules from a parallel corpus. We therefore avoid the need for human-written rules, which are usually written by trained experts, and explore a novel way to add morphological information to PBSMT. The rules inferred from corpora and used to enlarge the phrase table are shallow-transfer rules that build their output with the help of the bilingual dictionary and work on flat intermediate representations (see section 3.1); no syntactic parsing is consequently required.

The rest of the paper is organised as follows. The following section outlines related hybrid approaches. Section 3 formally defines the RBMT paradigm and summarises the method followed to automatically infer the shallow-transfer rules, whereas the enrichment of the phrase table is described in section 4. Sections 5 and 6 describe, respectively, the resources we used to build our submission and the results achieved for the English–French language pair. The paper ends with some concluding remarks.

2 Related work

Linguistic data from RBMT systems have already been used to enrich SMT systems (Tyers, 2009; Schwenk et al., 2009; Eisele et al., 2008; Sánchez-Cartagena et al., 2011a). We have already proved

---

[1]No other system was found statistically significantly better using the sign test at $p \leq 0.10$.

that using hand-written rules and dictionaries from RBMT yields better results than using only dictionaries (Sánchez-Cartagena et al., 2011a).

However, in the approach we present in this paper, rules are automatically inferred from a parallel corpus after converting it into the intermediate representation used by the Apertium RBMT platform (see section 3.3). It can be therefore seen as a novel method to add morphological information to SMT, as factored translation models do (Koehn and Hoang, 2007; Graham and van Genabith, 2010). Unlike factored models, we do not estimate independent statistical models for the translation of the different factors (lemmas, lexical categories, morphological inflection attributes, etc.) and for the generation of the final surface forms. Instead, we first infer a set of rules that deal with the grammatical divergences between the languages involved by performing operations such as reorderings, gender and number agreements, etc. Afterwards, we add synthetic phrase pairs generated from these rules and the Apertium dictionaries to the data from which the well-known, classical PBSMT models (Koehn, 2010) are estimated. The rules in our approach operate on the source-language (SL) morphological attributes of the input words and on the target-language (TL) morphological attributes of their translation according to a bilingual dictionary. In addition, they do no contain probabilities or scores, thus they increase the predictability of the output and can be easily corrected by humans. This fact also represents a significant difference with the probabilistic rules used by certain approaches that aim at improving the grammaticality of the SMT output (Riezler and Maxwell III, 2006; Bojar and Hajič, 2008).

With respect to the rule inference approach, other approaches such as those by Sánchez-Martínez and Forcada (2009) and Caseli et al. (2006) can be found in literature; however, our approach is the first strategy for shallow-transfer rule inference which generalises to unseen combinations of morphological inflection attributes in the training corpus (Sánchez-Cartagena et al., 2014).

3 Inferring shallow-transfer rules from parallel corpora

3.1 Shallow-transfer rule-based machine translation

The RBMT process can be split into three different steps (Hutchins and Somers, 1992): (i) analysis of the SL text to build an SL intermediate represen-

tation; (ii) transfer from that SL intermediate representation into a TL intermediate representation; and (iii) generation of the final translation from the TL intermediate representation.

Shallow-transfer RBMT systems use relatively simple intermediate representations, which are based on lexical forms consisting of lemma, part of speech and morphological inflection information of the words, and apply simple shallow-transfer rules that operate on sequences of lexical forms: this kind of systems do not perform full parsing. For instance, for translating the English sentence *I like Pierre's house* into French with the Apertium shallow-transfer RBMT platform we have used to build our submission, the following steps are carried out. First, the sentence is analysed as the following sequence of lexical forms:

I `PRN-p:1.num:sg`
like `VB-t:pres.p:ε:num:ε`
Pierre `PN`
's `POS`
house `N-gen:ε.num:sg`

This sequence is made up of a personal pronoun (`PRN`) in first person (`p:1`) singular (`num:sg`) with lemma *I*, the verb (`VB`) *like* in present tense (`t:pres`), a proper noun (`PN`) with lemma *Pierre*, the possessive ending (`POS`), and a noun (`N`) in singular with lemma *house*. Some morphological inflection attributes have an empty value ϵ because they do not apply to the corresponding language.

Then, structural transfer rules are applied to obtain the TL intermediate representation with the help of the bilingual dictionary, which provides the individual translation of each SL lexical form (including its morphological information). In this case, two rules are applied: the first one makes the verb to agree with the personal pronoun, while the second one translates the English possessive construction into French. The resulting sequence of TL lexical forms is:

Je `PRN-p:1.num:sg`
aime `VB-t:pres.p:1:num:sg`
le `DT-gen:f.num:sg`
maison `N-gen:f.num:sg`
de `PR`
Pierre `PN`

Note that a preposition (`PR`) with lemma *de* and a determiner (`DT`) with lemma *le* and the same gender and number as the common noun have been added by the rule. Finally, the translation into TL is generated from the TL lexical forms: *J'aime la maison de Pierre*.
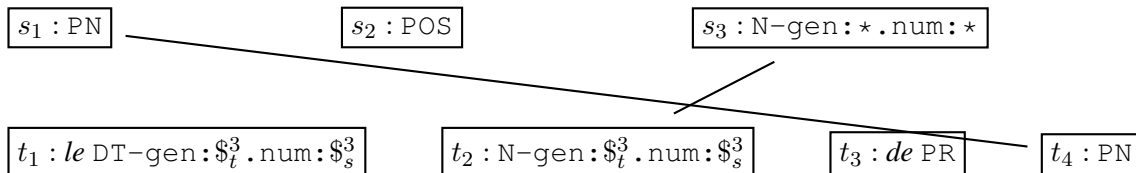
$$s_1:\texttt{PN} \qquad s_2:\texttt{POS} \qquad s_3:\texttt{N-gen:*.num:*}$$

$$t_1:\textit{le}\ \texttt{DT-gen:}\$_t^3\texttt{.num:}\$_s^3 \qquad t_2:\texttt{N-gen:}\$_t^3\texttt{.num:}\$_s^3 \qquad t_3:\textit{de}\ \texttt{PR} \qquad t_4:\texttt{PN}$$

Figure 1: Shallow-transfer rule for the translation of the English Saxon genitive construction into French.

3.2 A rule formalism suitable for rule inference

Figure 1 shows the second rule applied in the example from the previous section encoded with the formalism we have defined for rule inference (Sánchez-Cartagena et al., 2014). Each rule contains a sequence of SL word classes (depicted as the sequence of boxes at the top of the figure) and TL word classes (the sequence of boxes below them). The sequence of SL word classes defines the set of sequences of lexical forms which will match the rule. Each SL word class s_i defines the conditions that must be met by the i-th lexical form matching the rule and contains an optional lemma (no lemma means that any SL lemma is allowed), a lexical category and a set of morphological inflection attributes and their expected values. A wildcard (asterisk) as the value of a morphological inflection attribute means that it matches any possible value. Thus, the rule from the example matches any proper noun followed by a possessive ending and a noun, regardless of its gender and number.

As regards the TL word classes, they contain the same elements as the SL word classes and define the output of the rule. An empty lemma in a TL word class means that it is obtained by looking up in the bilingual dictionary the SL lexical form matching the aligned SL word class (alignments are represented as lines connecting SL and TL word classes). The reference value $\$_s^i$ means that the value of a morphological inflection attribute is copied from the SL lexical form matching the i-th SL word class, while the reference value $\$_t^i$ means that the value is taken from the TL lexical form obtained after looking up in the bilingual dictionary the aforementioned SL lexical form. The rule depicted in Figure 1 generates a sequence of four TL lexical forms. The first one is a determiner whose lemma is *le*, its gender is obtained from the gender of the TL lexical form resulting after looking up in the bilingual dictionary the third matching SL lexical form ($\$_t^3$), that is, the common noun, while its

number is directly obtained from the same SL lexical form before dictionary look-up ($\$_s^3$). Although they have not been used in this example, explicit values can be used in the morphological inflection attributes of the SL and TL word classes, thus restricting the SL lexical forms to which the rule can be applied to those having the values in the corresponding SL word classes,[2] and explicitly stating the value that the TL lexical forms produced by the rule will have, respectively.

3.3 Rule inference algorithm

The set of rules that will be used to generate the phrase pairs that will be integrated into the PB-SMT system's phrase table, encoded with the formalism presented in the previous section, are obtained from the parallel corpus by applying the steps described in this section. They are a subset of the steps followed by Sánchez-Cartagena et al. (2014) to infer shallow-transfer rules to be used in Apertium from small parallel corpora.

First, both sides of the parallel corpus are morphologically analysed and converted into the intermediate representations used by Apertium. Word alignments are then obtained by symmetrising (using the refined intersection method proposed by Och and Ney (2003)) the set of alignments provided by GIZA++ (Och and Ney, 2003) when it is run on both translations directions. Afterwards, the bilingual phrase pairs compatible with the alignments are extracted as it is usually done in SMT (Koehn, 2010, Sec. 5.2.3), and those that are not compatible with the bilingual dictionary of the Apertium English–French RBMT system[3] or

---

[2]In addition to that criterion, our formalism also permits restricting the application of a rule to the SL lexical forms that, after being looked up in the bilingual dictionary, the TL lexical forms obtained from them have specific morphological inflection attribute values (Sánchez-Cartagena et al., 2014) although no restrictions of this type are imposed in the rule depicted in Figure 1.

[3]If the words that belong to open lexical categories (those that carry the meaning of the sentence: nouns, verbs, adjectives, etc.) are aligned with other words that do not match the translation present in the bilingual dictionary, the rule in-

contain punctuation marks or unknown words are discarded. Finally, from each bilingual phrase pair, all the possible rules which correctly reproduce it —when the rule is applied to the SL side of the phrase pair, its TL side is obtained— are generated as follows. First, a very specific rule, which matches only the SL phrase in the bilingual phrase pair is generated; more general rules are then created by modifying this initial rule. The modifications to the initial rule consist of removing lemmas from the SL and TL word classes, introducing wildcard values in the morphological inflection attributes of the SL word classes and adding reference values in the morphological inflection attributes of the TL word classes. The result of this process is a huge set of rules with different levels of generalisation. Obviously, not all the rules in this set will be used: the best ones are automatically selected by considering all the rules obtained from the different bilingual phrase pairs extracted from the corpus and finding the minimum set of rules that meets the following two conditions:

1. Each bilingual phrase pair is correctly reproduced by at least one rule.

2. If a rule matches the SL side of bilingual phrase pair but does not correctly reproduce its TL side, there is another rule that is more specific (i.e. less general) than it, and correctly reproduces its TL side.

This minimisation problem is formulated as an integer linear programming[4] problem (Garfinkel and Nemhauser, 1972) and solved using the *branch and cut* algorithm (Xu et al., 2009).

From the small subset of the huge initial rules obtained by solving the minimisation problem, the rules whose effect can be achieved by combining shorter rules or by translating all or some of the words in isolation (i.e. word for word) are removed. In this way, the number of rules is further reduced and long rules, which are more prone to overgeneralisation because they are inferred from fewer bilingual phrase pairs, are discarded.[5]

4 Enhancing phrase-based SMT with shallow-transfer linguistic resources

The set of shallow-transfer rules inferred from the parallel corpus are integrated in the PBSMT system, together with the RBMT dictionaries, using the same method we used for our WMT 2011 shared translation task subsmission (Sánchez-Cartagena et al., 2011b). However, it is important to stress that, until now, this strategy had only been tested when the rules to be integrated were hand-written and not automatically obtained from corpora.

Our strategy involves adding to the phrase table of the PBSMT system all the bilingual phrase pairs which either match a shallow-transfer rule or an entry in the bilingual dictionary. Generating the set of bilingual phrase pairs which match bilingual dictionary entries is straightforward. First, all the SL surface forms that are recognised by Apertium and their corresponding lexical forms are generated. Then, these SL lexical forms are translated using the bilingual dictionary, and finally their TL surface forms are generated.

Bilingual phrase pairs which match structural transfer rules are generated in a similar way. First, the SL sentences to be translated are analysed with Apertium to get their SL lexical forms, and then the sequences of lexical forms that match a structural transfer rule are translated with that rule and passed through the rest of the Apertium pipeline in order to get their translations. If a sequence of SL lexical forms is matched by more than one structural transfer rule, it will be used to generate as many bilingual phrase pairs as different rules it matches. This differs from the way in which Apertium translates, as it only applies the longest rule. Note also that the test set is used to guide the phrase extraction in order to avoid the generation of an unmanageable set of phrase pairs.

We add these bilingual phrase pairs directly to the phrase table, rather than adding them to the training corpus and relying on the phrase extraction algorithm (Koehn, 2010, sec. 5.2.3), in order to avoid splitting the multi-word expressions provided by Apertium into smaller phrases (Schwenk et al., 2009, sec. 2). The bilingual phrase pairs are added only once to the list of corpus-extracted phrase pairs, and then the phrase translation probabilities are computed by relative frequency as usual (Koehn, 2010, sec. 5.2.5). A boolean feature

---

ference algorithm is likely to infer many very specific rules that try to correct that lexical mismatch. Since the aim of our approach is learning general rules that deal with the grammatical divergences between languages, the bilingual phrases that contain the aforementioned alignments are discarded. Words from closed lexical categories, that usually suffer deeper changes when the sentence is translated to a different language, are not subject to this restriction.

[4]An integer linear programming problem involves the optimisation (maximisation or minimisation) of a linear objective function subject to linear inequality constraints.

[5]Although longer rules contain more context information,

---

for our rule inferring algorithm there are fewer bilingual phrases from which to infer them, and consequently fewer evidence from which to extract the right reference attributes.

function to flag bilingual phrase pairs obtained from the RBMT resources is added to the phrase table in order to conveniently weight the synthetic RBMT phrase pairs.

5 System training

We built a baseline PBSMT Moses (Koehn et al., 2007) system[6] from a subset of the parallel corpora distributed as part of the WMT 2014 shared translation task, namely Europarl (Koehn, 2005), News Commentary and Common Crawl, and a subset of the French monolingual corpora, namely Common Crawl, Europarl, News Commentary and News Crawl. The language model was built with the KenLM language modelling toolkit (Heafield et al., 2013), which was used to train a 5-gram language model using interpolated Kneser-Ney discounting (Goodman and Chen, 1998). Word alignments were computed by means of GIZA++ (Och and Ney, 2003). The weights of the different feature functions were optimised by means of minimum error rate training (Och, 2003) on the 2013 WMT test set.[7]

The phrase table of this baseline system was then enriched with phrase pairs generated from rules automatically inferred from the concatenation of the test corpora distributed for the WMT 2008–2012 shared translation tasks, and from the English–French bilingual dictionary in the Apertium platform.[8] Since the minimisation problem which needs to be solved in order to obtain the rules is very time-consuming, we chose a small rule inference corpus similar to this year's test set. The bilingual dictionary, which contains mappings between SL and TL lemmas, consists of 13 088 entries and is quite small compared to the Spanish–English bilingual dictionary we used in our submission to WMT 2011 (Sánchez-Cartagena et al., 2011b), which consisted of 326 228 bilingual entries. This is because the English–French Apertium linguistic resources were automatically built by crossing data from other existing language pairs.

Table 1 summarises the data about the corpora used to build our submission, both for the PBSMT baseline system and for the rules used to enrich its phrase table.

The corpus used to automatically infer the rules

Task	Corpus	Sentences
Translation model	Europarl	2 007 723
	News Commentary	183 251
	Common Crawl	3 244 152
	Total	5 435 126
	Total clean	4 196 987
Language model	Common Crawl	3 244 152
	Europarl	2 190 579
	News Commentary	227 013
	News Crawl	30 451 749
	Total	36 113 493
Rule inference	newstest 2008–2012	13 071
Tuning	newstest2013	3 000
Test	newstest2014	3 003

Table 1: Size of the corpora used in the experiments. The bilingual training corpora was cleaned up to remove empty parallel sentences and those containing more than 40 tokens.

was split into two parts: the larger one (4/5 of the corpus) was used for actual rule inference as described in section 3.3; the remaining corpus was used as a development corpus as explained next. For each rule z, first the proportion $r(z)$ of bilingual phrase pairs correctly reproduced by the rule divided by the number of bilingual phrases it matches is computed. Rules whose proportion $r(z)$ is lower than a threshold value δ are then discarded before solving the minimisation problem. The value of δ is chosen so that it maximises, on the development corpus, the BLEU score (Papineni et al., 2002) obtained by an Apertium-based system which uses the inferred rules; in our submission $\delta = 0.15$. In addition, rules that do not correctly reproduce at least 100 bilingual phrase pairs were also discarded in order to make the minimisation problem computationally feasible.

6 Results and discussion

Table 2 reports the translation performance as measured by BLEU (Papineni et al., 2002), TER (Snover et al., 2006) and METEOR (Banerjee and Lavie, 2005) achieved by the baseline PBSMT, our submission (*UA-Prompsit*), Apertium when it uses the set of inferred rules, and Apertium when it uses no rules at all (word-for-word translation). The size of the phrase table and the amount of unknown words in the test set are also reported when applicable.

According to the three evaluation metrics, the translation performance of our submission is very close to that of the PBSMT baseline (slightly better according to BLEU and TER, and slightly worse according to METEOR). The difference between both systems computed by paired bootstrap

---

[6]No factored models were used.

[7]The corpora can be downloaded from `http://www.statmt.org/wmt14/translation-task.html`.

[8]`https://svn.code.sf.net/p/apertium/svn/incubator/apertium-en-fr`

system	BLEU	TER	METEOR	# of unknown words	phrase table size
baseline	0.3232	0.5807	0.5441	870	100 530 734
UA-Prompsit	0.3258	0.5781	0.5432	861	100 585 182
Apertium-rules	0.0995	0.7767	0.3168	4 743	-
Apertium-word-for-word	0.0631	0.8368	0.2617	4 743	-

Table 2: Case-insensitive BLEU, TER, and METEOR scores obtained, on the *newstest2014* test set, by the baseline PBSMT system (*baseline*), the hybrid system submitted to the WMT 2014 shared translation task (*UA-Prompsit*), Apertium when it uses the set of inferred rules (*Apertium-rules*), and Apertium when it uses no rules at all (*Apertium-word-for-word*). The number of unknown words and the size of the phrase table are also reported when applicable.

resampling (Koehn, 2004) is not statistically significant for any of the three evaluation metrics (1 000 iterations, $p = 0.05$).

An inspection of the 86 rules inferred shows that they encode some of the transformations that one would expect from a set of English–French rules, such as gender and number agreements between nouns, determiners and adjectives, preposition changes, and the introduction of the auxiliary verb *avoir* for the past tense. In addition, the improvement over word-for-word translation achieved when they are used by Apertium is statistically significant for the three evaluation metrics.

One of the reasons for not improving the baseline PBMT system might be the small coverage of the Apertium dictionaries. As already mentioned in the previous section, the English–French bilingual dictionary has a low number of entries compared to more mature language pairs in Apertium which have around 20 times more bilingual entries. Table 1 shows some effects of such a small dictionary: the number of unknown words for the Apertium-based system is really high, and with regards to *UA-Prompsit*, its coverage barely increases when compared to the PBSMT baseline. We plan to test the approach presented in this paper with language pairs for which more mature dictionaries are available in the Apertium project.

In addition to this, due to the tight schedule, we had to remove the rules not reproducing at least 100 bilingual phrase pairs in order to solve the minimisation problem in a short amount of time. This has clearly reduced the amount of rules inferred and prevented some useful information present in the parallel corpus from being incorporated in the form of rules. For instance, no rule matching a sequence longer than 3 lexical forms has been extracted (long bilingual phrases are less frequent than short ones). Future research directions for alleviating this problem include setting the minimum number of reproduced bilingual phrases independently for each sequence of SL lexical cate-

gories (Sánchez-Cartagena et al., 2014).

7 Concluding remarks

We have presented the MT system submitted jointly by the Departament de Llenguatges i Sistemes Informàtics at Universitat d'Alacant and Prompsit Language Engineering to the WMT 2014 shared translation task. We developed a hybrid system for the English–French language pair which enriches the phrase table of a standard PBSMT system with phrase pairs generated from the Apertium RBMT dictionaries and a set of shallow-transfer rules automatically inferred from a parallel corpus, also with the help of the dictionaries. This submission aims at solving one strong limitation of a previous submission of our team (Sánchez-Cartagena et al., 2011b): the need for a hand-crafted set of shallow-transfer rules, which can only be written by people with a deep knowledge of the languages involved. Our approach outperforms a standard PBSMT system built from the same data by a small, non statistically significant margin, according to two of the three evaluation metrics used. The low coverage of the dictionaries used and the aggressive pruning carried out when solving the minimisation problem needed to infer the rules are probably the reasons behind such a small improvement over the baseline.

Acknowledgements

References

S. Banerjee and A. Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72.

O. Bojar and J. Hajič. 2008. Phrase-based and deep syntactic English-to-Czech statistical machine translation. In *Proceedings of the third Workshop on Statistical Machine translation*, pages 143–146. Association for Computational Linguistics.

C. Callison-Burch, P. Koehn, C. Monz, and O. Zaidan. 2011. Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 22–64, Edinburgh, Scotland, July. Association for Computational Linguistics.

H. M. Caseli, M. G. V. Nunes, and M. L. Forcada. 2006. Automatic induction of bilingual resources from aligned parallel corpora: application to shallow-transfer machine translation. *Machine Translation*, 20(4):227–245. Published in 2008.

M. R. Costa-Jussà and M. Farrús. 2014. Statistical machine translation enhancements through linguistic levels: A survey. *ACM Comput. Surv.*, 46(3).

A. Eisele, C. Federmann, H. Saint-Amand, M. Jellinghaus, T. Herrmann, and Y. Chen. 2008. Using Moses to integrate multiple rule-based machine translation engines into a hybrid system. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 179–182, Columbus, Ohio.

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

R. S. Garfinkel and G. L. Nemhauser. 1972. *Integer programming*, volume 4. Wiley New York.

J. Goodman and S. F. Chen. 1998. An empirical study of smoothing techniques for language modeling. Technical Report TR-10-98, Harvard University, August.

Y. Graham and J. van Genabith. 2010. Factor templates for factored machine translation models. In *IWSLT 2010 : 7th International Workshop on Spoken Language Translation*, pages 275–283.

K. Heafield, I. Pouzyrevsky, J. H. Clark, and P. Koehn. 2013. Scalable modified Kneser-Ney language model estimation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 690–696, Sofia, Bulgaria, August.

W. J. Hutchins and H. L. Somers. 1992. *An introduction to machine translation*, volume 362. Academic Press New York.

P. Koehn and Hieu Hoang. 2007. Factored translation models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 868–876, Prague.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*.

P. Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 4, pages 388–395.

P. Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the Tenth Machine Translation Summit*, pages 12–16, Phuket, Thailand, September.

P. Koehn. 2010. *Statistical Machine Translation*. Cambridge University Press.

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

F. J. Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 160–167, Sapporo, Japan.

K. Papineni, S. Roukos, T. Ward, and W. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 311–318.

S. Riezler and J. T. Maxwell III. 2006. Grammatical machine translation. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 248–255. Association for Computational Linguistics.

V. M. Sánchez-Cartagena, F. Sánchez-Martínez, and J. A. Pérez-Ortiz. 2011a. Integrating shallow-transfer rules into phrase-based statistical machine translation. In *Proceedings of the XIII Machine Translation Summit*, pages 562–569, Xiamen, China, September.

V. M. Sánchez-Cartagena, F. Sánchez-Martínez, and J. A. Pérez-Ortiz. 2011b. The Universitat d'Alacant hybrid machine translation system for wmt 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 457–463, Edinburgh, Scotland, July. Association for Computational Linguistics.

V. M. Sánchez-Cartagena, J. A. Pérez-Ortiz, and F. Sánchez-Martínez. 2014. A generalised alignment template formalism and its application to the inference of shallow-transfer machine translation rules from scarce bilingual corpora. *Computer Speech and Language*. Submitted to the Special Issue on Hybrid Machine Translation.

F. Sánchez-Martínez and M. L. Forcada. 2009. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research*, 34(1):605–635.

H. Schwenk, S. Abdul-Rauf, L. Barrault, and J. Senellart. 2009. SMT and SPE Machine Translation Systems for WMT'09. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, StatMT '09, pages 130–134, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *In Proceedings of Association for Machine Translation in the Americas*, pages 223–231.

F. M. Tyers. 2009. Rule-based augmentation of training data in Breton–French statistical machine translation. In *Proceedings of the 13th Annual Conference of the European Association of Machine Translation*, pages 213–217.

Y. Xu, T. K. Ralphs, L. Ladányi, and M. J. Saltzman. 2009. Computational experience with a software framework for parallel integer programming. *INFORMS Journal on Computing*, 21(3):383–397.