Marker-based Filtering of Bilingual Phrase Pairs for SMT

Felipe Sánchez-Martínez
Dept. Llenguatges i Sistemes Informàtics
Universitat d'Alacant
E-03071 Alacant, Spain
fsanchez@dlsi.ua.es

Andy Way
NCLT, School of Computing
Dublin City University
Dublin 9, Ireland
away@computing.dcu.ie

Abstract

State-of-the-art statistical machine translation systems make use of a large translation table obtained after scoring a set of bilingual phrase pairs automatically extracted from a parallel corpus. The number of bilingual phrase pairs extracted from a pair of aligned sentences grows exponentially as the length of the sentences increases; therefore, the number of entries in the phrase table used to carry out the translation may become unmanageable, especially when online, 'on demand' translation is required in real time. We describe the use of closed-class words to filter the set of bilingual phrase pairs extracted from the parallel corpus by taking into account the alignment information and the type of the words involved in the alignments. On four European language pairs, we show that our simple yet novel approach can filter the phrase table by up to a third yet still provide competitive results compared to the baseline. Furthermore, it provides a nice balance between the unfiltered approach and pruning using stop words, where the deterioration in translation quality is unacceptably high.

1 Introduction

The state-of-the-art statistical approach to machine translation (MT) is the phrase-based model. Phrase-based statistical MT (PB-SMT) systems (Zens et al., 2002; Koehn et al., 2003) are based on the log linear model combination of several feature functions (Och and Ney, 2002), one of which is the *phrase translation probability* estimated after extracting bilingual phrase pairs from the parallel corpus.[1]

Bilingual phrase pairs are automatically extracted after computing the word alignments (Brown et al., 1993; Och and Ney, 2003). The set $\mathrm{BP}(s_1^J, t_1^I, A)$ of bilingual phrase pairs extracted from the word-aligned sentence pair $s_1^I = (s_1, \ldots, s_i, \ldots, s_I)$ and $t_1^J = (t_1, \ldots, t_j, \ldots, t_J)$ is defined as in (1) (Zens et al., 2002):

(1)

$$\mathrm{BP}(s_1^I, t_1^J, A) = \{(s_i^{i+n}, t_j^{j+m}) : \\ \forall (i', j') \in A : i \leq i' \leq i+n \Leftrightarrow j \leq j' \\ \leq j+m\},$$

where $A = \{(i,j) : i \in [1, I] \wedge j \in [1, J]\}$ is a set of pairs with the alignment information between the words in the source sentence s_1^I and the words in the target sentence t_1^J.

According to equation (1), all words within a bilingual phrase pair are consecutive and not aligned with words from outside the bilingual phrase pair. It is worth noting that bilingual phrase pairs may contain words that are not aligned at all, even at the beginning or the end of the phrase.

In order to make the extraction of bilingual phrase pairs computationally tractable, it is normally the case that only those possible pairs within a certain *n*-gram length are considered because the number of possible bilingual phrase pairs grows exponentially with the length of the sentences. In such cases, the amount of phrase pairs extracted from the whole training corpus may render the resulting translation table unmanageable in terms

[1]In the context of SMT, a phrase may be any sequence of consecutive words, not necessarily syntactic constituents.

of memory usage, even for large-scale system deployment.

The building of such large-scale systems is the norm for research groups participating in MT evaluations such as NIST or WMT, and it is generally fine to do this where once-off translation is required, such as in the bulk localisation scenario for large multinational software companies, for example.

However, where online 'on demand' translation is required, it is completely impractical to deploy the exact same systems in these workflows. Such situations include multilingual call centre scenarios, or where users from different languages want to interact in real time (consider a 'multilingual Facebook' scenario, for instance). In addition, the problem of access to information is increasing all the time in a world where both delivery and interface devices are changing massively to enable pervasive, on-the-move access to digital content. One such example is the CMU Transtac "eyes-free and hands-free" two-way speech-to-speech translation system for translation in the field between English–Iraqi Arabic and English–Farsi (Bach et al., 2007).

For all these reasons, therefore, many researchers have begun to investigate ways in which intelligible translations can be produced in real time. In this regard, we devised a simple yet novel filtering approach based on the "Marker Hypothesis" (Green, 1979) (cf. section 3). Essentially, we use linguistic information in the form of closed-class word lists to filter the set of bilingual phrase pairs, by taking into account the alignment information and the type of the words involved in the alignments.

The inspiration for the set of experiments carried out in this paper was that successful Example-Based MT (EBMT) systems (Nagao, 1984; Carl and Way, 2003) have been built using the Marker Hypothesis to segment source–target aligned sentence pairs into linguistically motivated bilingual chunks (cf. (Way and Gough, 2003; Gough and Way, 2004)). These systems have proven to be particularly useful where good translation performance is required with much smaller translation tables than are traditionally used in PB-SMT. For example, Groves and Way (2005a) showed that for a range of systems built with different amounts of data, on average the translation table of a PB-SMT system was about five times the size of the equivalent EBMT system. In a related paper, on a training set of 203K English–French aligned sentence pairs, Groves and Way (2005b) showed that seeding a PB-SMT system built using Pharaoh (Koehn, 2004a) with 403,317 EBMT alignments, a BLEU score of 36.43 was obtained, compared to a score of 37.53 with 1,732,715 phrase pairs built using Giza++ (Och and Ney, 2003).

The remainder of this paper is organised as follows. Section 2 reviews other research work that has also focused on the filtering of the bilingual phrase pairs. Then, in section 3 we describe our approach. Section 4 describes the experiments conducted on four language pairs and the results achieved. The paper ends with our concluding remarks together with avenues for further research.

2 Related Work

Previous approaches to filter the phrase pairs used in PB-SMT can be divided into two classes:

- those methods that filter the phrase table according to the text to be translated;
- those more general approaches that filter the set of bilingual phrase pairs extracted from the training corpus, or the translation table directly, without knowing in advance which texts are to be translated.

Our approach falls into this latter category (cf. section 3).

In the first group of approaches we find the work by Koehn (2004a), which performs a rudimentary analysis of the sentences to translate in order to minimise the number of entries in the phrase table to be loaded into memory. A more sophisticated approach is performed in (Badr et al., 2007), where the authors consider the relationship between different translation models to obtain a much smaller set of phrases associated with each sentence to translate.

The approach in (Lü et al., 2007) may be considered somewhat in-between the two filtering classes because while translation table pruning is performed at training time, the authors know in advance what text is to be translated once training is completed. Lü et al. (2007) use well-known information retrieval methods to select sentences from the training corpus that better match the domain of the test corpus. Then these sentences are used to optimise the distribution of the whole training corpus.

In the second group of approaches, Eck et al. (2005) sort the training sentences according to the frequency of unseen n-grams so as to select a reduced number of sentences for training SMT systems to run on small devices. They also propose the use of information retrieval methods for that purpose. Ma et al. (2007) use alignment-guided chunking to filter the size of the translation table by 78.6%, with a 2.93-point reduction (about 15%) in BLEU score for German–English experiments.[2] Somewhat more impressively, Johnson et al. (2007) perform significance testing (Agresti, 1996) to select the sentences to be used in the training phase. They report a 90% reduction in the phrase table on various language pairs of the Europarl (Koehn, 2005) parallel corpus without any reduction in the translation quality achieved, as measured by BLEU.

3 Marker-based Filtering of the Bilingual Phrase Pairs

All words in a language can be classified into two different categories, namely closed and open classes. Closed-class words (henceforth, *closed* words) such as prepositions, pronouns or articles, may be thought as the *core* words of a language, i.e. as the words providing the structure for well-formed sentences, but without any special intrinsic meaning. In contrast, open-class words (*open words*, such as nouns or verbs) may be thought of as the words which express the meaning of a sentence. This difference between closed words and open words explains why no new words are usually added to the set of closed words, while the set of open words can easily grow as a language evolves.

Having a set of words that provides the structure for the remaining words to express their meaning is known as the Marker Hypothesis (Green, 1979), which states that the syntactic structure of a language is *marked* at the surface level by a closed set of *marker* (closed) words.

As stated in the introduction, this paper is not the first approach that has used the Marker Hypothesis in MT. While most work has centred on building EBMT systems which relate source and target phrase pairs comprised of words (e.g. (Juola, 1994; Way and Gough, 2003; Gough and Way, 2004; Groves and Way, 2005a)), systems have also been successfully constructed where phrases consist of word–morpheme mappings (e.g. (Stroppa et al., 2006) for English–Basque, and (Labaka et al., 2007) for Spanish–Basque). Marker-based chunking still plays a significant role in the MaTrEx (Stroppa and Way, 2006; Hassan et al., 2007; Tinsley et al., 2008; Du et al., 2009) system, where the sets of marker words needed for bilingual chunking are extracted automatically rather than by assembling these by hand as its predecessors did (e.g. (Way and Gough, 2003; Gough and Way, 2004)).

To give the reader some idea of how marker words are used in practice in such systems, we revisit an example from (Groves and Way, 2005a), namely (2) (from (Koehn, 2005), Figure 2):

(2) that is almost a personal record for me this autumn!
⟶c' est pratiquement un record personnel pour moi , cet automne!

.

Once 7 sets of closed-class words (determiners, quantifiers, conjunctions, prepositions, wh-adverbs, possessive and personal pronouns, cf. (6) below) have been built for English and French, the marker words in (2) can be tagged, as in (3):

(3) <DET> that is almost <DET> a personal record
<PREP> for <PRON> me <DET> this autumn!
⟶<DET> c' est pratiquement <DET> un record personnel <PREP> pour <PRON> moi , <DET> cet automne!

Then using marker tag information (label, and relative sentence position), and lexical similarity (via mutual information), the marker chunks in (4) are automatically generated from the marker-tagged strings in (3):

(4) a. <DET> that is almost : <DET> c' est pratiquement
b. <DET> a personal record : <DET> un record personnel
c. <PREP> for me this autumn : <PREP> pour moi cet automne

Should they be required, the set of generalised templates in (5) can be derived automatically from the bilingual phrase pairs in (4):

[2]While these results do not appear in the paper, they were included in the presentation accompanying that paper. Thanks to Yanjun Ma for this clarification.

(5) a. <DET> is almost : <DET> est pratiquement

b. <DET> personal record : <DET> record personnel

c. <PREP> me this autumn : <PREP> cet automne

All of these resources, together with a bilingual lexicon (in later work, induced via Giza++), are brought to bear in translating new input strings to good effect.

In this work, our approach to filter the set of bilingual phrase pairs used to train a PB-SMT system is based on the Marker Hypothesis and the following intuitive idea: as closed words provide the structure and open words provide the meaning, accurate bilingual phrase pairs should have an alignment between the open words of the languages involved in the translation, while closed words may remain unaligned, as the syntactic structure changes from one language to another. This same idea was used by Sánchez-Martínez and Forcada (2009) to filter the bilingual phrase pairs used in the inference of shallow-transfer rules for the Apertium MT platform (Armentano-Oller et al., 2006).

We have explored two different criteria to filter the set of bilingual phrase pairs. The first one ("open words alig") discards those phrase pairs presenting open words, in one or both languages, not aligned with at least one open word of the other language. The second criterion ("open words alig+borders") is more restrictive, as it discards all phrase pairs discarded by the first criterion and also those phrase pairs whose first or last words are not aligned with any word of the other language—in this case no matter whether the first and last words in each language are closed words or open words. We experimented with this second criterion because, as a result of the bilingual phrase pairs extraction algorithm, unaligned words may appear at the beginning or the end of a phrase, and we wanted to test whether this introduces any noise in the translation table. Note that, after extracting the set of bilingual phrase pairs from a word-aligned sentence pair, two or more bilingual phrase pairs may only differ in that some of them contain unaligned words at the beginning or the end of the phrases.

4 Experiments

We tested the presented approach on the following language pairs: Spanish–English (`es-en`), English–Spanish (`en-es`), French–English (`fr-en`), and English–French (`en-fr`). We used the data distributed for the WMT09[3] Workshop on MT both for training and testing. Unfortunately, the test sets contain only one reference translation, which causes the scores obtained to be somewhat lower than might otherwise have been expected.

All the experiments were performed using the Moses open-source decoder for PB-SMT (Koehn et al., 2007) and the SRILM language modelling toolkit (Stolcke, 2002). Training was carried out as follows:

1. Word alignments were obtained using Giza++ (Och and Ney, 2003) and symmetrized in the usual way (Koehn et al., 2003).

2. Bilingual phrase pairs were extracted from the word-aligned sentence pairs.

3. Extracted phrase pairs were filtered following the approach presented in this paper, and then scored.

4. Weights were optimised using minimum error rate training (MERT) in the usual manner (Och, 2003).

With the two filtering criteria explained in Section 3, we tested different lists of words:

closed words: A list of closed words in each language is provided to the filtering algorithm. These lists contain determiners, prepositions, pronouns, coordinate and subordinate conjunctions, relative and possessive pronouns, and punctuation marks. They consist of 193 Spanish words, 174 French words and 185 English words. Examples include those in (6):

[3] `http://www.statmt.org/wmt09/`

(6) *English*:

⟨DET⟩: {the, a, some ...}
⟨PREP⟩: {on, at, in ...}
⟨PRON⟩: {you, he, she ...}

French:

⟨DET⟩: {le, la, les ...}
⟨PREP⟩: {sur, dans, par ...}
⟨PRON⟩: {vous, il, me ...}

Spanish:

⟨DET⟩: {el, la, los ...}
⟨PREP⟩: {de, para ...}
⟨PRON⟩: {yo, tú, usted ...}

closed words+vaux: In addition to the closed words discussed above, all inflected forms of auxiliary verbs and modal verbs in each language were used. The verbs considered are: *deber*, *haber*, *poder*, *querer* and *ser* for Spanish; *avoir*, *devoir*, *être*, *falloir*, *pouvoir* and *vouloir* for French; and *be* and *have* for English. In sum, they consist of 1,572 Spanish words, 490 French words and 201 English words. The large number of words in Spanish is due to inflected forms with enclitic pronouns attached.

stop words: With the aim of avoiding the need to manually build a list of closed words in each language, we tested the use of stop words automatically obtained from the training corpus (cf. (Stroppa and Way, 2006), who extract the required sets of marker words automatically from online dictionaries). The underlying assumption here is that closed words, as the core words of a language, are very frequent and, therefore, will appear in every list of stop words.

Table 1 shows for the two filtering criteria ("open words alig" and "open words alig+borders", see Section 3) and for the different lists of words explained above, the percentage of bilingual phrase pairs discarded and the translation performance of the resulting translation model as evaluated with BLEU (Bilingual evaluation understudy (Papineni et al., 2002)) and TER (Translation edit rate (Snover et al., 2006)).

In all cases the baseline system, i.e. when no filtering of the bilingual phrase pairs is done, performs better than our approach. However, a significant reduction in the number of phrase pairs (around 25% for es-en and en-es, and around 33% for fr-en and en-fr) can be achieved at the cost of a small loss in the translation performance (around 0.012 in BLEU for the former two pairs, and about 0.017 for the latter). That said, having conducted significance testing using boostrap resampling (Koehn, 2004b), these reductions in translation quality are significant. Of course, greater reductions in the number of phrase pairs can be achieved at an even higher cost in terms of translation quality.

With respect to which is the best filtering criterion, the results in Table 1 show that, as expected, a greater reduction—on average around 9-10%—in the number of bilingual phrase pairs is achieved through the "open words alig+borders" criterion. Nevertheless, this greater reduction is at the cost of a higher loss in translation quality, although only around 0.5 BLEU points on average across the board.

As for the list of words used by the filtering algorithm, it can be concluded that using a list of closed words gives better performance than the remaining lists of words. For example, for es-en, around a quarter of the phrase pairs are filtered with just over one BLEU point reduction in performance. For en-fr, the gap is even larger, with a one third reduction in the size of the translation table and competitive MT performance.

However, note that in some language pairs, adding auxiliary and modal verbs to the list of closed words provided slightly better results. For all language pairs bar es-en, although fewer phrase pairs were filtered, MT performance was better when auxiliaries and modals were included in the set of marker words, although the differences in MT performance were not always statistically significant.

Compared with the translation performance achieved when using stop words, it becomes clear that the use of closed words provides better results. For es-en and en-es, around 50% relatively more phrase pairs are filtered using stop words, but performance decreases by up to 3 BLEU points compared to when closed words are using as the filter. For fr-en and en-fr, we observe a drop in performance of around 1.5 BLEU points when

		open words alig			open words alig+borders		
Lang. pair	List of words	filtered pairs	BLEU	TER	filtered pairs	BLEU	TER
es-en	baseline		0.2355	0.6416		0.2355	0.6416
	closed words	24.73%	0.2232	0.6570	34.80%	0.2170	0.6673
	closed words+vaux	23.72%	0.2188	0.6644	34.69%	0.2157	0.6675
	200 stop words	36.42%	0.1952	0.6889	46.34%	0.1921	0.6885
	100 stop words	36.59%	0.1991	0.6818	45.96%	0.1942	0.6882
	50 stop-words	31.63%	0.2090	0.6705	41.15%	0.2037	0.6766
en-es	baseline		0.2208	0.6588		0.2208	0.6588
	closed words	24.72%	0.2090	0.6701	34.71%	0.2032	0.6823
	closed words+vaux	23.69%	0.2112	0.6713	34.59%	0.2039	0.6796
	200 stop words	36.38%	0.1845	0.6975	46.24%	0.1807	0.7077
	100 stop words	36.57%	0.1888	0.6935	45.86%	0.1838	0.7021
	50 stop words	31.64%	0.2014	0.6826	41.98%	0.1943	0.6897
fr-en	baseline		0.2331	0.6476		0.2331	0.6476
	closed words	33.04%	0.2128	0.6700	41.26%	0.2072	0.6747
	closed words+vaux	30.74%	0.2130	0.6693	40.16%	0.2076	0.6763
	200 stop words	36.20%	0.1947	0.6882	47.08%	0.1878	0.6932
	100 stop words	34.61%	0.2027	0.6795	44.89%	0.1968	0.6825
	50 stop words	35.14%	0.2082	0.6742	44.31%	0.2029	0.6790
en-fr	baseline		0.2105	0.6993		0.2105	0.6993
	closed words	33.08%	0.1965	0.7125	41.20%	0.1928	0.7208
	closed words+vaux	30.75%	0.1990	0.7114	40.07%	0.1957	0.7155
	200 stop words	36.17%	0.1807	0.7297	46.97%	0.1760	0.7345
	100 stop words	34.65%	0.1865	0.7239	44.81%	0.1798	0.7352
	50 stop words	35.18%	0.1903	0.7244	44.24%	0.1885	0.7241

Table 1: For each language pair, percentage of bilingual phrase pairs discarded and translation performance, as evaluated by BLEU and TER, for the two filtering criteria explained in Section 3, and for the different lists of words used in the experiments.

200 stop words are used, but with only around 10% relatively extra phrase pairs being filtered.

5 Conclusion and Further Work

We have presented a simple yet novel approach that may be used to filter the bilingual phrase pairs extracted from the parallel training corpus for deployment in PB-SMT in situations where a smaller system footprint is required. Our approach is based on the Marker Hypothesis and on the intuitive idea that the open-class words in a bilingual phrase pair should be aligned because they are responsible for the meaning, while it is less costly for closed-class words to remain unaligned.

The approach was widely tested on four European language pairs using different lists of closed words and two different filtering criteria. The results show that more than one quarter of the bilingual phrase pairs can be ruled out at the cost of a small (yet statistically significant) loss in translation quality. Despite this drop in performance, it is clear that more and more real examples are coming to the fore where a smaller translation table is absolutely necessary, such as the integration of PB-SMT systems in mobile devices, or to enable online on-demand translation between speakers having no common language.

As for future work, we plan to test whether the results may be improved if prepositions are not considered as closed-class words when they are part of a phrasal verb. In these cases the preposition changes the meaning of the verb and, therefore, does not play the role of a closed-class word in terms of the Marker Hypothesis.

We will also test our approach for the translation from English to a non-European language such as Chinese, Japanese or Hindi. Chinese is the more difficult language, since it lacks some markers that would help to identify when a noun phrase is started. Japanese is easier, since in the creation of the language (from Chinese), some markers were introduced to facilitate reading. Hindi, as is the case with all Indian languages, has a one-to-one mapping to English word classes, and so we are confident that similar benefits may accrue as for the European languages tested in this paper.

Finally, we plan to deploy our system in a multilingual chat environment with a well-known multinational software company, as well as develop a 'multilingual Facebook'-type demonstration system. It will be interesting to see to what extent our distinction between open and closed words proves particularly instrumental under such conditions.

Acknowledgements

We thank Mikel L. Forcada for his advice and encouragement in the generation of this paper. We also thank Yanjun Ma, Tsuyoshi Okita, Sandipan Dandapat, Rejwanul Haque and Sudip Naskar for fruitful discussions on the application of the Marker Hypothesis to Asian languages. We would also like to thank the anonymous reviewers for useful comments which served to improve our work. Finally, we would like to thank our funding agencies: the first author thanks the support of the Spanish Ministry of Education and Science (project TIN2006-15071-C03-01); the second author acknowledges the support from Science Foundation Ireland (`http://www.sfi.ie`) through grants 05/IN/1732 and 06/RF/CMS064.

References

Agresti, A. 1996. *An introduction to categorical data analysis*. Wiley, New York.

Armentano-Oller, C., R. C. Carrasco, A. M. Corb-Bellot, M. L. Forcada, M. Ginestí-Rosell, S. Ortiz-Rojas, J. A. Pérez-Ortiz, G. Ramírez-Sánchez, F. Sánchez-Martínez, and M. A. Scalco. 2006. 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*, volume 3960 of *Lecture Notes in Computer Science*, pages 50–59. Springer-Verlag, Itatiaia, Brazil, May.

Bach, N., M. Eck, P. Charoenpornsawat, T. Köhler, S. Stüker, T. Nguyen, R. Hsiao, A. Waibel, S. Vogel, T. Schultz, and A. Black. 2007. The CMU Transtac 2007 eyes-free and hands-free two-way speech-to-speech translation system. In IWSLT 2007*, Proceedings of the 4th International Workshop on Spoken Language Translation*, pages 29–36, Trento, Italy.

Badr, G., E. Joanis, S. Larkin, and R. Kuhn, 2007. *Manageable Phrase-based Statistical Machine Translation Models*, chapter Computer Recognition Systems 2 (CORES 2007). Book Series: Advances in Soft Computing, pages 437–444. Springer, Berlin, Germany, October.

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

Carl, M. and A. Way, editors. 2003. *Recent Advances in Example-Based Machine Translation*, volume 21 of *Text, Speech and Language Technology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Du, J., Y. He, S. Penkale, and A. Way. 2009. Matrex: the DCU MT system for WMT 2009. In *Proceedings of the Fourth Workshop on Statistical Machine Translation, EACL 2009*, page in press, Athens, Greece.

Eck, M., S. Vogel, and A. Waibel. 2005. Low cost portability for statistical machine translation based on n-gram coverage. In *MT Summit X*, pages 227–234, Phuket, Thailand.

Gough, N. and A. Way. 2004. Robust large-scale EBMT with marker-based segmentation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 95–104, Baltimore, MD.

Green, T. 1979. The necessity of syntax markers. two experiments with artificial languages. *Journal of Verbal Learning and Behavior*, 18:481–496.

Groves, D. and A. Way. 2005a. Hybrid data-driven models of machine translation. *Machine Translation*, 19(3–4):301–323.

Groves, D. and A. Way. 2005b. Hybrid example-based SMT: the best of both worlds? In *Proceedings of the Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond, ACL 2005*, pages 183–190, Ann Arbor, MI.

Hassan, H., Y. Ma, and A. Way. 2007. MATREX: the DCU machine translation system for IWSLT 2007. In IWSLT 2007*, Proceedings of the 4th International Workshop on Spoken Language Translation*, pages 69–75, Trento, Italy.

Johnson, H., M. Joel, G. Foster, and R. Kuhn. 2007. Improving translation quality by discarding most of the phrasetable. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 967–975, Prague, Czech Republic.

Juola, P. 1994. A psycholinguistic approach to corpus-based machine translation. In *Proceedings of the Third International Conference on the Cognitive Science of Natural Language Processing (CSNLP-94), [no page numbers]*.

Koehn, P. 2004a. Pharaoh: A beam search decoder for phrase-based statistical machine translation models. In *Proceedings of the 6th biennial conference of the Association for Machine Translation in the Americas*, pages 115–124, Washington, DC.

Koehn, P. 2004b. Statistical significance tests for machine translation. In *2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 388–395, Barcelona, Spain.

Koehn, P. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the Tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. `http://www.statmt.org/europarl/`.

Koehn, P., F. J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2003*, pages 48–54, Edmonton, AL., Canada, May-June.

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic.

Labaka, G., N. Stroppa, A. Way, and K. Sarasola. 2007. Comparing rule-based and data-driven approaches to spanish-to-basque machine translation. In *MT Summit XI*, pages 297–30, Copenhagen, Denmark.

Lü, Y., J. Huang, and Q. Liu. 2007. Improving statistical machine translation performance by training data selection and optimization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007)*, pages 343–350, Prague, Czech Republic.

Ma, Y., N. Stroppa, and A. Way. 2007. Alignment-guided chunking. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 114–121, Skövde, Sweden.

Nagao, M. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In Elithorn, Alick and Ranan Banerji, editors, *Artificial and Human Intelligence*, pages 173–180, Amsterdam, The Netherlands. North-Holland.

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

Och, F. J. and H. Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In *40th Annual Meeting of the Association for Computational Lingustics (ACL)*, pages 295–302, Philadelphia, PA.

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

Papineni, K., S. Roukos, T. Ward, and W. J. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *40th Annual meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA., July.

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

Snover, M., B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, "Visions for the Future of Machine Translation"*, pages 223–231, Cambridge, MA., August.

Stolcke, A. 2002. SRILM – an extensible language modeling toolkit. In *Proceedings of the International Conference on Spoken Language Processing*, pages 901–904, Denver, CO., June.

Stroppa, N. and A. Way. 2006. MaTrEx: the DCU machine translation system for IWSLT 2006. In *Proceedings of IWSLT 2006 Workshop*, pages 31–36, Kyoto, Japan.

Stroppa, N., D. Groves, A. Way, and K. Sarasola. 2006. Example-based machine translation of the Basque language. In *Proceedings of the 7th biennial conference of the Association for Machine Translation in the Americas*, pages 232–241, Cambridge, MA.

Tinsley, J., Y. Ma, S. Ozdowska, and A. Way. 2008. Matrex: the DCU MT system for WMT 2008. In *Proceedings of the Third Workshop on Statistical Machine Translation, ACL 2008*, pages 171–174, Columbus, OH.

Way, A. and N. Gough. 2003. *wEBMT*: Developing and validating an EBMT system using the world wide web. *Computational Linguistics*, 29(3):421–457.

Zens, R., F. J. Och, and H. Ney. 2002. Phrase-based statistical machine translation. In *KI 2002: Advances in Artificial Intelligence: Proceedings 25th Annual German Conference on AI*, volume 2479 of *Lecture Notes in Computer Science*, pages 18–32, Berlin, Germany. Springer-Verlag.