Building hybrid machine translation systems by using an EBMT preprocessor to create partial translations

Mikel Artetxe, Gorka Labaka, Kepa Sarasola
IXA NLP Group, University of the Basque Country (UPV/EHU)
{martetxe003@ikasle., gorka.labaka@, kepa.sarasola@}ehu.eus

Abstract

This paper presents a hybrid machine translation framework based on a preprocessor that translates fragments of the input text by using example-based machine translation techniques. The preprocessor resembles a translation memory with named-entity and chunk generalization, and generates a high quality partial translation that is then completed by the main translation engine, which can be either rule-based (RBMT) or statistical (SMT). Results are reported for both RBMT and SMT hybridization as well as the preprocessor on its own, showing the effectiveness of our approach.

1 Introduction

The traditional approach to Machine Translation (MT) has been rule-based (RBMT), but it has been progressively replaced by Statistical Machine Translation (SMT) since the 1990s (Hutchins, 2007). Example-Based Machine Translation (EBMT), the other main MT paradigm, has never attracted that much attention: even though it gives excellent results with repetitive text for which accurate matches are found in the parallel corpus, its quality quickly degrades as more generalization is needed. Nevertheless, it has been argued that, along with the raise of hybrid systems that try to combine multiple paradigms, EBMT can help to overcome some of the weaknesses of the other approaches (Dandapat et al., 2011)[1].

In this paper, we propose one such system based on a multi-pass system combination: an EBMT preprocessor translates those fragments of the input text for which accurate matches are found in the parallel corpus, generating a high-quality partial translation that is then completed by the main translator, which can be either rule-based or statistical.

The function of the EBMT preprocessor is therefore similar to that of Translation Memories (TM), with the difference that previously made translations are not reused to aid human translators but a MT engine. Needless to say, if the EBMT preprocessor was only able to reuse full sentences as traditional TM systems do at the most basic level, the quality of its partial translations would match that of humans, but its contribution would be negligible in most situations. At the same time, trying to increase the coverage by generalizing too much at the expense of translation quality, as traditional EBMT systems do, would make the whole system pointless if the preprocessor is not able to outperform the main MT engine for the fragments it translates. This way, for our approach to work as intended, it is necessary to find a trade-off between coverage and translation quality. In this work, we take a preprocessor that reuses full sentences as our starting point and explore two generalization techniques similar to those used by second and third generation TM systems (Gotti et al., 2005):

- **Named-entity (NE) generalization**, giving the option to replace NEs like proper names and numerals in the parallel corpus with any other found in the text to translate.

[1]This paper refers to as hybridization to any combination of MT paradigms, no matter if they are integrated in a single engine or not. However, some authors distinguish between hybridization for systems that meet this requirement and combination for systems that do not.

- **Chunk generalization**, giving the option to reuse examples in a subsentential level.

Several other methods that combine EBMT and TM with other MT paradigms have been proposed in the literature. Koehn and Senellart (2010) use an SMT system to fill the mismatched parts from a fuzzy search in a TM. Similarly, Shirai et al. (1997) use a RBMT engine to complete the mismatched fragments from an EBMT system and smooth the resulting output using linguistic rules. On the other hand, Dandapat et al. (2012) integrate SMT phrase tables into an EBMT framework. Following the opposite approach, Groves and Way (2005) feed an SMT system with alignments obtained using EBMT techniques. Sánchez-Martínez et al. (2009) use EBMT techniques to obtain bilingual chunks that are then integrated into a RBMT system. Lastly, Alegria et al. (2008) propose a multi-engine system that selects the best translation created by a RBMT, an SMT and an EBMT engine. However, to the best of our knowledge, the use of a generic multi-pass hybridization method for EBMT that works with both SMT and RBMT has never been reported so far.

The remaining of this paper is structured as follows. The proposed method is presented in Section 2. Section 3 explains the experimental settings under which the system was tested, and the results obtained are then discussed in Section 4. Section 5 concludes the paper.

2 Method

Our method follows the so-called compiled approach to EBMT, which differs from runtime or pure EBMT in that it requires a training phase to compile translation units below the sentence level (Dandapat, 2012). Therefore, the system we propose consists of three elements: the compiling component presented in Section 2.1, which analyzes and aligns the parallel corpus to be used by the EBMT preprocessor; the EBMT preprocessor itself as presented in Section 2.2, which creates a high-quality partial translation of the input text using the data created by the previous module; and the integration with the main translator presented in Section 2.3, which completes the partial translation given by the previous module by using either a RBMT or an SMT engine.

2.1 Compiling

The compiling phase involves processing a sentence-aligned parallel corpus to be used by the EBMT preprocessor. Two steps are required for this: the analysis step, presented in Section 2.1.1, and the alignment step, presented in Section 2.1.2. The resulting data is encoded in a custom binary format based on suffix arrays (Manber and Myers, 1990) for its efficient retrieval by the EBMT preprocessor.

2.1.1 Analysis

The analysis step involves the tokenization, NE recognition and classification, lemmatization and parsing of each side of the parallel corpus. We have used Freeling (Padró and Stanilovsky, 2012) as our analyzer for Spanish, Stanford CoreNLP (Socher et al., 2013) for English and Eustagger (Ezeiza et al., 1998) for Basque, with a custom regex-based handling for numerals. The resulting constituency-based parse tree is simplified by removing inner nodes that correspond to part-of-speech tags and representing NEs as single leaves. In the case of Basque, our analyzer is only capable of shallow parsing, so we have generated a dummy tree in which chunks are the only inner nodes.

2.1.2 Alignment

The alignment step involves establishing the translation relationships among the tokens[2] and NEs of the parallel corpus. This is done separately because the latter serves as the basis for NE generalization as discussed in Section 1, so we allow the option of not aligning NEs in this level if there is not enough evidence to do so.

This way, word-alignment produces a set A_n for each nth sentence pair where $(i, j) \in A_n$ if and only if there is a translation relationship between the ith token in the source language and the jth token in the target language, as well as the lexical weightings or translation probabilities in both directions, that is, a set of $p(e|f)$ and $p(f|e)$ probabilities that express the likelihood of the f token to be translated as e and the e token to be translated as f, respectively. Our system has been integrated both with GIZA++ (Och and Ney, 2003) and Berkeley Aligner (Liang et al., 2006).

As for NE alignment, we align NEs if and only if they have the same written form, are equivalent

---

[2] We refer as tokens to the leaves of the parse tree obtained in the analysis phase, which implies that NEs are considered (multiword) tokens.

numerals or are found in either of the following dictionaries:

- A manually built dictionary, mostly consisting of translation relationships between proper names like countries.

- An automatically generated dictionary from Wikipedia article titles with support for redirections.

- An automatically generated dictionary from word-alignment, consisting of every NE pair $f - e$ for which $\frac{p(e|f)+p(f|e)}{2} > \theta$ and f and e appear a minimum of l times in the corpus.[3]

2.2 EBMT preprocessing

The goal of the EBMT preprocessing is to create a high-quality partial translation of the input text. As it is common in EBMT, this is done in three steps: matching, alignment and recombination, which are described in the following subsections.

2.2.1 Matching

The matching phase involves looking for fragments of the input text in the training corpus. For this purpose, the input text is first analyzed as described in Section 2.1.1, and chunks of each of the input sentences are then searched in the parallel corpus according to the following criteria:

1. The searched chunks must be syntactic units (either inner nodes or groups of consecutive inner siblings).

2. The searched chunks must contain a minimum of k tokens to avoid trivial translations that would have a negative impact on the overall translation quality. After some preliminary experiments, we have set k to 4.

3. The search process is hierarchical, that is, nodes that are closer to the root have priority over the rest in case of overlapping matches. If overlapping matches are found in the same level of the parse tree, the chunk with the biggest number of tokens has priority over the rest.

4. Full syntactic match requirement, that is, not only the leaves of the searched chunks have to match but also their corresponding subtrees.

---

[3]Based on some preliminary experiments, we set $\theta = 0.5$ and $l = 10$.

5. The generalization of aligned NEs in the training corpus. According to this criterion, aligned NEs in the training corpus are considered to be valid matches for any NE in the input text, whereas unaligned NEs are processed as plain tokens.

2.2.2 Alignment

The next step in the EBMT preprocessing is to build a translation for each match, filtering those that are not valid. For that purpose, we first identify the translation that corresponds to each match in the parallel corpus, and we then translate the aligned NEs it contains.

For the first point, given a match of a chunk in the source language, we select the shortest sequence in the target language that satisfies the following conditions. If there is no possible translation that satisfies all these conditions for a given match, the match is rejected.

1. It must contain at least one aligned token.

2. No token in either fragment can be aligned with a token outside the other fragment.

3. The translation must be a syntactic unit as defined in Section 2.2.1, but without the requirement for the matched nodes to be inner ones (i.e. they could also be leaves).

Due to NE generalization, the translation generated this way might contain NEs that do not correspond to the searched ones. These NEs are translated as follows:

1. Identify the searched NE for each aligned NE in the translation. This is done by following the translation relationships as defined by NE alignment in the compiling phase.

2. Translate the lemma of the searched NEs. The set of dictionaries described in Section 2.1.2 is used for that purpose with a custom processing for numerals. NEs that cannot be translated by these means are left unchanged, as they would presumably correspond to proper names of persons or locations.

3. Inflect the translated lemma by applying the same morphological tags that the aligned NE had. We only apply this step for morphologically rich languages as it is the case of Basque.

For instance, if "Putin claims victory in Russia elections" is matched with "Peña Nieto claims victory in Mexico elections", and "Peña Nietok Mexikoko hauteskundeak irabazi ditu" is selected as its translation, we would first identify that the Basque "Peña Nietok" is aligned with the English "Peña Nieto", which was matched with "Putin", and "Mexikoko" is aligned with "Mexico", which was matched with "Russia". We would then translate "Putin" as "Putin" and "Russia" as "Errusia" according to the dictionaries described in Section 2.1.2. Lastly, we would inflect these lemmas to match the lexical form of their corresponding aligned NE. In this case, "Peña Nietok" was the ergative form of "Peña Nieto", so we would inflect "Putin" in ergative giving "Putinek". Similarly, "Mexikoko" was the local-genitive form of "Mexiko", so we would inflect "Errusia" in local-genitive giving "Errusiako". This way, we would obtain the final translation "Putinek Errusiako hauteskundeak irabazi ditu".

2.2.3 Recombination

After the alignment phase, it is possible to have either zero, one, or several translation candidates for each searched chunk. Thanks to the hierarchical searching process, it is guaranteed that these translations will not overlap, so rather than combining them we try to select the best candidate for each searched chunk. For that purpose, we choose the most frequent translation in each case and, in case of a tie, the one with the highest lexical weighting.

2.3 Integration

As discussed in the previous section, the EBMT preprocessor creates a partial translation of the input text by translating chunks that are matched in the training corpus. The next and last phase involves building the full translation by completing it with the help of the main MT system. This is done differently depending on the type of system it is:

- When hybridizing with RBMT systems, the input text is translated as it is, and a postprocessor replaces translation fragments that correspond to matched chunks with the ones proposed by the EBMT preprocessor. In order to identify these fragments, the original chunks are marked with XML tags that the main MT system keeps in the translation it generates.

- When hybridizing with SMT systems, Moses' XML markup is used in its "inclusive" mode to make the translations generated by the EBMT preprocessor compete with the entries in the phrase table. It is remarkable that the "exclusive" and "constraint" modes, which force the decoder to choose the proposed translation or others that contain it, respectively, gave consistently worse results. We speculate that this could be due to the boundary friction problem, as the EBMT system translates fragments without taking their context into account, and the language model might be able to choose a better translation for the given context.

3 Experimental settings

As discussed in Section 1, it is expected that the performance of our method will greatly depend on the similarity between the input text and the examples given in the training corpus. Taking that into account, we decided to train our system in two different domains: the particularly repetitive domain of collective bargaining agreements, and the more common domain of parliamentary proceedings. For the former, we used the Spanish-Basque IVAP corpus, consisting of a total of 81 collective bargaining agreements to which we added the larger Elhuyar's administrative corpus to aid word-alignment. For the latter, we used the Spanish-English Europarl corpus as given in the shared task of the ACL 2007 workshop on statistical machine translation, consisting of proceedings of the European Parliament. Table 1 summarizes their details. As for the testing data, we used an in-domain test set for each corpus as well as an out-of-domain one for Europarl as shown in Table 2.

In order to evaluate the performance of our method we carried out the following two experiments:

- **A manual evaluation of the EBMT preprocessor** to measure both the coverage and the quality of its partial translations. For this purpose, we randomly selected 100 sentences for each in-domain test set and asked 5 volunteers to score the quality of each translated fragment in its context in a scale between 1 (incorrect translation) and 4 (correct translation).

- **An automatic evaluation of the whole system** using the Bilingual Evaluation Under-

	Language	Domain	Sentences
IVAP + Elhuyar	es-eu	collective bargaining agreements + administrative	50,824 + 4,747,332
Europarl	es-en	parliament proceedings	1,254,414

Table 1: Training corpus

	Language	Domain	In domain?	Sentences	Tokens	Tokens / sentence
IVAP	es-eu	collective bargaining agreement	yes	1,928	39,625	20.55
Europarl	es-en	parliamentary proceedings	yes	2,000	56,213	28.01
News commentary	es-en	news	no	2,007	61,341	30.67

Table 2: Test set

	Full sentences	Full sentences with NE	Chunks with NE		
			GIZA++	Berkeley (HMM)	Berkeley (synt.)
IVAP	18,284 (46.14%)	18,691 (47.17%)	23,962 (60.47%)	26,436 (66.72%)	-
Europarl	379 (0.62%)	548 (0.89%)	10,565 (17.22%)	10,986 (17.91%)	9,653 (15.74%)
News commentary	12 (0.02%)	12 (0.02%)	5,365 (9.54%)	5,566 (9.90%)	4,674 (8.31%)

Table 3: Tokens translated by the EBMT preprocessor

study (BLEU) metric (Papineni et al., 2002). For this automatic evaluation, we hybridized our system both with a RBMT and an SMT system. Our RBMT translator of choice was Matxin (Mayor et al., 2011) for Spanish-Basque and Apertium (Forcada et al., 2011) for Spanish-English, whereas we used Moses (Koehn et al., 2007) as our SMT engine for both language pairs.

4 Results and discussion

This section presents the outcomes of the experiments described in Section 3. The results for the quality and coverage experiment are discussed in Section 4.1, and the RBMT and SMT hybridization in Sections 4.2 and 4.3.

4.1 Quality and coverage of EBMT

Table 3 shows the number of tokens translated by the EBMT preprocessor according to each generalization mechanism. In the case of chunk generalization, we tried both GIZA++ and Berkeley aligner with and without syntactic tailoring (DeNero and Klein, 2007), which could presumably generate more chunk alignments that meet the restrictions of our translation process. However, contrary to our expectations syntactic tailoring gave the worst results by far both in terms of coverage and translation quality, apparently because it is still an experimental feature, and it was the default HMM mode of Berkeley Aligner which clearly outperformed the rest. We will consequently refer to the results obtained by this aligner in the remaining of this section.

As we expected, Table 3 reflects that the cover-age of the EBMT preprocessing clearly depends on the similarity between the input text and the training corpus. For the domain of collective bargaining agreements, our EBMT preprocessor is able to translate around two thirds of the input tokens. Even though the results we obtain for the other test sets are poorer, the impact of our method is still very significant, as the EBMT preprocessor is able to translate 17.91% and 9.90% of the tokens in the in-domain and out-of-domain test sets for Europarl, respectively. As for the distribution of these partial translations, we observe that most of the translations in IVAP come from the traditional TM behavior of our preprocessor[4], but the relative contribution of the generalization mechanisms gets considerably higher as the distance between the input text and the training corpus increases[5].

As far as the quality of the partial translations is concerned, Tables 4 and 5 show the results of the manual evaluation we carried out for both in-domain test sets. The overall results are very positive in both cases, with an average score of 3.45 and 3.39 out of 4 for IVAP and Europarl, respectively. In spite of the average scores being similar, it is worth mentioning that there is a considerable difference in the variance of the evaluations, with Europarl obtaining much more coherent scores than IVAP (3.30-3.49 range for Europarl and 3.02-3.73 range for IVAP). We believe

---

[4]69.16% of the tokens translated by the EBMT preprocessor when using all the generalization mechanisms correspond to full sentences (18,284 out of 26,436 as shown in Table 3)

[5]Only 3.45% and 0.22% of the tokens translated by the EBMT preprocessor when using all the generalization mechanisms correspond to full sentences in Europarl and News commentary, respectively (379 out of 10,986 and 12 tokens out of 5,566 as shown in Table 3)

	1	2	3	4	Average
Evaluator 1	2 (1.56%)	5 (3.91%)	19 (14.84%)	102 (79.69%)	3.73
Evaluator 2	5 (3.91%)	4 (3.13%)	18 (14.06%)	101 (78.91%)	3.68
Evaluator 3	11 (8.59%)	8 (6.25%)	9 (7.03%)	100 (78.13%)	3.55
Evaluator 4	13 (10.16%)	14 (10.94%)	25 (19.53%)	76 (59.38%)	3.28
Evaluator 5	19 (14.96%)	23 (18.11%)	21 (16.54%)	64 (50.39%)	3.02
Average	10 (7.82%)	10.8 (8.45%)	18.4 (14.4%)	88.6 (69.33%)	3.45

Table 4: Results of the manual evaluation in IVAP (es-eu)

	1	2	3	4	Average
Evaluator 1	8 (4.79%)	11 (6.59%)	40 (23.95%)	108 (64.67%)	3.49
Evaluator 2	14 (8.38%)	11 (6.59%)	28 (16.77%)	114 (68.26%)	3.45
Evaluator 3	11 (6.71%)	20 (12.2%)	25 (15.25%)	108 (65.85%)	3.40
Evaluator 4	16 (9.58%)	14 (8.38%)	38 (22.75%)	99 (59.28%)	3.32
Evaluator 5	17 (10.24%)	20 (12.05%)	25 (15.06%)	104 (62.65%)	3.30
Average	13.2 (7.94%)	15.2 (9.15%)	31.2 (18.77%)	106.6 (64.14%)	3.39

Table 5: Results of the manual evaluation in Europarl (es-en)

	RBMT baseline	RBMT + full sentences	RBMT + full sentences with NE	RBMT + chunks with NE (Berkeley HMM)
IVAP	0.0498	0.3350	0.3330	0.3168
Europarl	0.1755	0.1786	0.1790	0.1983
News commentary	0.2173	0.2173	0.2173	0.2227

Table 6: BLEU scores with RBMT hybridization

Source	Finalmente, **Señorías**, los medios de comunicacin deben **jugar también un papel importante** en esta tarea.
Baseline	Finally, **Señorías**, the media have to **play also an important paper** in this task.
System	Finally, **ladies and gentlemen**, the media have to **play an important role too** in this task.
Reference	Finally, **ladies and gentlemen**, the media must **also play an important role** in this task.

Table 7: An example of RBMT hybridization in Europarl

that the reason behind that is the unfamiliarity of some evaluators with machine translation and the register used for legal documents in Basque, which could have made them penalize minor mistakes that were sometimes even found in the reference translations too severely[6]. As a matter of fact, some full sentence translations that were equal to the reference ones got 1 and 2 scores. In any case, the reported results reflect that our EBMT preprocessor produces high-quality partial translations, with less than 20% of them obtaining a negative (1 or 2) score in average for both test sets.

4.2 RBMT hybridization

Table 6 shows the BLEU scores obtained when hybridizing with RBMT translators. As it can be seen, we obtain very good results, with our system outperforming the baseline in all the test sets. The gain in BLEU is particularly remarkable in the case of IVAP, with an improvement of 26.7 points, but still notable for the other more standard in-domain and out-of-domain test sets, with an improvement of 2.28 and 0.54 points, respectively.

As far as the contribution of each generalization

step is concerned, it can be observed that, in the case of IVAP, all the improvement comes from the TM behavior of our preprocessor, and the generalization steps themselves have a negative impact. We believe that this is due to an integration problem with Matxin, as we find that it often misplaces our XML tags in its translations, yielding to senseless replacements that have a negative impact in the overall translation quality. In the case of both Apertium test sets, which do not suffer from this problem, the generalization steps work as expected and, in fact, practically all the improvement comes from them. Table 7 shows one such case, where the proposed system is able to properly translate the out-of-vocabulary word "Señorías" and the idiomatic expression "jugar un papel importante" unlike the baseline.

4.3 SMT hybridization

The BLEU scores obtained with SMT hybridization are shown in Table 8. As it can be seen, our system is not able to beat the baseline for either of the Spanish-English test sets, although there are instances in which the hybrid system gives better results as it is the case of the example in Table 9. We think that, as shown in Table 10, the reason behind

---
[6]Note that not all the evaluators for both test sets were the same

	SMT baseline	SMT + full sentences	SMT + full sentences with NE	SMT + chunks with NE (Berkeley HMM)
IVAP	0.3368	0.4483	0.4472	0.4593
Europarl	0.3307	0.3307	0.3304	0.3251
News commentary	0.2984	0.2982	0.2982	0.2967

Table 8: BLEU scores with SMT hybridization

Source	**De ser así, se comete un error**, ya que se trata de la credibilidad y fiabilidad que tiene la Unión Europea [...]
Baseline	**For example, we are making a mistake**, because that is the credibility and reliability of the European Union [...]
System	**If that is the case, it is a mistake**, because that is the credibility and reliability of the European Union [...]
Reference	**If it were to be the case then it is a miscalculation** because this is about the credibility and reliability of the European Union [...]

Table 9: An example of SMT hybridization in Europarl

	Full sentences	Full sentences with NE	Chunks with NE
IVAP	15.10	11.31	8.09
Europarl	7.02	9.39	5.10
News commentary	6.00	-	4.74

Table 10: Average length of the fragments translated by the EBMT preprocessor

that is that the fragments translated by the EBMT preprocessor are too short for these test sets, as the baseline SMT system would be able to properly handle this size n-grams. Increasing the minimum number of tokens k to be searched by the EBMT preprocessor as discussed in Section 2.2.1 would solve this problem, but it would also decrease its coverage, considerably reducing the impact of the whole system.

Nevertheless, we obtain very good results in IVAP, where we achieve an overall improvement of 12.25 BLEU points from which 1.1 come from the generalization steps. We therefore conclude that our system works with SMT hybridization as long as the domain is repetitive enough to reuse long text chunks that traditional SMT systems are not able to handle effectively.

5 Conclusions and future work

In summary, this paper develops a generic multi-pass hybridization method based on an EBMT preprocessor that creates partial translations making use of NE and chunk generalization. The effectiveness of the preprocessor is experimentally demonstrated both in terms of coverage and translation quality. Furthermore, our experiments show that the proposed method considerably improves the baseline with RBMT hybridization, and we also obtain very good results with SMT hybridization in repetitive enough domains.

In the future, we intend to further optimize our system by using heuristics to detect wrong alignments, improve our processing for Spanish contractions, which often led to parsing errors, and introduce a better handling for NEs with common nouns, which were incorrectly left unchanged when not found in any dictionary. In addition, we plan to improve SMT integration by increasing the minimum number of tokens to be translated by the EBMT preprocessor and optimizing the weight assigned to our partial translations. We also want to explore the possibility of selecting more than one translation for each chunk that would then compete with each other and the rest of the entries in the phrase table. Furthermore, we would like to fix the integration problems with Matxin and use a full syntactic analyzer for Basque. We also intend to try more metrics to better understand the behavior of the whole system. Lastly, we plan to release our system as an open source project.

Acknowledgments

References

Alegria, Iñaki, Arantza Casillas, Arantza Díaz De Ilarraza, Jon Igartua, Gorka Labaka, Mikel Lersundi, Aingeru Mayor, Kepa Sarasola, Xabier Saralegi, and B Laskurain. 2008. Mixing Approaches to MT for Basque: Selecting the best output from RBMT, EBMT and SMT. *MATMT 2008: Mixing Approaches to Machine Translation*, pages 27–34.

Dandapat, Sandipan, Sara Morrissey, Andy Way, and Mikel L Forcada. 2011. Using example-based MT

to support statistical MT when translating homogeneous data in a resource-poor setting. In *Proceedings of the 15th annual meeting of the European Association for Machine Translation (EAMT 2011)*, pages 201–208.

Dandapat, Sandipan, Sara Morrissey, Andy Way, and Joseph van Genabith. 2012. Combining EBMT, SMT, TM and IR technologies for quality and scale. In *Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*, pages 48–58. Association for Computational Linguistics.

Dandapat, Sandipan. 2012. *Mitigating the Problems of SMT using EBMT*. Ph.D. thesis, Dublin City University.

DeNero, John and Dan Klein. 2007. Tailoring Word Alignments to Syntactic Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 17–24, Prague, Czech Republic, June. Association for Computational Linguistics.

Ezeiza, Nerea, Iñaki Alegria, José María Arriola, Rubén Urizar, and Itziar Aduriz. 1998. Combining stochastic and rule-based methods for disambiguation in agglutinative languages. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 380–384. Association for Computational Linguistics.

Forcada, Mikel L, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim ORegan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, and Francis M Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144.

Gotti, Fabrizio, Philippe Langlais, Elliott Macklovitch, Didier Bourigault, Benoit Robichaud, and Claude Coulombe. 2005. 3GTM: A third-generation translation memory. In *Proceedings of the 3rd Computational Linguistics in the North-East Workshop*, pages 8–15.

Groves, Declan and Andy Way. 2005. Hybrid example-based SMT: the best of both worlds? In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pages 183–190. Association for Computational Linguistics.

Hutchins, John. 2007. Machine translation: A concise history. *Computer aided translation: Theory and practice*.

Koehn, Philipp and Jean Senellart. 2010. Convergence of translation memory and statistical machine translation. In *Proceedings of AMTA Workshop on MT Research and the Translation Industry*, pages 21–31.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180. Association for Computational Linguistics.

Liang, Percy, Ben Taskar, and Dan Klein. 2006. Alignment by agreement. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 104–111. Association for Computational Linguistics.

Manber, Udi and Gene Myers. 1990. Suffix Arrays: A New Method for On-line String Searches. In *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '90, pages 319–327, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.

Mayor, Aingeru, Iñaki Alegria, Arantza Díaz De Ilarraza, Gorka Labaka, Mikel Lersundi, and Kepa Sarasola. 2011. Matxin, an open-source rule-based machine translation system for Basque. *Machine translation*, 25(1):53–82.

Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

Padró, Lluís and Evgeny Stanilovsky. 2012. FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Sánchez-Martınez, Felipe, Mikel L Forcada, and Andy Way. 2009. Hybrid rule-based-example-based MT: feeding Apertium with sub-sentential translation units. In *3rd International Workshop on Example-Based Machine Translation*, page 11. Citeseer.

Shirai, Satoshi, Francis Bond, and Yamato Takahashi. 1997. A hybrid rule and example-based method for machine translation. In *Proceedings of NLPRS*, volume 97, pages 49–54. Citeseer.

Socher, Richard, John Bauer, Christopher D Manning, and Andrew Y Ng. 2013. Parsing with compositional vector grammars. In *In Proceedings of the ACL conference*. Citeseer.