Hybrid Machine Translation Guided by a Rule–Based System

Cristina España-Bonet,
Lluís Màrquez
Universitat Politècnica de Catalunya
cristinae@lsi.upc.edu,
lluism@lsi.upc.edu

Gorka Labaka, Arantza Díaz de Ilarraza,
Kepa Sarasola
University of the Basque Country
gorka.labaka@ehu.es,
jipdisaa@ehu.es, kepa.sarasola@ehu.es

Abstract

This paper presents a machine translation architecture which hybridizes Matxin, a rule-based system, with regular phrase-based Statistical Machine Translation. In short, the hybrid translation process is guided by the rule-based engine and, before transference, a set of partial candidate translations provided by SMT subsystems is used to enrich the tree-based representation. The final hybrid translation is created by choosing the most probable combination among the available fragments with a statistical decoder in a monotonic way. We have applied the hybrid model to a pair of distant languages, Spanish and Basque, and according to our evaluation (both automatic and manual) the hybrid approach significantly outperforms the best SMT system on out-of-domain data.
Keywords: hybrid MT models, RBMT, phrase-based SMT, Spanish-Basque MT

1 Introduction

It is well known that rule-based and phrase-based statistical machine translation paradigms (RBMT and SMT, respectively) have complementary strengths and weaknesses. First, RBMT systems tend to produce syntactically better translations and deal with long distance dependencies, agreement and constituent reordering in a better way, since they perform the analysis, transfer and generation steps based on syntactic principles. On the bad side, they usually have problems with lexical selection due to a poor handling of word ambiguity. Also, in cases in which the input sentence has an unexpected syntactic structure, the parser may fail and the quality of the translation decrease dramatically.

On the other side, phrase-based SMT models usually do a better job with lexical selection and general fluency, since they model lexical choice with distributional criteria and explicit probabilistic language models. However, SMT systems usually generate structurally worse translations, since they model translation more locally and have problems with long distance reordering. They also tend to produce very obvious errors, which are annoying for regular users, e.g., lack of gender and number agreement, bad punctuation, etc. Moreover, the SMT systems can experience a severe degradation of performance when applied to corpora different from those used for training (*out-of-domain* evaluation).

In this work we present a hybrid architecture that tries to get the best of both worlds. Contrary to many of the previous works on MT hybridization (HMT), we do not make a posterior combination of the output of both systems, nor use the RBMT system to enrich the translation table of the SMT system. Instead, we use Matxin (Alegria et al., 2007), a RBMT system, for guiding the main translation steps and an SMT system to feed the transferred syntactic structure with several translation options at different levels of granularity, just before the transference step. The creation of final translation implies the selection of a subset of translated units, and it is performed by a monotonic statistical decoder. This is why we refer to the system as 'Statistical Matxin Translator', or SMatxinT in short.

The main idea is that the RBMT system should perform parsing and rule-based transfer and reordering to produce a good structure for the output, while SMT helps the lexical selection by providing multiple translation suggestions for the pieces of the source language corresponding to the tree constituents. The final decoding accounts also for fluency by using language models, and can be monotonic (and so, fast) because the structure has been already decided by the RBMT component.

As a proof of concept we have instantiated and applied the SMatxinT architecture to a pair of structurally and morphologically distant languages, Spanish and Basque. The results obtained on several benchmark corpora show that the hybrid approach is able to significantly improve the out-of-domain results of the best individual SMT system in terms of BLEU (Papineni et al., 2002) and TER (Snover et al., 2006) scores. A manual evaluation has been performed on a set of 100 samples from the test set verifying the significant advantage of the SMatxinT hybrid system. More detailed analyses reveal that all the components of the hybrid system play an important role in the system (i.e., RBMT structural translation, SMT translation candidates and RBMT original translation). We think that the improvement obtained is remarkable given the simple statistical decoding process implemented so far. Indeed, the upper bound performance for the hybrid method calculated with the current setting reveals that there is still a large room for improvement.

One issue that has not been addressed by SMatxinT is the strong dependence on the syntactic analysis of the source sentence. So, it still suffers from one of the problems of the RBMT systems. We plan to address this problem in the future work.

The rest of the paper is organized as follows. Section 2 overviews some related work on RBMT-SMT system hybridization. Section 3 presents the SMatxinT architecture describing in detail all its various components. Section 4 describes the experimental setting, and discusses the results obtained. Finally, Section 5 concludes and outlines future research lines.

2 Related Work

System combination, either serial or by a posterior combination of systems' outputs, is a first step towards hybridization. Although it has been shown to help in improving translation quality, the combination does not represent a real hybridization since systems do not interact among them (see Thurmair (2009) for a classification of HMT architectures).

In the case of actual interdependences, one of the systems in action leads the translation process and the other ones strengthen it. Much work has been done in building systems where the statistical component is in charge of the translation and the companion system provides complementary information. For instance, Eisele et al. (2008) and Chen and Eisele (2010) introduce lexical information coming from a rule-based translator into an SMT system, in the form of new phrase pairs for the translation table. In both cases results are positive on out-of-domain tests.

The opposite direction, that is, where the RBMT system leads the translation and the SMT system provides complementary information, has been less explored. Habash et al. (2009) enrich the dictionary of a RBMT system with phrases from an SMT system. Federmann et al. (2010) use the translations obtained with a RBMT system and substitute selected noun phrases by their SMT counterparts. Globally, their results improve the individual systems when the hybrid system is applied to translate into languages with a richer morphology than the source.

Similar in spirit to Federmann et al. (2010), translations given by SMatxinT are controlled by the RBMT system in a way that will be clarified in the following sections, but SMatxinT is enriched with a wider variety of SMT translation options.

3 A Hybrid MT Model Guided by RBMT

3.1 Individual MT Systems

Our hybrid model builds on three individual machine translation systems, a rule-based Spanish-Basque system and two variants of regular phrase based statistical MT systems. These three subsystems are described below.

SMT basic system (SMTb) The development of the baseline system was carried out using available state-of-the-art tools: GIZA++ toolkit (Och, 2003), SRILM toolkit (Stolcke, 2002) and Moses Decoder (Koehn et al., 2007). More particularly, we used a log-linear combination of several common feature functions: phrase translation probabilities (in both directions), word-based translation probabilities (lexicon model, in both directions), a phrase length penalty and the target language model. The language model is a simple 3-gram language model

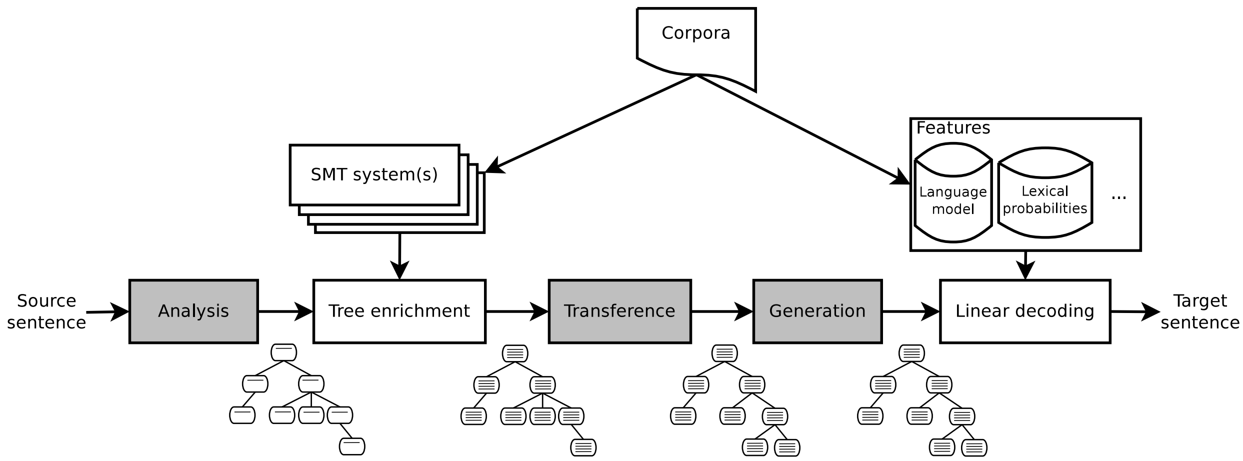

Figure 1: General architecture of SMatxinT. The RBMT modules which guide the MT process are the grey boxes

with modified Kneser-Ney smoothing. We also used a lexical reordering model ('msd-bidirectional-fe' training option). Parameter optimization was done following the usual practice, i.e., Minimum-Error-Rate Training over the BLEU measure.

Morpheme-based SMT system (SMTm) A second variant of the SMT system was used to address the rich morphology of Basque. In this system, words are split into several morphemes by using a Basque morphological analyzer/lemmatizer, aiming at reducing the sparseness produced by the agglutinative nature of Basque and the small amount of parallel corpora. Adapting the baseline system to work at the morpheme level mainly consists of training Moses on the segmented text (the exact training options of SMTb are used for SMTm). The SMT system trained on segmented words will generate a sequence of morphemes. So, in order to obtain the final Basque text from the segmented output, a word-generation post-process is applied. Details on this system can be found in (Labaka, 2010).

Rule-based system (Matxin) Matxin is an open-source Spanish-Basque RBMT engine (Alegria et al., 2007), following the traditional transfer model. Matxin consists of three main components: 1) analysis of the source sentence into a dependency tree structure; 2) transfer from the source language dependency tree to a target language dependency structure; and 3) generation of the output translation from the target dependency structure.

Matxin reuses several open tools and it is based on an unique XML format for the flow between the different modules, which makes easier the interaction among different developers of tools and resources. The result is an open source software which can be downloaded from matxin.sourceforge.net, and it has an on-line demo[1] available since 2006.

3.2 Architecture of SMatxinT

The SMatxinT architecture is based on the three following principles: 1) generally, the final translation should be based on RBMT system's syntactic rearrangements; 2) the hybrid system must have the chance of using SMT-based local translations to improve lexical selection; and 3) it should be able to recover from potential problems encountered in the analysis, using longer SMT translations.

Following principle 1, SMatxinT adopts the architecture and data structures from Matxin, the above described RBMT system. The internal representation is a dependency parse tree, where the boundaries of each phrase are marked. In order to add hybrid functionality we introduce two new modules to the RBMT architecture: 1) *tree enrichment*, which incorporates SMT additional translations to each phrase of the syntactic tree, and 2) *linear decoder*, which is responsible for generating the final translation by selecting among RBMT and SMT partial translation candidates from the enriched tree. Figure 1 shows the general architecture of the SMatxinT system. The two genuinely new modules are described in the following subsections.

[1] `http://www.opentrad.com`

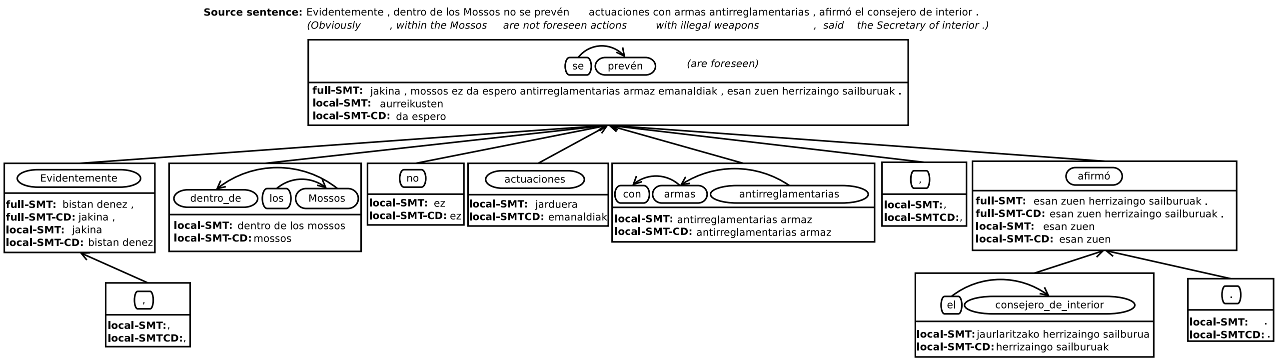

Figure 2: Example of the analysis tree enriched with SMT translations.

3.3 Tree enrichment

After syntactic analysis and before the transfer module modifies the tree, the tree enrichment module uses one (or several) SMT systems to assign complementary translations to each phrase. These partial translations can be used by the linear decoder at the end of the translation process instead of the translation created by the RBMT system. For this process, we relied on the phrase segmentation created by the RBMT analyzer and incorporated, for each phrase, two types of translations: 1) the SMT translation of that phrase, and 2) the translation of the entire subtree dependent on that phrase. For example, by looking at Figure 2, one sees the enriched analysis where this module incorporated the translation for the phrase itself (*local-SMT*) and the translation of the entire subtree dependent on it (*full-SMT*).

Thus, we intend to satisfy the conditions that we set in the design of the hybrid translator. In its expected behavior, the system should maintain the constituents in the order specified by the RBMT system, but choosing for each phrase the most adequate translation (RBMT or SMT). Besides, the hybrid system will have the chance of using longer pure SMT translations, avoiding potential analysis errors. As a limit case, the hybrid system uses the full sentence's SMT translation, since this is included as an alternative of the subtree dependent to the root node.

It seems interesting to discuss the appropriateness of the length of the fragments used to generate SMT alternative translations. In our design, the tree is enriched with statistical translations of the phrases and combination of phrases marked by the analyzer. In principle, the statistical translator might not have enough context to create useful translations, and the use of smaller translations would increase the complexity of the system without improving the results. Similarly, one can argue that in many cases, a phrase may also be too short for the statistical translator. So, in order to address these limitations, we also extract local translations from longer SMT translations. Those context discriminated phrases are marked as *CD* in the example in Figure 2. In that example, we can see how the local CD translation of the phrase 'se prevén' (**local-SMT-CD**=*'da espero'*) does not match with the one achieved by translating the phrase in isolation (**local-SMT**=*'aurreikusten'*), since the sentence is negative and the verb chain must be adapted.

3.4 Linear decoding

Once the alternative SMT translations are incorporated, the RBMT translation process continues as usual (we had to slightly modify the transfer and generation modules to keep the SMT translations incorporated by the previous module). Once the generation process is finished, we have a tree structure where the order of the different phrases is defined, but with more than one possible translation for each phrase. At this point, we need a new module to decide what translation use for each phrase.

This decision process is similar to the search process conducted by an SMT decoder, but simplified, since rearrangements are not allowed. Although we initially considered the possibility of implementing our own search engine, we finally decided to use an already available statistical decoder (Moses) for the monotone search.

The approach here is quite different to that of Federmann et al. (2010). We do not substitute possible bad chunk translations from the RBMT system by its

SMT counterpart, but let a decoder choose among all the available options. The election will be better the more informed the decoder is. Besides, we do not face all the problems associated to the alignment between systems because, as said, the tree is enriched with phrases that are obtained by running an SMT system for each of the segments (or subtrees) given by the RBMT system. We also use context discriminated translations which are extracted from larger translations but always within the same system.

3.5 Features of the decoder

We have defined the following features to guide the decoder in selecting the fragments from the enriched tree to construct the best target output.

Usual SMT features

Language Model (LM): Same n-gram based Language Model used in the SMT systems.

Word Penalty (WP): Count of words used in the translation.

Phrase Penalty (PP): Count of phrases used in the translation.

Source/Consensus features

Counter ($1...n$): Indicates how many different systems generated this phrase.

SMT (1, e): Indicates whether the phrase has been generated by an SMT system or not.

RBMT (1, e): Indicates whether the phrase has been generated by the RBMT system or not.

BOTH ($e^{\#}$): where # is the number of source words covered by those phrases generated by both MT paradigm systems (SMT and RBMT).

Lexical probability features

Corpus Lexical Probability (both directions): This feature is based on the lexical probability commonly used in statistical decoding, but with a modification aimed to face the unknown alignments. Our hybrid system relies on the internal alignments created by different individual systems to calculate this feature. So, since each system uses different data sources to generate the translation, these alignments are not directly compatible with the IBM-1 probabilities and we have to define a mechanism for dealing with unknown alignments. The morpheme-based SMT and the RBMT systems generate alignments that are not present in the training corpora (they both tend to associate the Spanish preposition with the Basque inflected word). Those alignments that where not present in the corpus are ignored, and for those words which all their alignments are ignored, the probability assigned to the NULL alignment is used. Unknown words that would not be present in the IBM-1 probabilities table use a default NULL alignment probability (10^{-10}).

Dictionary Lexical Probability (both directions): Lexical probability inferred from the dictionary. This lexical feature uses the previous mechanism to deal with unknown alignments, but instead of using corpus probabilities, probabilities extracted from the RBMT dictionary are used. The dictionary organize equivalences in senses and duplicates the translations that can be used for more than one sense. We define the dictionary lexical probability according to those repetitions. Moreover, in order to model the RBMT system preferences, the first equivalence in the dictionary (the one that the RBMT system predominantly uses) is assigned with a double probability than other alternative translations.

4 Experiments and Results

4.1 Corpora

In this work we have used a heterogeneous bilingual corpus including four aligned corpora in Basque and Spanish: six reference books, a collection of articles, and translation memories with administrative and TV program descriptions. That makes a total of 491,853 sentences with 7,966,419 tokens in Spanish and 6,062,911 tokens in Basque. It is worth noting that the bilingual corpus is rather small, so we could expect significant sparseness in pure statistical approaches. The number of tokens on the Basque side is much lower due to the rich morphology and the agglutinative nature of the language.

The training corpus is basically made up of administrative documents and descriptions of TV programs. For development and testing we extracted some administrative data for the *in-domain* evaluation and selected two collections of news for the *out-of-domain* study, totaling four sets:

Elhuyardevel and *Elhuyartest*: 1,500 segments each, extracted from the administrative documents.

EITBtest: 1,500 sentences with one reference translation extracted from the news collection of the

		Elhuyar		EITB		NEWS	
		BLEU	**TER**	**BLEU**	**TER**	**BLEU**	**TER**
Individual systems	**Matxin**	5.10	83.32	5.77	87.73	11.72	82.04
	SMTb	**14.96**	70.20	8.03	83.27	14.74	78.63
	SMTm	13.71	71.64	7.64	85.59	14.58	78.90
Control system	**Google**	7.32	78.43	6.73	86.32	12.01	81.84
Oracles	**SMatxinT$_0$**	19.42	62.21	11.96	76.64	20.71	73.36
	SMatxinT	20.35	60.47	13.23	74.20	23.16	71.25
Hybrid systems	**SMatxinT$_0$**	14.50	69.73	8.45	82.17	14.90	77.29
	SMatxinT	14.73	**69.18**	**8.81**	**81.33**	**15.31**	**76.54**

Table 1: BLEU and TER results of all individual and hybrid systems (including oracles for the latter)

Origin system	SMatxinT		sBLEU Oracle	
	chunks	**tokens**	**chunks**	**tokens**
SMT	2,682 (44.2%)	11,391 (65.4%)	3,202 (38.4%)	9,043 (51.2%)
SMT-CD	523 (8.6%)	1,737 (10.0%)	779 (9.3%)	1,890 (10.7%)
RBMT	401 (6.6%)	1,279 (7.3%)	969 (11.6%)	2,554 (14.4%)
BOTH	2,454 (40.5%)	3,013 (17.3%)	3,389 (40.6%)	4,192 (23.7%)
Total	6,060 (100%)	17,420 (100%)	8,339 (100%)	17,679 (100%)

Table 2: Number of chunks and lexical tokens coming from each of the single MT systems observed in the translations. The first block shows the distribution resulting from SMatxinT; the second block refers to the Oracle

Basque News and Information Channel[2].

NEWStest: 1,000 sentences collected from Spanish newspapers with two references.

Additionally, we collected a 21 million word monolingual corpus, which together with the Basque side of the parallel bilingual corpora, builds up a 28 million word corpus to train the language model.

4.2 Systems

We have evaluated and compared six different systems. The first three correspond to the individual systems described in Section 3.1 (*SMTb*, *SMTm* and *Matxin*), which are used as baselines.

Regarding the hybrid engine, we consider two different versions: *SMatxinT*$_0$, in which Matxin is used to analyze the source and reorder the target, but the lexical realization of the translation is done by using only phrases supplied by the statistical systems (both regular and context-aware); 2) *SMatxinT*, the full hybrid system, in which also the RBMT translations are available to the linear decoder.

Finally, we evaluate also the freely available Google translate[3], which has the Spanish–Basque pair available and allows us to control the quality of the rest of the systems.

[2] `http://www.eitb24.com/en`

[3] `http://translate.google.com/`

4.3 Results and Discussion

Table 1 presents all the results along systems and corpora, using BLEU and TER scores (best results per corpora highlighted in boldface). Regarding the individual systems, we can observe that for in-domain tests (Elhuyar corpus), the statistical systems SMTb and SMTm clearly outperform the rule-based system Matxin (by 8-9 BLEU points and 12-13 TER points), while in the out-of-domain tests, this difference reduces notably (to 1-3 BLEU points and 3-4 TER points). Actually, the performance of Matxin remains quite stable across corpora, while the SMT systems generalize badly to out-of-domain corpora, showing a significant drop in performance. This is explained mainly due to a lower coverage of the test lexicon.[4]

In absolute terms, BLEU and TER scores are quite low for the translation tasks. The reason is the large differences between the two languages, and also the use of a lexical evaluation metric, such as BLEU, on a target language that is morphologically rich and highly agglutinative. Note that the phrase-based Google SMT translator obtains comparable scores to Matxin and lower than our SMT systems,

[4] This overfitting to the training corpus, which leads to slightly over-estimate the performance of statistical systems, is a well known phenomenon in the MT literature.

even in the out-of-domain corpora, indicating that the results of our systems are state-of-the-art.

Regarding the hybrid systems, there are two clear scenarios. In the *in-domain* evaluation, results of the hybrid system are comparable but not better than the SMT subsystems (BLEU and TER metrics differ in their preferences). This is explained due to the comparatively much lower performance of Matxin with respect to the SMT systems in the in-domain test set. However, in the *out-of-domain* tests, SMatxinT and SMatxinT_0 are consistently better than all the single engines. For EITBtest and NEWStest, the improvement of SMatxinT with respect to the best individual system is of 0.8 and 0.6 BLEU points and of 2 and 2.1 TER points, respectively. All these differences are statistically significant according to paired bootstrap resampling test (Koehn, 2004). It is also worth noting that, in all cases, the results of SMatxinT are slightly better than those of SMatxinT_0, indicating that not only the syntactic rearrangements of the RBMT system are important but also its partial translations. Overall, these results are consistent with those of Federmann et al. (2010). A direct comparison is not possible because the language pairs involved differ, but in both papers the hybrid system outperforms the SMT one when translating into a morphologically richer language.

Table 2 shows the proportion of chunks and words coming from the different individual translation systems that are present in the SMatxinT translations. As it can be seen, only a 6.6% of the chunks come exclusively from the RBMT translation. A large percentage (40.5%, mainly punctuation and conjunctions) come from both, and the largest percentage corresponds to chunks supplied exclusively from the SMT systems (52.8%). So, the RBMT system segments and rearranges the sentence, but the lexicon is mostly taken from the SMT system. This is explaining the small difference in scores between SMatxinT and SMatxinT_0, although we have already seen that this small 6.6% makes a positive difference.

This tendency is confirmed by an inspection of the best attainable translations with the hybrid system (right hand side of Table 2). This *oracle* translation is approximated by picking the highest sBLEU[5] translation from the $10,000$-best translations provided by the decoder. The oracle has still a majority of SMT chunks, but the difference is an increment of a 5 points in the percentage of fragments coming from Matxin, giving some more credit to the RBMT translations. Another important difference is that the oracle is finer grained and uses shorter chunks to construct the best solution (especially with the fragments coming only from SMT). However, this difference is compensated by a larger number of chunks for a very similar final translation average length. With all these differences, the oracle achieves much higher scores (at least 5 points improvement in both metrics) evincing that there is still a large room for improvement in the hybrid system.

[5]Smoothed BLEU score at the sentence level.

Assessments	SMT	Tied	SMatxinT
All	45 (22.5%)	64 (32.0%)	**91 (45.5%)**
Agreement	13 (21.3%)	17 (27.9%)	**31 (50.8%)**

Table 3: Manual evaluation for a random subset of 100 sentences of NEWStest.

Finally, we also conducted a manual evaluation on 100 sentences randomly selected from NEWStest (with lengths between 10 and 40 words). Four native speakers evaluated 50 translations each, so that every sentence has two assessments comparing SMatxinT with the best SMT system. Results confirm the outcome of the automatic evaluation. The annotators agreed on the evaluation of 61 out of the 100 sentences, being SMatxinT the best system in 31 cases, SMTm the best one in 13 and both systems tied in 17 cases. That is, SMatxinT is the preferred system 50.8% of the times. Considering all 200 assessments together the percentage is 45.5%. Concrete values can be read in Table 3.

5 Conclusions

In this paper we have presented SMatxinT, a hybrid machine translation system that combines RBMT with phrase-based SMT. The RBMT system Matxin leads the translation process and generates the syntactic structure in the target language. The SMT system generates multiple candidate translations of any fragment in this tree, and a posterior linear decoder selects the best combination to create the final output. In this way, the SMT subsystem contributes to alleviate one of the main weaknesses of RBMT systems, that is lexical selection. Besides, the fact that the decoder works with translation options at all the levels provides some robustness to the system. The final decoder has available SMT translations that cover groups of RB-chunks or even the full translation of the source sentence. Those could be useful in case that the order of the target given by Matxin is not the correct one.

SMatxinT achieves statistically significant improvements on out-of-domain test sets for the Spanish-to-Basque language pair, according to BLEU and TER evaluation metrics. This advantage has been corroborated by a manual evaluation conducted on a set of 100 samples. Although this first version of SMatxinT is already obtaining significantly better results, the analysis of the oracles shows that there is still a large room for improvement. Oracle translations tend to be composed by more and shorter chunks, and a larger proportion of chunks coming from Matxin. So, a plausible way to enhance the system would be to define new features for the linear decoder that take this fact into account. In general, we think that decoding could be also improved with more linguistically based features.

Another line of work to alleviate the strong dependence of SMatxinT on the initial syntactic parsing is the incorporation of basic information into the system such as multiple syntactic trees from the side of the rule based system and n-best translations from the SMT subsystems (note that other available SMT systems could be incorporated as well). Also important is a more detailed manual comparison of the outputs of the different systems, in particular to document improvements in sentence structure vs word choice over baseline systems. That would allow to better understand the strong and weak points of the hybrid system. Finally, broadening the study to other language pairs would also contribute to corroborate the value of the proposed architecture.

Acknowledgements

This work has been partially funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 247914 (MOLTO project, FP7-ICT-2009-4-247914) and the Spanish Ministry of Science and Innovation (OpenMT-2 project, TIN2009-14675-C03-01).

References

I. Alegria, A. Díaz de Ilarraza, G. Labaka, M. Lersundi, A. Mayor, and K. Sarasola. 2007. Transfer-based MT from spanish into basque: Reusability, standardization and open source. *Lecture Notes in Computer Science*, 4394:374–384.

Y. Chen and A. Eisele. 2010. Hierarchical hybrid translation between english and german. In Viggo Hansen and Francois Yvon, editors, *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, pages 90–97. EAMT, EAMT, 5.

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*, StatMT '08, pages 179–182.

C. Federmann, A. Eisele, Y. Chen, S. Hunsicker, J. Xu, and H. Uszkoreit. 2010. Further experiments with shallow hybrid mt systems. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 77–81, July.

N. Habash, B. Dorr, and C. Monz. 2009. Symbolic-to-statistical hybridization: extending generation-heavy machine translation. *Machine Translation*, 23:23–63.

P. Koehn, 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 *Annual Meeting of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic, June.

P. Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, Barcelona, Spain.

G. Labaka. 2010. *EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language. Its use in SMT-RBMT-EBMT hybridation*. Ph.D. thesis, University of the Basque Country.

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

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 Association for Computational Linguistics*, pages 311–318.

M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of AMTA*, pages 223–231.

A. Stolcke. 2002. SRILM - An Extensible Language Modeling Toolkit. In *Proceedings of the International Conference of Spoken Language Processing*, volume 2, pages 901–904.

G. Thurmair. 2009. Comparing different architectures of hybrid machine translation systems. In *Proc MT Summit XII*.