

An English-to-Hungarian Morpheme-based Statistical Machine Translation System with Reordering Rules

László J. Laki, Attila Novák, Borbála Siklósi

MTA-PPKE Language Technology Research Group

Pázmány Péter Catholic University, Faculty of Information Technology

50/a Práter Street, 1083 Budapest, Hungary

{surname.firstname}@itk.ppke.hu

Abstract

Phrase-based statistical machine translation systems can generate translations of reasonable quality in the case of language pairs with similar structure and word order. However, if the languages are more distant from a grammatical point of view, the quality of translations is much behind the expectations, since the baseline translation system cannot cope with long distance reordering of words and the mapping of word internal grammatical structures. In our paper, we present a method that tries to overcome these problems in the case of English-Hungarian translation by applying reordering rules prior to the translation process and by creating morpheme-based and factored models. Although automatic evaluation scores do not reliably reflect the improvement in all cases, human evaluation of our systems shows that readability and accuracy of the translations were improved both by reordering and applying richer models.

1 Introduction

Phrase-based statistical machine translation systems rely on statistical observations derived from phrase alignments automatically extracted from parallel bilingual corpora. The main advantage of applying SMT is its language-independence. The phrase-based model works well for language pairs with similar syntactic structure and word order.

However, phrase-based models fail to handle great word-order differences adequately. We describe our attempt to improve performance by transforming source language (English) sentences to a structure similar to that of the corresponding target (Hungarian) sentence. We also describe our approach for handling data sparseness due to the

inadequate coverage of linguistic structures by the limited training corpus. It is a common problem in the case of translation to agglutinating languages like Hungarian, where a much greater amount of training data would be necessary to provide adequate statistics than what is necessary for closely related language pairs involving only morphologically less complex languages.

2 Machine Translation from English to Hungarian

English and Hungarian are rather distant regarding morphological and syntactic structure and word order. Hungarian, like Finnish or Turkish, is an agglutinating and compounding language, which morphological processes yield a huge number of different word forms. This, combined with free word order of main grammatical constituents and systematically different word order in NP's and PP's, results in poor performance of traditional phrase-based SMT systems. In order to have an SMT system produce correct translations of high quality, it is required to have a relevant statistical model acquired from bilingual corpora. Thus, even if a corpus of a substantial size were available (which is not the case), both the alignment phase of constructing a translation model and translation itself would be compromised by the high number of seldom or never seen word forms.

3 Related work

For language pairs having very different syntactic structure and word order, research has shifted towards using hierarchical models or the use of hybrid methods, such as augmenting purely statistical approaches by handmade rules as a preprocessing step. Such extensions have proved to improve results significantly in systems translating from English to German, Arabic or Turkish and several other languages (Yeniterzi and Oflazer, 2010; Go-

jun and Fraser, 2012; Collins et al., 2005). The hybrid models applied to English-Hungarian machine translation that we present in this paper belong to the latter line of research.

We applied both reordering and morphological segmentation in order to handle both word order problems and data sparseness caused by agglutination. Luong et al. (2010) applied only morphological analysis in the case of translation from English to Finnish. On the other hand, Yeniterzi and Oflazer (2010) described an approach for English to Turkish translation, in which they applied both syntactic source-side reordering and morphological segmentation. In their work, morphemes constructing a single word were joined during the translation process, but in our experiments, this method increased data sparseness in the training set, decreasing the quality of the final translation rather than improving it. Another difference between Yeniterzi and Oflazer (2010)’s and our work is that they applied the morphological generator integrated in the SMT system, while we used our computational morphology on SMT output as a word form generator, generating final word forms in cases, where the SMT system was not able to find it.

Relying on recent trends and results of research in the field of machine translation, we believe that neither a purely rule-based nor a statistical method by itself is an optimal way to handle the problem. Our work reflects this attitude by applying hand-made language-specific rules. Some works, such as (Jiang et al., 2010; Holmqvist et al., 2012; Genzel, 2010) have also tried deriving such reordering rules automatically.

A further method to apply would be using a hierarchical tree-based translation system, also augmented by reordering rules and morphological segmentation. Such a method is presented in (Gao et al., 2011), but focusing on a narrower problem and applying it to Chinese to English translation.

4 Hybrid morpheme-based machine translation system with reordering rules

In order to mitigate the aforementioned difficulties regarding word order and data sparseness, we created a hybrid system with different preprocessing and decoding solutions. First we applied reordering rules in order to transform the source sentence to a structure more appropriate for word alignment

		Test	Train
# of sentences		1000	1,026,836
Words	en	14.137	14.173
(AVG per sent.)	hu	11.672	11.764
Morphemes	en	16.764	16.768
(AVG per sent.)	hu	18.391	18.429

Table 1: Size of training and test datasets measured in the number of sentences, average number of words per sentences and the average number of morphemes per sentences on the English and Hungarian sides.

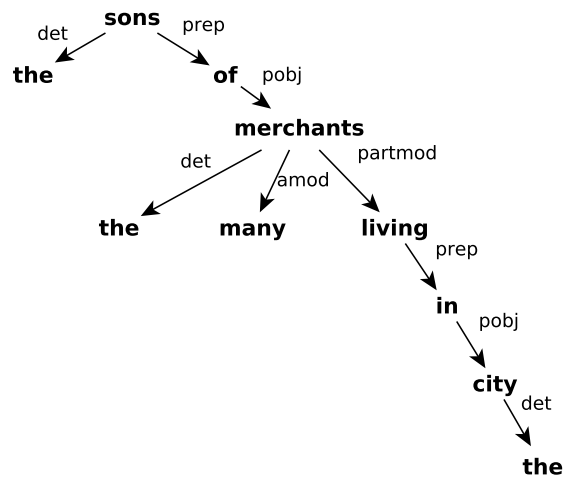
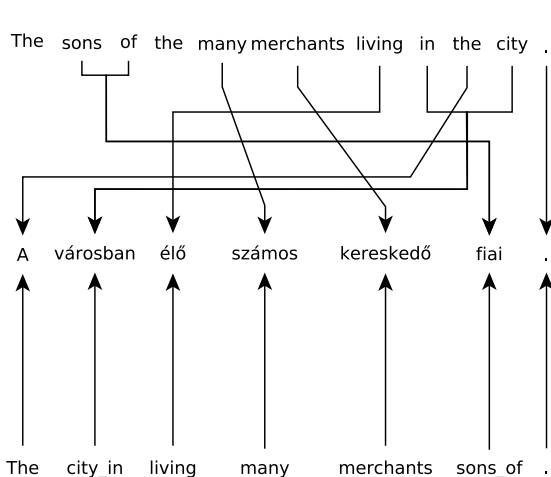
and phrase extraction. The problem of lexical granularity (i.e. the relatively substantial difference in the number of words in the corresponding sentences, see Table 1) was also to be solved. We explored two approaches: a) increasing the number of tokens on both sides using morphemes instead of words and b) decreasing the number of word tokens on the English side to approximate that of the corresponding Hungarian sentences.

4.1 Reordering rules

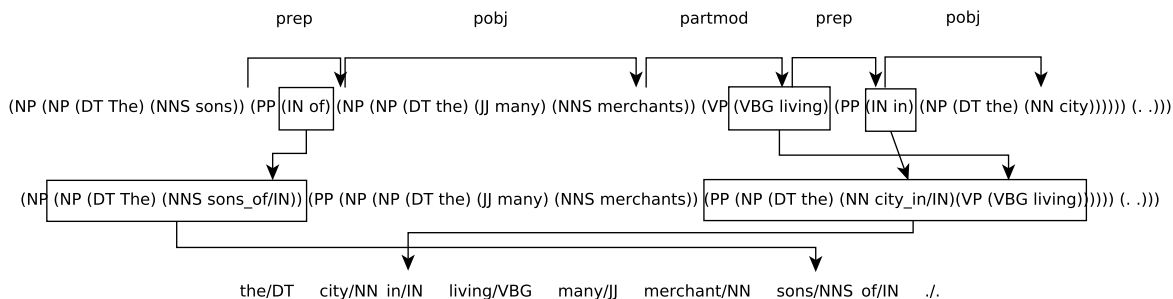
In order to augment the phrase-based SMT system, we defined reordering rules as a preprocessing step. The goal of these transformations is to move words in the English source sentence to positions that correspond to their place in the Hungarian translation. Fig. 1 illustrates the transformation process on the phrase *the sons of the many merchants living in the city*. E.g., the subphrase *living in the city* is transformed to the order *the city in living* corresponding to the Hungarian translation “*a város+ban élő*” as shown in Fig. 1a. Our rules apply only to those word order differences, which are systematically present between the two grammars (e.g. prepositions vs. case endings/postpositions). We did not intend to handle free word order variations of Hungarian, where the same meaning can be expressed with several different orderings, since the actual word order in a sentence is not only determined by syntactic rules, but also by pragmatic factors.

Dependency structure: Reordering rules are guided by dependency relations. After generating a context-free parse, these relations are extracted by the Stanford parser (Marneffe et al., 2006) that we used in our experiments. The dependency structure of our example is shown in Fig. 1b.

Thus the example phrase *merchants living in the city* is transformed along the relations PART-



(a) Word alignment of a sentence pair before and after reordering (b) Dependency structure of the sentence: *The sons of the many merchants living in the city*



(c) The process of reordering along dependency relations.

Figure 1: Word alignment, dependency relations and reordering

MOD(merchant, living)¹, PREP(living, in)¹ and POBJ(in, city)¹. First the preposition is attached to the child of the POBJ relation, then they are positioned before the noun phrase preceding it as shown in Fig. 1c. The resulting word order *the city in living merchants* corresponds to the Hungarian structure "a város+ban élő kereskedők".

Since these levels of analysis depend on each other, errors arising at each phase propagate and cumulate through the whole process having a significant effect on reordering. Even though we used the lexicalized version of the Stanford parser, which is reported to work more accurately, it still very often generates ungrammatical parses with agreement errors and odd PoS sequences as shown in Table 2 (showing only the generated PoS tag sequences here).

¹PARTMOD=participial modifier, PREP=prepositional modifier, POBJ=object of preposition. The full list of dependency relations can be found in http://nlp.stanford.edu/software/dependencies_manual.pdf

-/: 100/CD million/CD sound/NN good/JJ to/TO me/PRP ./.
For/IN airline/NN personnel/NNS ./, we/PRP cash/NN personal/JJ checks/VBZ up/RP to/TO \$/\$ 100/CD ./.

Table 2: Examples of low level errors (verbs tagged as nouns and vice versa) that affect reordering and translation

Morpheme-based restructuring: Due to the agglutinating nature of Hungarian, many function words in English are expressed as suffixes in the Hungarian translation. In order to enable the phrase-based system to have them correspond to each other, we applied morphological analysis on the Hungarian sentences segmenting each word to their morphological constituents. To annotate the Hungarian side of the corpus, we used the PurePos automated morphological annotation system (Orosz and Novák, 2012). A simple example is a phrase like *in my house*, which is

transformed to the form *house my in* corresponding to the single word “*házamban*” in Hungarian. The morphological segmentation of this word is *ház[N]+am[PxSI]+ban[Ine]*¹. Defining and applying the rules for such short phrases is not particularly difficult. However, related words in longer sentences can be much further separated from each other and they may be involved in more than one relation, which often results in an interaction of word order constraints. In a similar manner, some rules insert morphological elements corresponding to those present in the Hungarian sentence, but not explicitly expressed in English, such as the accusative case suffix or subject agreement of verbs. These pieces of implicit structural information can be induced from the dependency relations. For example, in the English phrase *giving/VBG a/DT present/NN*, the word *present* is tagged as *acc* (based on its object role) corresponding to the Hungarian accusative *-t* suffix resulting in the re-ordered phrase of *giving a present+acc* now perfectly aligning to the Hungarian structure of “*adni egy ajándék+ot*”

4.2 Lexical granularity

The number of words is often rather different in a pair of Hungarian and English sentences enforcing the alignment module of the SMT system to create one-to-many or many-to-many alignments, or simply leave tokens unaligned. Such alignments often result in missing or ‘hallucinated’ words in the translation. Table 1 shows the differences in the average number of words and morphemes in our parallel corpus. The average number of words is smaller in Hungarian than in the English sentences. On the other hand, at least at the granularity of the morphological analysis we applied to our data, the number of morphemes is higher in Hungarian than in English. The number of tokens on both sides can be made more similar by either decreasing the number of words on the English side by joining function words corresponding to Hungarian suffixes or by increasing the number on both sides using morphemes as tokens.

As the difference is primarily due to the fact that some English function words are represented as suffixes in Hungarian, the relative difference between the number of morphemes in the corresponding sentences is lower than that of the words. So one possible approach to solving the

lexical granularity difference problem is to use morphemes instead of words. One problem with morpheme-based translation is that it is often the case in longer sentences that instances of the same functional morpheme belong to more than one different word in the sentence. This causes indeterminacies in the alignment process (because the models implemented in the Giza++ word aligner cannot be forced to assume locally monotone alignment at the places where we in fact know that the alignment should be monotone), which often results in erroneous phrases being extracted from the training corpus. For example, if there are two nouns in a sentence and one of them is plural, then the [PL] tag corresponding to this feature might land at another noun.

The difficulty of aligning very frequent functional morphemes is illustrated by the fact that in the Giza++ alignments created from our training corpus, 39% of the nominal plural ([PL]) morphemes remained unaligned, 13% was not attached to the noun it should have been attached to, because the alignment was not monotone, while 1% was aligned to several (up to eight) instances of the corresponding morpheme. Alignment is not the only problem: some indivisible morpheme sequences (like noun+plural) should always stay together but we had concerns that, unless it is constrained to monotone decoding, the baseline distortion model of the decoder will often scatter suffixes throughout the sentence instead. A lexicalized reordering model can be expected to solve this problem, thus we used lexical reordering in our models but for comparison we also tested how each model performs when the decoder is constrained to monotone decoding.

Another approach we tested was fusing separate words on the English side that correspond to a single word in the Hungarian sentence (modeling English as an agglutinating language) to avoid the aligner connecting these morphemes to some other words on the Hungarian side and using a factored model to try to solve the data sparseness issues this move results in. For example, possessive determiners are attached to the head noun as suffixes in this model like the corresponding possessive suffixes in Hungarian : the phrase *my/PRP\$ own/JJ mother/NN* is transformed to the form *own/JJ mother/NN_my/PRP\$*, which corresponds to the Hungarian phrase *saját anyá.m*.

By applying either of the morpheme-token-

¹*PxSI*=Possessor: *ISg*=‘my’, *Ine*=*Inessive*=‘in’

based or the factored morphosyntactic-feature-based solution, the translations generated by the SMT system contain sequences of lemmas and morphosyntactic tags, thus, in order to get the final form of the translated sentence, the surface form of the words have to be generated from the morpheme sequence. In our experiments, we applied the word form generator module of the Humor morphological analyzer to the output of the decoder (Novák, 2003; Prózszéky and Kis, 1999).

4.3 Factored translation

The Moses SMT toolkit (Koehn et al., 2007), which we used in our experiments, is suitable for implementing factored translation models. Instead of relying on just the surface form of the words, further annotations such as morphological analysis can be used in the process of a factored translation. Translation factors might be the surface form of each word, its lemma, its main PoS tag and its morphosyntactic features. During factored translation, there is an opportunity to use multiple translation models, generation models or contextual language models. Since the system has the possibility to use any combination of these, in theory it is able to generate better translations using sparse linguistic data than a word-based baseline system. This feature is vital in cases where some abstraction is necessary, because some words in the sentence to be translated or generated are missing from the training set.

To see how well a factored model performs in the case of translation to an agglutinating language, we also trained a factored translation system combined with our reordering rules. The factors in our case were of the form: `lemma/PoS | PoS+morphtags`, where `PoS` is the main part-of-speech tag and `morphtags` are the rest of the morphological features and extra morphemes attached to the word as described in Section 4.2. Training the system with this combination of factors to handle data sparseness issues seems reasonable in theory; however, translation of lexical and grammatical factors is compromised by a serious weakness of the factored translation implementation in Moses. If the two factors are treated as connected at training time, then if a certain combination of a lemma and its morphology is not present in the translation models, which is very frequent in the case of an agglutinating language, then it can not be translated even if both the lemma and the

morphological feature set are represented in the training corpus separately. In such cases none of the factors are translated and the source word is copied to the output untranslated.

Another method of training a factored model is to translate factors independently. This could indeed solve data sparseness problems, but, as we noted during our experiments, another problem arises in this case: at translation time, translations of morphological tags often land at wrong lemmas. This is due to the fact when translating a phrase, the system selects a translation having one word order, e.g. [Det N V], for one factor (the lemmas) and another, e.g. [V Det N] for the other (the morphosyntactic tags). This results in ill-formed structures, such as nominal morphosyntactic features landing on verbs and verbal morphosyntactic features landing on nouns etc., thus, although the translation might contain the relevant translations regarding both lemmas and morphological features, the final sentence will be an inconsistent mixture of them, making generation of the right word forms impossible. Due to word order variations in Hungarian, this situation turned out to be rather frequent, affecting 21% of our 1000 test sentences.

In order to improve translations compromised by inconsistent mapping of lemmas and morphology, we introduced a postprocessing step extracting and restoring the proper positions of the morphological tags in the result of factored translations. Relying on the alignment information, the proper position of each morphological tag in the sequence can be found. At translation time, Moses can output which source words each target phrase was translated from. We introduced two auxiliary factors to the phrase table that represent alignments of our two main factors. If the alignments in the two factors mismatch, we can realign them using the auxiliary alignment factors (using the word order in the lemma factor as pivot). Once having the factors rematched, the two factors of the target translation are unified and the morphological generator can be applied to generate the final word forms. As it is evident from the evaluation data presented in Section 5, the realignment of factors consistently improved the quality of translations produced by all factored models.

5 Experiments and results

We performed experiments on word-based, morpheme-based and factored translations from English to Hungarian with and without applying our reordering rules as a preprocessing step. We also contrasted the performance of our experimental systems with that of some commercial systems: the rule-based MetaMorpho (Novák et al., 2008; Novák, 2009) and the major commercial translation services, Google Translate and Bing Translator, which apply their language independent statistical systems trained on huge parallel corpora. Low BLEU scores of translations generated by these systems (compared to those usually obtained for other languages) indicate that machine translation to Hungarian is indeed a difficult task.

In all of our experiments, the Moses (Koehn et al., 2007) toolkit was used for building the translation models and performing the translation task itself, using IRSTLM (Federico et al., 2008) to build language models. Wherever it was necessary, PurePos (Orosz and Novák, 2012) was used for morphological analysis and generation, and the Stanford Parser (Marneffe et al., 2006) for constituent and dependency parsing.

5.1 Datasets

As training data, we used the Hunglish (Varga et al., 2005) corpus, created by BME MOKK² and the Research Institute for Linguistics of the Hungarian Academy of Sciences. This corpus contains parallel texts from the following domains: literature and magazines, legal texts and movie subtitles. There is a great degree of variation in the quality of different parts of the corpus. We automatically eliminated sentence pairs from the corpus that caused technical problems, but overall translation quality was not checked.

The corpus we used for training the system consists of 1,026,836 parallel sentences with 14,553,765 words on the English side and 12,079,557 on the Hungarian side. For testing purposes, a 1000-sentence-long portion was selected from the same corpus with one reference translation. Automatic evaluation was performed on this set using the BLEU evaluation metric. Results for each system are listed in Table 3.

²MOKK Centre for Media Research and Education at the Department of Sociology and Communication, Budapest University of Technology and Economics

5.2 Baseline systems

We built a word-based, a morpheme-based, and a factored baseline system (featured as w , m and f in Table 3), not using the reordering rules described in Section 4.1, each trained using Moses.

For the word-based baseline model w , the only preprocessing we applied was standard tokenization and lowercasing. A phrase table with a phrase length limit of 7 was extracted, and a 5-gram language model was built. A lexicalized reordering model with a distortion limit of 6 was used in this baseline model (and all other models with non-monotone decoding).

We evaluated this system using two automatic metrics: the usual word-based BLEU (w-BLEU) and, in order to have a relevant base of comparison to the other systems, a morpheme-based score (mm-BLEU), which in the case of the word-based baseline was computed applying morphological analysis to the translations. mm-BLEU is based on counts of identical abstract morpheme sequences in the generated and the reference translations instead of identical word sequences. Note that this differs from m-BLEU as used in e.g. (Clifton and Sarkar, 2011), which is BLEU applied to pseudo-morphs generated by an unsupervised segmenter. mm-BLEU measures the ability of the system to generate the correct morphemes in the translations.

The second baseline system m was trained on morphologically segmented sentences, thus the output of the decoder is a sequence of morphemes. A BLEU score computed on the output of the decoder in this case is mm-BLEU. The morphological generator was applied to the output of the Moses decoder in order to acquire the final word forms. The morpheme-based system m performed better in terms of mm-BLEU, although it got a lower w-BLEU score.

The third, factored baseline model f was outperformed by the two other models both in terms of w-BLEU and mm-BLEU, even when the problem caused by a different word order in the factors was fixed as described in Section 4.3 (the system fx).

5.3 Reordered models

Based on considerations described in Sections 4.1 and 4.2, we performed reordering as a preprocessing step both at training and translation time. Models using this configuration were also evaluated applying the same w-BLEU and mm-BLEU

ID	w-BLEU	mm-BLEU
w-based baseline (<i>w</i>)	14.57%	59.32%
m-based baseline mon. (<i>mm</i>)	11.69%	63.18%
m-based baseline (<i>m</i>)	12.19%	63.87%
factored baseline monotone (<i>fm</i>)	9.70%	56.00%
factored baseline mon. fixed (<i>fm_x</i>)	9.84%	57.09%
w-based reord. (<i>wre</i>)	14.83%	58.06%
w-based reord. joined (<i>wre₋</i>)	13.05%	57.21%
m-based reord. mon. (<i>mrem</i>)	12.01%	64.24%
m-based reord. (<i>mre</i>)	12.22%	64.94%
fact. reord. mon. (<i>frem</i>)	10.50%	59.56%
fact. reord. mon. fixed (<i>frem_x</i>)	10.64%	60.28%
fact. reord. (<i>fre</i>)	10.78%	59.97%
fact. reord. fixed (<i>fre_x</i>)	10.88%	60.83%
Google Translate (<i>goo</i>)	15.68%	55.86%
Bing Translator (<i>bing</i>)	12.16%	53.05%
MetaMorpho (<i>mmo</i>)	6.86%	50.97%

Table 3: Automatic evaluation scores for systems tested in the experiments.

metrics. We implemented various morpheme-based, factored and word-based reordered models. The two word-based setups performed the same transformations moving function words, the difference between the two was only whether the moved words were kept as distinct words (*wre*) or joined to the target word as suffixes to form a single word form (*wre₋*). The models allowed further reordering during decoding using a lexicalized reordering model.

The morpheme-based (*mre*) and the factored models (*fre* and *fre_x*, the latter with factor misalignment fixed) were contrasted with alternative setups where the decoder was constrained to monotone decoding (*mrem*, *frem*, *frem_x*). We had concerns that in the case of the morpheme-based model the decoder might move suffixes to incorrect positions. However, using a lexicalized reordering model prevented these problems and the systems with reordering during decoding performed consistently better. Monotone decoding blocked the decoder from fixing word order in the preverbal field of the comment part of Hungarian sentences, where strict word order constraints apply in contrast to the free word order of the topic and the postverbal part of the comment. While our reordering rules did not capture these constraints depending on various subtle features of the actually selected translation that cannot be reliably inferred from the English original, the lexically constrained reordering performed by the decoder did manage to generate translations that conformed to them at least to some extent.

The results presented in Table 3 show that the reordered *wre*, *mre* and *fre_x* models obtained consistently higher BLEU scores than the corresponding baseline models (the only exception being the mm-BLEU score of the *wre* model). Although the BLEU scores do not show this clearly, the translations generated by the *wre₋* model are far worse than the output of any other system due to a high number of untranslated “agglutinating English” words with function words attached to content words as suffixes.

Figure 4 shows the translation results of our different systems. As it can be seen, *mre* performed the best, regarding fluency and reflecting the original meaning.

6 Human evaluation

It has been shown that system rankings based on single reference BLEU scores often do not correspond to how humans evaluate the translations. For this reason, automatic evaluation has for a long time not been used to officially rank systems at Workshops on Statistical Machine Translation (WMT) (Callison-Burch et al., 2007). In our work, we presented results of automated evaluation using a single reference BLEU metrics, but we also investigated translations generated by each system using human evaluation, applying the ranking scheme used at WMT workshops to officially rank systems.

300 sentences were randomly chosen from the test set for the purpose of human evaluation. Five annotators evaluated translations generated by each of the above described systems plus the reference translation in the corpus with regard to translation quality (considering both adequacy and fluency in a single quality ranking). The order of translations was randomized for each sentence and a balanced number of comparisons was performed for each system pair. The systems were ranked based on a score that was defined as the number of times the output of a system was deemed not worse than that of the other in pairwise comparisons divided by the number of pairwise comparisons. The aggregate results of human evaluation are listed in Table 5.

Manual investigation of the translation outputs revealed that the system incorporating morphological and syntactic information are better at capturing grammatical relations in the original text and rendering them in the translation by generating the

original English	After you were picked up at sea , our listening post in Malta intercepted that fax .
reordered English	after/[IN] you/[PRP] be/[VB] [Past] pick/[VB] [PPart] up/[RP] at/[IN] sea/[NN] ./[.], our/[PRP\$] listen/[VB] [ING] post/[NN] in/[IN] malta/[NNP] intercept/[VB] [PPart] that/[DT] fax/[NN] ./[.].
morpheme based translation	miután/[KOT] felvesz/[IGE] [Past] [t3] [Def] maga/[FN_NM] [e3] [ACC] a/[DET] tenger/[FN] [SUP] ./[PUNCT] hallgat/[IGE] [Past] [e3] [Def] a/[DET] hely/[FN] [PSt1] ./[PUNCT] hogy/[KOT] máltá/[FN] [INE] áll/[IGE] [Past] [e3] [Def] ez/[FN_NM] [ACC] a/[DET] fax/[FN] [ACC] ./[PUNCT]
final translation back-translation	Miután felvették magát a tengeren , hallgatta a helyünk , hogy máltá állta ezt a faxot . After you were picked up at sea, our listening post caught the fax in Malta.
baseline translation back-translation	Azután , hogy felvette a tengeren , a máltai hallgatta az emelkedő , hogy fax . After you, he picked it up at the sea, and that Malta were caught, that it is a fax.
Hungarian reference back-translation	Miután önt kihalászták , ezt fogták el egy máltai postán . After you were fished out, this was caught at a post in Malta.

Table 4: Translation results of our systems with hand made backtranslations for comparison with the reference.

ref	mmo	goo	bing	mre	frex	m	fx	w	wre	wre_
88.33	76.30	72.80	61.66	55.60	55.42	54.28	52.03	51.33	50.89	37.57

Table 5: Human evaluation ranking of systems measured as percentage of generating a translation not worse than the other in pairwise comparisons

appropriate inflected forms. Rule-based reordering also improved quality when using linguistically rich models. The only ones that performed worse than the baseline were the word-based re-ordered solutions, especially the one based on “agglutinating English”, the poor performance of which came as no surprise. BLEU scores do not correspond well to human judgments. Of our models, the *wre* system had the highest BLEU score, however, human evaluation ranked that worse than any of the morpheme-based systems. Moreover, MetaMorpho, the commercial system having highest rank had by far the lowest BLEU score.

Considering all the systems in the ranking procedure, it can be observed that the reference translation used also for measuring BLEU score does not always represent the best translation either according to our evaluators. It is worth noting though that there was a rather significant variance in the ranking of reference translations due to some evaluators ranking them much less favourably than others (75.29% vs. 92.98%).

7 Conclusion

We performed several experiments on English-Hungarian machine translation. Automatic evaluation consistently scored models including rule-based reordering higher than systems not including it. Human evaluation confirmed that applying reordering and morphological segmentation does

improve translation quality in the case of translating to an agglutinating language like Hungarian.

Our models are not yet on par with commercial systems. The rather limited amount of training corpus that also has serious quality problems is certainly one factor playing a role in this. Our future plans include enlarging and improving our training corpus, improving alignment and components of the syntactic annotation and reordering chain as well as experimenting with combination of morpheme-based and factored models.

Acknowledgement

This work was partially supported by TÁMOP – 4.2.1.B – 11/2/KMR-2011-0002 and TÁMOP – 4.2.2./B – 10/1-2010-0014.

References

- Chris Callison-Burch, Cameron Forgyce, Philipp Koehn, Christof Monz, and Josh Schroeder. 2007. (Meta-)evaluation of machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT ’07, pages 136–158, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ann Clifton and Anoop Sarkar. 2011. Combining morpheme-based machine translation with post-processing morpheme prediction. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT ’11, pages 32–42, Stroudsburg,

- PA, USA. Association for Computational Linguistics.
- Michael Collins, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 531–540, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marcello Federico, Nicola Bertoldi, and Mauro Cettolo. 2008. Irstlm: an open source toolkit for handling large scale language models. In *INTER-SPEECH*, pages 1618–1621.
- Yang Gao, Philipp Koehn, and Alexandra Birch. 2011. Soft dependency constraints for reordering in hierarchical Phrase-Based translation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 857–868, Edinburgh, Scotland, UK., jul. Association for Computational Linguistics.
- Dmitriy Genzel. 2010. Automatically learning source-side reordering rules for large scale machine translation. In *COLING*, pages 376–384.
- Anita Gojun and Alexander Fraser. 2012. Determining the placement of German verbs in English-to-German SMT. In Walter Daelemans, Mirella Lapata, and Lluís Mrquez, editors, *EACL*, pages 726–735. The Association for Computer Linguistics.
- Hieu Hoang. 2007. Factored translation models. In *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 868–876.
- Maria Holmqvist, Sara Stymne, Lars Ahrenberg, and Magnus Merkel. 2012. Alignment-based reordering for SMT. In Nicoletta Calzolari (Conference Chair) et al., editor, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Jie Jiang, Jinhua Du, and Andy Way. 2010. Source-side syntactic reordering patterns with functional words for improved phrase-based SMT. In *Proceedings of SSST-4, Fourth Workshop on Syntax and Structure in Statistical Translation*, pages 19–27, Beijing.
- 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 *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 177–180, Prague. Association for Computational Linguistics.
- Minh-Thang Luong, Preslav Nakov, and Min-Yen Kan. 2010. A hybrid morpheme-word representation for machine translation of morphologically rich languages. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 148–157, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marie-Catherine De Marneffe, Bill Maccartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *In LREC 2006*.
- Attila Novák, László Tihanyi, and Gábor Prószték. 2008. The MetaMorpho translation system. In *Proceedings of the Third Workshop on Statistical Machine Translation*, StatMT '08, pages 111–114, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Attila Novák. 2003. What is good Humor like? In *I. Magyar Számítógépes Nyelvészeti Konferencia*, pages 138–144, Szeged. SZTE.
- Attila Novák. 2009. MorphoLogic's submission for the WMT 2009 Shared Task. In *Proceedings of the Fourth Workshop on Statistical Machine Translation at EACL 2009*, Athens, Greece.
- György Orosz and Attila Novák. 2012. PurePos – an open source morphological disambiguator. In *Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science.*, Wrocław, Poland.
- Kishore Papineni, 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*, ACL '02, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Gábor Prószték and Balázs Kis. 1999. A unification-based approach to morpho-syntactic parsing of agglutinative and other (highly) inflectional languages. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 261–268, Stroudsburg, PA, USA. Association for Computational Linguistics.
- D. Varga, L. Németh, P. Halácsy, A. Kornai, V. Trón, and V. Nagy. 2005. Parallel corpora for medium density languages. In *Recent Advances in Natural Language Processing (RANLP 2005)*, pages 590–596.
- Reyyan Yeniterzi and Kemal Oflazer. 2010. Syntax-to-morphology mapping in factored phrase-based statistical machine translation from English to Turkish. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 454–464, Stroudsburg, PA, USA. Association for Computational Linguistics.