Syntax-to-Morphology Mapping in Factored Phrase-Based Statistical Machine Translation from English to Turkish

Reyyan Yeniterzi
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, 15213, USA
reyyan@cs.cmu.edu

Kemal Oflazer
Computer Science
Carnegie Mellon University-Qatar
PO Box 24866, Doha, Qatar
ko@cs.cmu.edu

Abstract

We present a novel scheme to apply factored phrase-based SMT to a language pair with very disparate morphological structures. Our approach relies on syntactic analysis on the source side (English) and then encodes a wide variety of local and non-local syntactic structures as *complex structural tags* which appear as additional factors in the training data. On the target side (Turkish), we only perform morphological analysis and disambiguation but treat the complete complex morphological tag as a factor, instead of separating morphemes. We incrementally explore capturing various syntactic substructures as complex tags on the English side, and evaluate how our translations improve in BLEU scores. Our maximal set of source and target side transformations, coupled with some additional techniques, provide an 39% relative improvement from a baseline 17.08 to 23.78 BLEU, all averaged over 10 training and test sets. Now that the syntactic analysis on the English side is available, we also experiment with more long distance constituent reordering to bring the English constituent order close to Turkish, but find that these transformations do not provide any additional consistent tangible gains when averaged over the 10 sets.

1 Introduction

Statistical machine translation into a morphologically complex language such as Turkish, Finnish or Arabic, involves the generation of target words with the proper morphology, in addition to properly ordering the target words. Earlier work on translation from English to Turkish (Oflazer and Durgar-El-Kahlout, 2007; Oflazer, 2008; Durgar-El-Kahlout and Oflazer, 2010) has used an approach which relied on identifying the contextually correct parts-of-speech, roots and any morphemes on the English side, and the complete sequence of roots and overt derivational and inflectional morphemes for each word on the Turkish side. Once these were identified as separate tokens, they were then used as "words" in a standard phrase-based framework (Koehn et al., 2003). They have reported that, given the typical complexity of Turkish words, there was a substantial percentage of words whose morphological structure was incorrect: either the morphemes were not applicable for the part-of-speech category of the root word selected, or the morphemes were in the wrong order. The main reason given for these problems was that the same statistical translation, reordering and language modeling mechanisms were being employed to *both* determine the morphological structure of the words *and*, at the same time, get the global order of the words correct. Even though a significant improvement of a standard word-based baseline was achieved, further analysis hinted at a direction where morphology and syntax on the Turkish side had to be dealt with using separate mechanisms.

Motivated by the observation that many local and some nonlocal syntactic structures in English essentially map to morphologically complex words in Turkish, we present a radically different approach which does not segment Turkish words into morphemes, but uses a representation equivalent to the full word form. On the English side, we rely on a full syntactic analysis using a dependency parser. This analysis then lets us abstract and encode many local and some nonlocal syntactic structures as complex tags (dynamically, as opposed to the static complex tags as proposed by Birch et *al.* (2007) and Hassan et *al.* (2007)). Thus

Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 454–464,
Uppsala, Sweden, 11-16 July 2010.

we can bring the representation of English syntax closer to the Turkish morphosyntax.

Such an approach enables the following: (i) Driven by the pattern of morphological structures of full word forms on the Turkish side represented as root words and complex tags, we can identify and reorganize phrases on the English side, to "align" English syntax to Turkish morphology wherever possible. (ii) Continuous and discontinuous variants of certain (syntactic) phrases can be conflated during the SMT phrase extraction process. (iii) The length of the English sentences can be dramatically reduced, as most function words encoding syntax are now abstracted into complex tags. (iv) The representation of both the source and the target sides of the parallel corpus can now be mostly normalized. *This facilitates the use of factored phrase-based translation that was not previously applicable due to the morphological complexity on the target side and mismatch between source and target morphologies.*

We find that with the full set of syntax-to-morphology transformations and some additional techniques we can get about 39% relative improvement in BLEU scores over a word-based baseline and about 28% improvement of a factored baseline, all experiments being done over 10 training and test sets. We also find that further constituent reordering taking advantage of the syntactic analysis of the source side, does not provide tangible improvements when averaged over the 10 data sets.

This paper is organized as follows: Section 2 presents the basic idea behind syntax-to-morphology alignment. Section 3 describes our experimental set-up and presents results from a sequence of incremental syntax-to-morphology transformations, and additional techniques. Section 4 summarizes our constituent reordering experiments and their results. Section 5 presents a review of related work and situates our approach.

We assume that the reader is familiar with the basics of phrase-based statistical machine translation (Koehn et al., 2003) and factored statistical machine translation (Koehn and Hoang, 2007).

2 Syntax-to-Morphology Mapping

In this section, we describe how we map between certain source language syntactic structures and target words with complex morphological structures. At the top of Figure 1, we see a pair of (syntactic) phrases, where we have (positionally) aligned the words that should be translated to each other. We can note that the function words *on* and

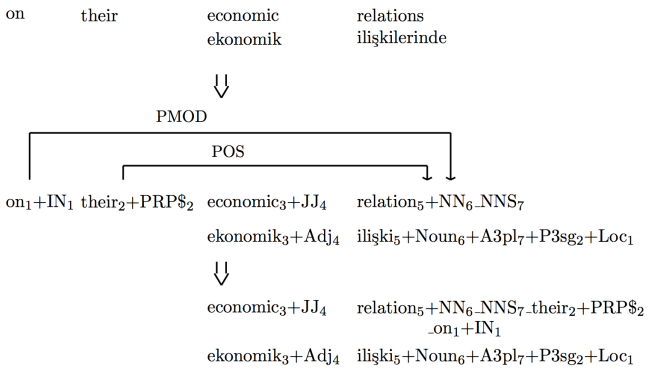

Figure 1: Transformation of an English prepositional phrase

their are not really aligned to any of the Turkish words as they really correspond to two of the morphemes of the last Turkish word.

When we tag and syntactically analyze the English side into dependency relations, and morphologically analyze and disambiguate the Turkish phrase, we get the representation in the middle of Figure 1, where we have co-indexed components that should map to each other, and some of the syntactic relations that the function words are involved in are marked with dependency links.[1]

The basic idea in our approach is to take various function words on the English side, whose syntactic relationships are identified by the parser, and then package them as complex tags on the related content words. So, in this example, if we move the first two function words from the English side and attach as *syntactic tags* to the word they are in dependency relation with, we get the aligned representation at the bottom of Figure 1.[2,3] Here we can note that all root words and tags that correspond to each other are nicely structured and are in the same relative order. In fact, we can treat each token as being composed of two factors: the roots and the accompanying tags. The tags on the Turkish side encode morphosyntactic information encoded in the morphology of the words, *while the*

[1]The meanings of various tags are as follows: Dependency Labels: **PMOD** - Preposition Modifier; **POS** - Possessive. Part-of-Speech Tags for the English words: **+IN** - Preposition; **+PRP$** - Possessive Pronoun; **+JJ** - Adjective; **+NN** - Noun; **+NNS** - Plural Noun. Morphological Feature Tags in the Turkish Sentence: **+A3pl** - 3rd person plural; **+P3sg** - 3rd person singular possessive; **+Loc** - Locative case. Note that we mark an English plural noun as **+NN_NNS** to indicate that the root is a noun and there is a plural morpheme on it. Note also that *economic* is also related to *relations* but we are not interested in such content words and their relations.

[2]We use _ to prefix such syntactic tags on the English side.

[3]The order is important in that we would like to attach the same sequence of function words in the same order so that the resulting tags on the English side are the same.

(complex) tags on the English side encode local (and sometimes, non-local) syntactic information. Furthermore, we can see that before the transformations, the English side has 4 words, while afterwards it has only 2 words. We find (and elaborate later) that this reduction in the English side of the training corpus, in general, is about 30%, and is correlated with improved BLEU scores. We believe the removal of many function words and their folding into complex tags (which do not get involved in GIZA++ alignment – we only align the root words) seems to improve alignment as there are less number of "words" to worry about during that process.[4]

Another interesting side effect of this representation is the following. As the complex syntactic tags on the English side are based on syntactic relations and not necessarily positional proximity, the tag for *relations* in a phrase like *in their cultural, historical and economic relations* would be exactly the same as above. Thus phrase extraction algorithms can conflate all constructs like *in their ... economic relations* as one phrase, regardless of the intervening modifiers, assuming that parser does its job properly.

Not all cases can be captured as cleanly as the example above, but most transformations capture local and nonlocal syntax involving many function words and then encode syntax with complex tags resembling full morphological tags on the Turkish side. *These transformations, however, are not meant to perform sentence level constituent reordering on the English side.* We explore these later.

We developed set of about 20 linguistically-motivated syntax-to-morphology transformations which had variants parameterized depending on what, for instance, the preposition or the adverbial was, and how they map to morphological structure on the Turkish side. For instance, one general rule handles cases like *while ... verb* and *if ... verb* etc., mapping these to appropriate complex tags. It is also possible that multiple transformations can apply to generate a single English complex tag: a portion of the tag can come from a verb complex transformation, and another from an adverbial phrase transformation involving a marked such as *while*. Our transformations handle the following cases:

- Prepositions attach to the head-word of their complement noun phrase as a component in its complex tag.
- Possessive pronouns attach to the head-word they specify.
- The possessive markers following a noun (separated by the tokenizer) attached to the noun.
- Auxiliary verbs and negation markers attach to the lexical verb that they form a verb complex with.
- Modals attach to the lexical verb they modify.
- Forms of *be* used as predicates with adjectival or nominal dependents attach to the dependent.
- Forms of *be* or *have* used to form passive voice with past participle verbs, and forms of *be* used with *-ing* verbs to form present continuous verbs, attach to the verb.
- Various adverbial clauses formed with *if*, *while*, *when*, etc., are reorganized so that these markers attach to the head verb of the clause.

As stated earlier, these rules are linguistically motivated and are based on the morphological structure of the *target* language words. Hence for different target languages these rules will be different. The rules recognize various local and nonlocal syntactic structures in the source side parse tree that correspond to complex morphological of target words and then remove source function words folding them into complex tags. For instance, the transformations in Figure 1 are handled by scripts that process Malt Parser's dependency structure output and that essentially implement the following sequence of rules expressed as pseudo code:

```
1) if  (<Y>+PRP$ POS <Z>+NN<TAG>)
   then {
        APPEND <Y>+PRP$ TO <Z>+NN<TAG>
        REMOVE <Y>+PRP$
     }

2) if (<X>+IN  PMOD <Z>+NN<TAG>)
   then {
        APPEND <X>+IN TO <Z>+NN<TAG>
        REMOVE <X>+IN
     }
```

Here <X>, <Y> and <Z> can be considered as Prolog like-variables that bind to patterns (mostly root words), and the conditions check for specified dependency relations (e.g., PMOD) between the left and the right sides. When the condition is satisfied, then the part matching the function word is removed and its syntactic information is appended to form the complex tag on the noun (<TAG> would either match null string or any previously appended function word markers.)[5]

[4] Fraser (2009) uses the first four letters of German words after morphological stripping and compound decomposition to help with alignment in German to English and reverse translation.

[5] We outline two additional rules later when we see a more complex example in Figure 2.

There are several other rules that handle more mundane cases of date and time constructions (for which, the part of the date construct which the parser attaches a preposition, is usually different than the part on the Turkish side that gets inflected with case markers, and these have to be reconciled by overriding the parser output.)

The next section presents an example of a sentence with multiple transformations applied, after discussing the preprocessing steps.

3 Experimental Setup and Results

3.1 Data Preparation

We worked on an English-Turkish parallel corpus which consists of approximately 50K sentences with an average of 23 words in English sentences and 18 words in Turkish sentences. This is the same parallel data that has been used in earlier SMT work on Turkish (Durgar-El-Kahlout and Oflazer, 2010). Let's assume we have the following pair of parallel sentences:

E: if a request is made orally the authority must make a record of it
T: istek sözlü olarak yapılmışsa yetkili makam bunu kaydetmelidir

On the English side of the data, we use the Stanford Log-Linear Tagger (Toutanova et al., 2003), to tag the text with Penn Treebank Tagset. On the Turkish side, we perform a full morphological analysis, (Oflazer, 1994), and morphological disambiguation (Yuret and Türe, 2006) to select the contextually salient interpretation of words. We then remove any morphological features that are not explicitly marked by an *overt morpheme*.[6] So for both sides we get,

E: if+IN a+DT request+NN is+VBZ made+VBN orally+RB the+DT authority+NN must+MD make+VB a+DT record+NN of+IN it+PRP
T: istek+Noun sözlü+Adj olarak+Verb+ByDoingSo yap+Verb+Pass+Narr+Cond yetkili+Adj makam+Noun bu+Pron+Acc kaydet+Verb+Neces+Cop

Finally we parse the English sentences using MaltParser (Nivre et al., 2007), which gives us labeled dependency parses. On the output of the parser, we make one more transformation. We replace each word with its root, and possibly add an additional tag for any inflectional information conveyed by overt morphemes or exceptional forms. This is done by running the TreeTagger (Schmid, 1994) on the English side which provides the roots in addition to the tags, and then carrying over this information to the parser output. For example, `is` is tagged as `be+VB_VBZ`, `made` is tagged as `make+VB_VBN`, and a word like `books` is tagged as `book+NN_NNS` (and not as `books+NNS`). On the Turkish side, each marker with a preceding + is a morphological feature. The first marker is the part-of-speech tag of the root and the remainder are the overt inflectional and derivational markers of the word. For example, the analysis `kitap+Noun+P2pl+A3pl+Gen` for a word like `kitap-lar-ınız-ın`[7] (*of your books*) represents the root `kitap` (*book*), a `Noun`, with third person plural agreement `A3pl`, second person plural possessive agreement, `P2pl` and genitive case `Gen`.

The sentence representations in the middle part of Figure 2 show these sentences with some of the dependency relations (relevant to our transformations) extracted by the parser, explicitly marked as labeled links. The representation at the bottom of this figure (except for the co-indexation markers) corresponds to the final transformed form of the parallel training and test data. The co-indexation is meant to show which root words on one side map to which on the other side. Ultimately we would want the alignment process to uncover the *root word alignments* indicated here. *We can also note that the initial form of the English sentence has 14 words and the final form after transformations, has 7 words (with complex tags).*[8]

3.2 Experiments

We evaluated the impact of the transformations in factored phrase-based SMT with an English-Turkish data set which consists of 52712 parallel sentences. In order to have more confidence in the impact of our transformations, we randomly generated 10 training, test and tune set combinations. For each combination, the latter two were 1000 sentences each and the remaining 50712 sentences were used as training sets.[9,10]

We performed our experiments with the Moses toolkit (Koehn et al., 2007). In order to encourage long distance reordering in the decoder, we used a distortion limit of -1 and a distortion weight of

[6] For example, the morphological analyzer outputs `+A3sg` to mark a singular noun, if there is no explicit plural morpheme. Such markers are removed.

[7] – shows surface morpheme boundaries.

[8] We could give two more examples of rules to process the if-clause in the example in Figure 2. These rules would be applied sequentially: The first rule recognizes the passive construction mediated by `be+VB<AGR>` forming a verb complex (`VC`) with `<Y>+VB_VBN` and appends the former to the complex tag on the latter and then deletes the former token. The second rule then recognizes `<X>+IN` relating to `<Y>+VB<TAGS>`with `VMOD` and appends the former to the complex tag on the latter and then deletes the former token.

[9] The tune set was not used in this work but reserved for future work so that meaningful comparisons could be made.

[10] It is possible that the 10 test sets are not mutually exclusive.

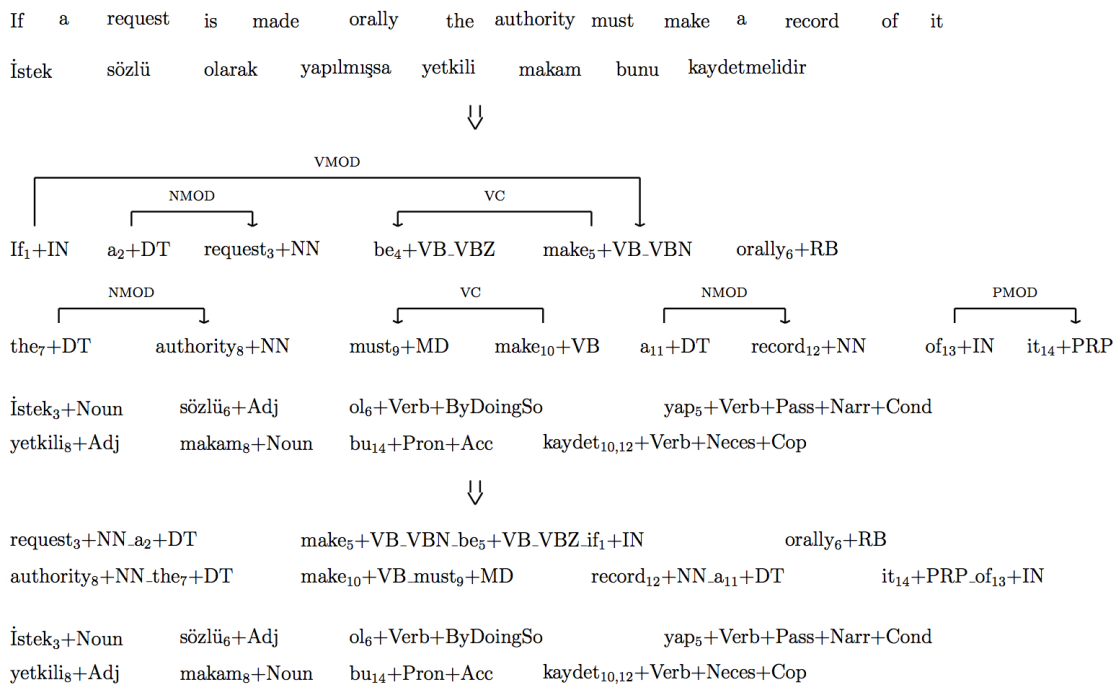

Figure 2: An English-Turkish sentence pair with multiple transformations applied

0.1.[11] We did not use MERT to further optimize our model.[12]

For evaluation, we used the BLEU metric (Papineni et al., 2001). Each experiment was repeated over the 10 data sets. Wherever meaningful, we report the average BLEU scores over 10 data sets along with the maximum and minimum values and the standard deviation.

[11]These allow and do not penalize unlimited distortions.

[12]The experience with MERT for this language pair has not been very positive. Earlier work on Turkish indicates that starting with default Moses parameters and applying MERT to the resulting model does not even come close to the performance of the model with those two specific parameters set as such (*distortion limit -1* and *distortion weight 0.1*), most likely because the default parameters do not encourage the range of distortions that are needed to deal with the constituent order differences. Earlier work on Turkish also shows that even when the *weight-d* parameter is initialized with this specific value, the space explored for distortion weight and other parameters do not produce any improvements on the test set, even though MERT claims there are improvements on the tune set.

The other practical reasons for not using MERT were the following: at the time we performed this work, the discussion thread at `http://www.mail-archive.com/moses-support@mit.edu/msg01012.html` indicated that MERT was not tested on multiple factors. The discussion thread at `http://www.mail-archive.com/moses-support@mit.edu/msg00262.html` claimed that MERT does not help very much with factored models. With these observations, we opted not to experiment with MERT with the multiple factor approach we employed, given that it would be risky and time consuming to run MERT needed for 10 different models and then not necessarily see any (consistent) improvements. MERT however is orthogonal to the improvements we achieve here and can always be applied on top of the best model we get.

3.2.1 The Baseline Systems

As a baseline system, we built a standard phrase-based system, using the surface forms of the words without any transformations, and with a 3-gram LM in the decoder. We also built a second baseline system with a factored model. Instead of using just the surface form of the word, we included the root, part-of-speech and morphological tag information into the corpus as additional factors alongside the surface form.[13] Thus, a token is represented with three factors as `Surface|Root|Tags` where `Tags` are complex tags on the English side, and morphological tags on the Turkish side.[14]

Moses lets word alignment to align over any of the factors. We aligned our training sets using only the root factor to conflate statistics from different forms of the same root. The rest of the factors are then automatically assumed to be aligned, based on the root alignment. Furthermore, in factored models, we can employ different language models for different factors. For the initial set of experiments we used 3-gram LMs for all the factors.

For factored decoding, we employed a model whereby we let the decoder translate a surface form directly, but if/when that fails, the decoder can back-off with a generation model that builds a target word from independent translations of the root and tags.

[13]In Moses, factors are separated by a '|' symbol.

[14]Concatenating `Root` and `Tags` gives the `Surface` form, in that the surface is unique given this concatenation.

The results of our baseline models are given in top two rows of Table 1. As expected, the word-based baseline performs worse than the factored baseline. We believe that the use of multiple language models (some much less sparse than the surface LM) in the factored baseline is the main reason for the improvement.

3.2.2 Applying Syntax-to-Morphology Mapping Transformations

To gauge the effects of transformations separately, we first performed them in batches on the English side. These batches were (i) transformations involving nouns and adjectives (*Noun+Adj*), (ii) transformations involving verbs (*Verb*), (iii) transformations involving adverbs (*Adv*), and (iv) transformations involving verbs and adverbs (*Verb+Adv*).

We also performed one set of transformations on the Turkish side. In general, English prepositions translate as case markers on Turkish nouns. However, there are quite a number of lexical *postpositions* in Turkish which also correspond to English prepositions. To normalize these with the handling of case-markers, we treated these postpositions as if they were case-markers and attached them to the immediately preceding noun, and then aligned the resulting training data (*PostP*).[15]

The results of these experiments are presented in Table 1. We can observe that the combined syntax-to-morphology transformations on the source side provide a substantial improvement by themselves and a simple target side transformation on top of those provides a further boost to 21.96 BLEU which represents a 28.57% relative improvement over the word-based baseline and a 18.00% relative improvement over the factored baseline.

Experiment	Ave.	STD	Max.	Min.
Baseline	17.08	0.60	17.99	15.97
Factored Baseline	**18.61**	**0.76**	**19.41**	**16.80**
Noun+Adj	21.33	0.62	22.27	20.05
Verb	19.41	0.62	20.19	17.99
Adv	18.62	0.58	19.24	17.30
Verb+Adv	19.42	0.59	20.17	18.13
Noun+Adj +Verb+Adv	**21.67**	**0.72**	**22.66**	**20.38**
Noun+Adj+Verb +Adv+PostP	**21.96**	**0.72**	**22.91**	**20.67**

Table 1: BLEU scores for a variety of transformation combinations

We can see that every transformation improves the baseline system and the highest performance is attained when all transformations are performed. However when we take a closer look at the individual transformations performed on English side, we observe that not all of them have the same effect. While *Noun+Adj* transformations give us an increase of 2.73 BLEU points, *Verbs* improve the result by only 0.8 points and improvement with *Adverbs* is even lower. To understand why we get such a difference, we investigated the correlation of the decrease in the number of tokens on both sides of the parallel data, with the change in BLEU scores. The graph in Figure 3 plots the BLEU scores and the number of tokens in the two sides of the training data as the data is modified with transformations. We can see that as the number of tokens in English decrease, the BLEU score increases. In order to measure the relationship between these two variables statistically, we performed a correlation analysis and found that there is a strong negative correlation of -0.99 between the BLEU score and the number of English tokens. We can also note that the largest reduction in the number of tokens comes with the application of the *Noun+Adj* transformations, which correlates with the largest increase in BLEU score.

It is also interesting to look at the n-gram precision components of the BLEU scores (again averaged). In Table 2, we list these for words (actual BLEU), roots (BLEU-R) to see how effective we are in getting the root words right, and morphological tags, (BLEU-M), to see how effective we are in getting just the morphosyntax right. It

		1-gr.	2-gr.	3-gr.	4-gr.
BLEU	21.96	55.73	27.86	16.61	10.68
BLEU-R	27.63	68.60	35.49	21.08	13.47
BLEU-M	27.93	67.41	37.27	21.40	13.41

Table 2: Details of Word, Root and Morphology BLEU Scores

seems we are getting almost 69% of the root words and 68% of the morphological tags correct, but not necessarily getting the combination equally as good, since only about 56% of the full word forms are correct. One possible way to address is to use longer distance constraints on the morphological tag factors, to see if we can select them better.

3.2.3 Experiments with higher-order language models

Factored phrase-based SMT allows the use of multiple language models for the target side, for different factors during decoding. Since the number of possible distinct morphological tags (the morphological tag vocabulary size) in our training data

[15]Note than in this case, the translations would be generated in the same format, but we then split such postpositions from the words they are attached to, during decoding, and *then* evaluate the BLEU score.

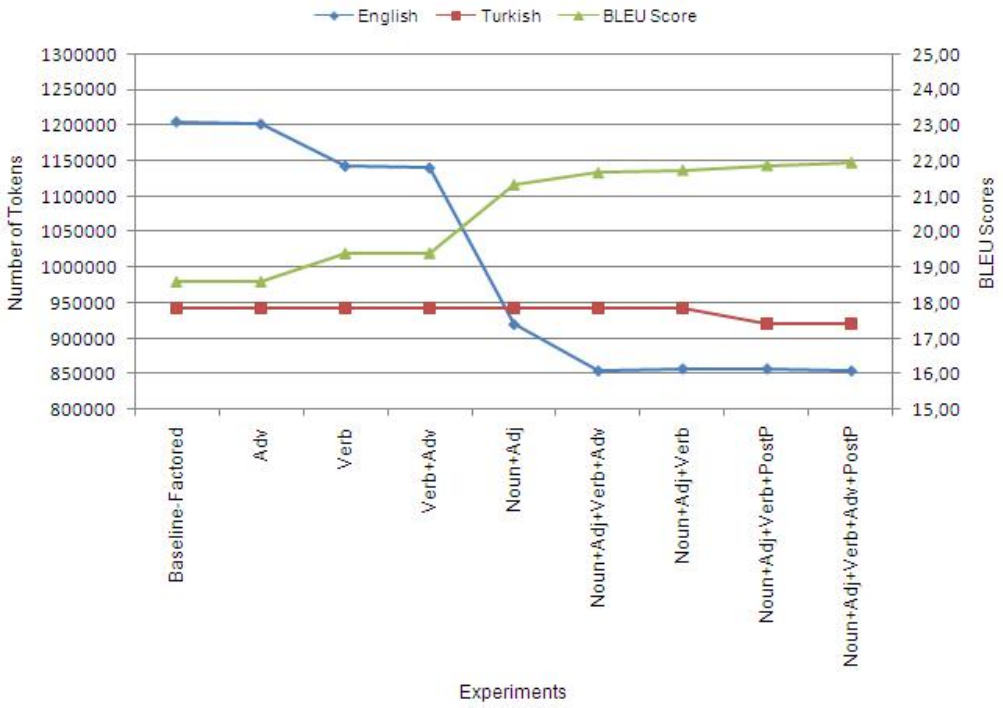

Figure 3: BLEU scores vs number of tokens in the training sets

(about 3700) is small compared to distinct number of surface forms (about 52K) and distinct roots (about 15K including numbers), it makes sense to investigate the contribution of higher order n-gram language models for the morphological tag factor on the target side, to see if we can address the observation in the previous section.

Using the data transformed with *Noun+Adj+Verb+Adv+PostP* transformations which previously gave us the best results overall, we experimented with using higher order models (4-grams to 9-grams) during decoding, for the morphological tag factor models, keeping the surface and root models at 3-gram. We observed that for all the 10 data sets, the improvements were consistent for up to 8-gram. The BLEU with the 8-gram *for only the morphological tag factor* averaged over the 10 data sets was **22.61** (max: 23.66, min: 21.37, std: 0.72) compared to the 21.96 in Table 1. Using a 4-gram *root* LM, considerably less sparse than word forms but more sparse that tags, we get a BLEU score of **22.80** (max: 24.07, min: 21.57, std: 0.85). The details of the various BLEU scores are shown in the two halves of Table 3. It seems that larger n-gram LMs contribute to the larger n-gram precisions contributing to the BLEU but not to the unigram precision.

3-gram root LM		**1-gr.**	**2-gr.**	**3-gr.**	**4-gr.**
BLEU	22.61	55.85	28.21	17.16	11.36
BLEU-R	28.21	68.67	35.80	21.55	14.07
BLEU-M	28.68	67.50	37.59	22.02	14.22
4-gram root LM		**1-gr.**	**2-gr.**	**3-gr.**	**4-gr.**
BLEU	22.80	55.85	28.39	17.34	11.54
BLEU-R	28.48	68.68	35.97	21.79	14.35
BLEU-M	28.82	67.49	37.63	22.17	14.40

Table 3: Details of Word, Root and Morphology BLEU Scores, with 8-gram tag LM and 3/4-gram root LMs

3.2.4 Augmenting the Training Data

In order to alleviate the lack of large scale parallel corpora for the English–Turkish language pair, we experimented with augmenting the training data with reliable phrase pairs obtained from a previous alignment. Phrase table entries for the surface factors produced by Moses after it does an alignment on the roots, contain the English (e) and Turkish (t) parts of a pair of aligned phrases, and the probabilities, $p(e|t)$, the conditional probability that the English phrase is e given that the Turkish phrase is t, and $p(t|e)$, the conditional probability that the Turkish phrase is t given the English phrase is e. Among these phrase table entries, those with $p(e|t) \approx p(t|e)$ and $p(t|e) + p(e|t)$ larger than some threshold, can be considered as reliable mutual translations, in that they mostly translate to each other and not much to others. We extracted

from the phrase table those phrases with $0.9 \leq p(e|t)/p(t|e) \leq 1.1$ and $p(t|e) + p(e|t) \geq 1.5$ and added them to the training data to further bias the alignment process. The resulting BLEU score was **23.78** averaged over 10 data sets (max: 24.52, min: 22.25, std: 0.71).[16]

4 Experiments with Constituent Reordering

The transformations in the previous section *do not perform any constituent level reordering*, but rather *eliminate* certain English function words as tokens in the text and fold them into complex syntactic tags. That is, no transformations reorder the English SVO order to Turkish SOV,[17] for instance, or move postnominal prepositional phrase modifiers in English, to prenominal phrasal modifiers in Turkish. Now that we have the parses of the English side, we have also investigated a more comprehensive set of reordering transformations which perform the following constituent reorderings to bring English constituent order more in line with the Turkish constitent order at the top and embedded phrase levels:

- Object reordering (*ObjR*), in which the objects and their dependents are moved in front of the verb.
- Adverbial phrase reordering (*AdvR*), which involve moving post-verbal adverbial phrases in front of the verb.
- Passive sentence agent reordering (*PassAgR*), in which any post-verbal agents marked by *by*, are moved in front of the verb.
- Subordinate clause reordering (*SubCR*) which involve moving postnominal relative clauses or prepositional phrase modifers in front of any modifiers of the head noun. Similarly any prepositional phrases attached to verbs are moved to in front of the verb.

We performed these reorderings on top of the data obtained with the *Noun+Adj+Verb+Adv+PostP* transformations earlier in Section 3.2.2 and used the same decoder parameters. Table 4 shows the performance obtained after various combination of reordering operations over the 10 data sets. Although there were some improvements for certain cases, none of reordering gave consistent improvements for all the data sets. A cursory examinations of the alignments produced after these reordering transformations indicated that the resulting root alignments were not necessarily that close to being monotonic as we would have expected.

Experiment	Ave.	STD	Max.	Min.
Baseline	**21.96**	**0.72**	**22.91**	**20.67**
ObjR	21.94	0.71	23.12	20.56
ObjR+AdvR	21.73	0.50	22.44	20.69
ObjR+PassAgR	21.88	0.73	23.03	20.51
ObjR+SubCR	21.88	0.61	22.77	20.92

Table 4: BLEU scores of after reordering transformations

5 Related Work

Statistical Machine Translation into a morphologically rich language is a challenging problem in that, on the target side, the decoder needs to generate both the right sequence of constituents and the right sequence of morphemes for each word. Furthermore, since for such languages one can generate tens of hundreds of inflected variants, standard word-based alignment approaches suffer from sparseness issues. Koehn (2005) applied standard phrase-based SMT to Finnish using the Europarl corpus and reported that translation to Finnish had the worst BLEU scores.

Using morphology in statistical machine translation has been addressed by many researchers for translation from or into morphologically rich(er) languages. Niessen and Ney (2004) used morphological decomposition to get better alignments. Yang and Kirchhoff (2006) have used phrase-based backoff models to translate unknown words by morphologically decomposing the unknown source words. Lee (2004) and Zolmann et *al.* (2006) have exploited morphology in Arabic-English SMT. Popovic and Ney (2004) investigated improving translation quality from inflected languages by using stems, suffixes and part-of-speech tags. Goldwater and McClosky (2005) use morphological analysis on the Czech side to get improvements in Czech-to-English statistical machine translation. Minkov et *al.* (2007) have used morphological postprocessing on the target side, to improve translation quality. Avramidis and Koehn (2008) have annotated English with additional morphological information extracted from a syntactic tree, and have used this in translation to Greek and Czech. Recently, Bisazza and Federico (2009) have applied morphological segmentation in *Turkish-to-English* statistical machine translation and found that it provides nontrivial BLEU

[16] These experiments were done on top of the model in 3.2.3 with a 3-gram word and root LMs and 8-gram tag LM.

[17] Although Turkish is a free-constituent order language, SOV is the dominant order in text.

score improvements.

In the context of translation from English to Turkish, Durgar-El Kahlout and Oflazer (2010) have explored different representational units of the lexical morphemes and found that selectively splitting morphemes on the target side provided nontrivial improvement in the BLEU score. Their approach was based on splitting the target Turkish side, into constituent morphemes while our approach in this paper is the polar opposite: we do not segment morphemes on the Turkish side but rather join function words on the English side to the related content words. Our approach is somewhat similar to recent approaches that use complex syntactically-motivated complex tags. Birch et *al.* (2007) have integrated more syntax in a factored translation approach by using CCG supertags as a separate factor and have reported a 0.46 BLEU point improvement in Dutch-to-English translations. Although they used supertags, these were obtained not via syntactic analysis but by supertagging, while we determine, on the fly, the appropriate syntactic tags based on syntactic structure. A similar approach based on supertagging was proposed by Hassan et *al.* (2007). They used both CCG supertags and LTAG supertags in Arabic-to-English phrase-based translation and have reported about 6% relative improvement in BLEU scores. In the context of reordering, one recent work (Xu et al., 2009), was able to get an improvement of 0.6 BLEU points by using source syntactic analysis and a constituent reordering scheme like ours for English-to-Turkish translation, but without using any morphology.

6 Conclusions

We have presented a novel way to incorporate source syntactic structure in English-to-Turkish phrase-based machine translation by parsing the source sentences and then encoding many local and nonlocal source syntactic structures as additional complex tag factors. Our goal was to obtain representations of source syntactic structures that parallel target morphological structures, and enable us to extend factored translation, in applicability, to languages with very disparate morphological structures.

In our experiments over a limited amount training data, but repeated with 10 different training and test sets, we found that syntax-to-morphology mapping transformations on the source side sentences, along with a very small set of transformations on the target side, coupled with some additional techniques provided about 39% relative improvement in BLEU scores over a word-based baseline and about 28% improvement of a factored baseline. We also experimented with numerous additional syntactic reordering transformation on the source to further bring the constituent order in line with the target order but found that these did not provide any tangible improvements when averaged over the 10 different data sets.

It is possible that the techniques presented in this paper may be less effective if the available data is much larger, but we have reasons to believe that they will still be effective then also. The reduction in size of the source language side of the training corpus seems to be definitely effective and there no reason why such a reduction (if not more) will not be observed in larger data. Also, the preprocessing of English prepositional phrases and many adverbial phrases usually involve rather long distance relations in the source side syntactic structure[18] and when such structures are coded as complex tags on the nominal or verbal heads, such long distance syntax is effectively "localized" and thus can be better captured with the limited window size used for phrase extraction.

One limitation of the approach presented here is that it is not directly applicable in the reverse direction. The data encoding and set-up can directly be employed to generate English "translation" expressed as a sequence of root and complex tag combinations, but then some of the complex tags could encode various syntactic constructs. To finalize the translation after the decoding step, the function words/tags in the complex tag would then have to be unattached and their proper positions in the sentence would have to be located. The problem is essentially one of generating multiple candidate sentences with the unattached function words ambiguously positioned (say in a lattice) and then use a second language model to rerank these sentences to select the target sentence. This is an avenue of research that we intend to look at in the very near future.

Acknowledgements

We thank Joakim Nivre for providing us with the parser. This publication was made possible by the generous support of the Qatar Foundation through Carnegie Mellon University's Seed Research program. The statements made herein are solely the responsibility of the authors.

[18] For instance, consider the example in Figure 2 involving *if* with some additional modifiers added to the intervening noun phrase.

References

Eleftherios Avramidis and Philipp Koehn. 2008. Enriching morphologically poor languages for statistical machine translation. In *Proceedings of ACL-08/HLT*, pages 763–770, Columbus, Ohio, June.

Alexandra Birch, Miles Osborne, and Philipp Koehn. 2007. CCG supertags in factored translation models. In *Proceedings of SMT Workshop at the 45th ACL*.

Arianna Bisazza and Marcello Federico. 2009. Morphological pre-processing for Turkish to English statistical machine translation. In *Proceedings of the International Workshop on Spoken Language Translation*, Tokyo, Japan, December.

İlknur Durgar-El-Kahlout and Kemal Oflazer. 2010. Exploiting morphology and local word reordering in English to Turkish phrase-based statistical machine translation. *IEEE Transactions on Audio, Speech, and Language Processing*. To Appear.

Alexander Fraser. 2009. Experiments in morphosyntactic processing for translating to and from German. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 115–119, Athens, Greece, March. Association for Computational Linguistics.

Sharon Goldwater and David McClosky. 2005. Improving statistical MT through morphological analysis. In *Proceedings of HLT/EMNLP-2005*, pages 676–683, Vancouver, British Columbia, Canada, October.

Hany Hassan, Khalil Sima'an, and Andy Way. 2007. Supertagged phrase-based statistical machine translation. In *Proceedings of the 45th ACL*, pages 288–295, Prague, Czech Republic, June. Association for Computational Linguistics.

Philipp Koehn and Hieu Hoang. 2007. Factored translation models. In *Proceedings of EMNLP*.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT/NAACL-2003*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th ACL–demonstration session*, pages 177–180.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit X*.

Young-Suk Lee. 2004. Morphological analysis for statistical machine translation. In *Proceedings of HLT/NAACL-2004 – Companion Volume*, pages 57–60.

Einat Minkov, Kristina Toutanova, and Hisami Suzuki. 2007. Generating complex morphology for machine translation. In *Proceedings of the 45th ACL*, pages 128–135, Prague, Czech Republic, June. Association for Computational Linguistics.

Sonja Niessen and Hermann Ney. 2004. Statistical machine translation with scarce resources using morpho-syntatic information. *Computational Linguistics*, 30(2):181–204.

Joakim Nivre, Hall Johan, Nilsson Jens, Chanev Atanas, Gülşen Eryiğit, Sandra Kübler, Marinov Stetoslav, and Erwin Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering Journal*, 13(2):99–135.

Kemal Oflazer and İlknur Durgar-El-Kahlout. 2007. Exploring different representational units in English-to-Turkish statistical machine translation. In *Proceedings of Statistical Machine Translation Workshop at the 45th Annual Meeting of the Association for Computational Linguistics*, pages 25–32.

Kemal Oflazer. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2):137–148.

Kemal Oflazer. 2008. Statistical machine translation into a morphologically complex language. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, pages 376–387.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2001. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th ACL*, pages 311–318.

Maja Popovic and Hermann Ney. 2004. Towards the use of word stems and suffixes for statistical machine translation. In *Proceedings of the 4th LREC*, pages 1585–1588, May.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of HLT/NAACL-2003*, pages 252–259.

Peng Xu, Jaeho Kang, Michael Ringgaard, and Franz Och. 2009. Using a dependency parser to improve SMT for subject-object-verb languages. In *Proceedings HLT/NAACL-2009*, pages 245–253, June.

Mei Yang and Katrin Kirchhoff. 2006. Phrase-based backoff models for machine translation of highly inflected languages. In *Proceedings of EACL-2006*, pages 41–48.

Deniz Yuret and Ferhan Türe. 2006. Learning morphological disambiguation rules for Turkish. In *Proceedings of HLT/NAACL-2006*, pages 328–334, New York City, USA, June.

Andreas Zollmann, Ashish Venugopal, and Stephan Vogel. 2006. Bridging the inflection morphology gap for Arabic statistical machine translation. In *Proceedings of HLT/NAACL-2006 – Companion Volume*, pages 201–204, New York City, USA, June.