Exploiting Morphology and Local Word Reordering in English-to-Turkish Phrase-Based Statistical Machine Translation

İlknur Durgar El-Kahlout and Kemal Oflazer

***Abstract*—In this paper, we present the results of our work on the development of a phrase-based statistical machine translation prototype from English to Turkish—an agglutinative language with very productive inflectional and derivational morphology. We experiment with different morpheme-level representations for English–Turkish parallel texts. Additionally, to help with word alignment, we experiment with local word reordering on the English side, to bring the word order of specific English prepositional phrases and auxiliary verb complexes, in line with the morpheme order of the corresponding case-marked nouns and complex verbs, on the Turkish side. To alleviate the dearth of the parallel data available, we also augment the training data with sentences just with content word roots obtained from the original training data to bias root word alignment, and with highly reliable phrase-pairs from an earlier corpus alignment. We use a morpheme-based language model in decoding and a word-based language model in re-ranking the n-best lists generated by the decoder. Lastly, we present a scheme for *repairing* the decoder output by *correcting* words which have incorrect morphological structure or which are out-of-vocabulary with respect to the training data and language model, to further improve the translations. We improve from 15.53 BLEU points for our word-based baseline model to 25.17 BLEU points for an improvement of 9.64 points or about 62% relative.**

***Index Terms*—Complex morphology, English, statistical machine translation (SMT), Turkish, word reordering.**

I. Introduction

MACHINE translation is one of the major, oldest, and the most active areas in natural language processing. Following the lack of success of the earlier symbolic or rule-based approaches in developing wide-coverage machine translation systems [1], the availability of large amounts of parallel electronic texts and the increase in the computational power have motivated researchers to shift from rule-based to corpus-based paradigms. The major paradigm in machine translation in nearly the last twenty years has been *statistical machine translation* (SMT) that started with the seminal work at IBM [2], [3]. Statistical machine translation continues to be a very active research area, continually bringing in new techniques, additional sources of information, refinements, and language(pair)-specific variations.

Most recent statistical machine translation approaches rely on the language model to enforce word-order constraints in target language sentences, but differ on how they formulate the translation model. Early approaches have employed a word-based approach, treating words as translation units [3]. Later, phrase-based approaches have used "phrases," which, in this context, denote any sequence of tokens (that may or may not be linguistically meaningful), and have modeled the translation model as combinations of other component models[4]–[8]. More recently, factored models [9], have considered exploiting richer linguistic information such as word roots, parts-of-speech and other morphological information. Currently, there is substantial work in exploiting syntactic information on the source [10], or the target side [11] or on both sides of the translation ([12], [13]).

Statistical translation decoders essentially take a source sentence and segment it into all possible words/phrases. These are then translated and moved around into many possible target language word/phrase sequences with probabilities provided by the components of the translation model, and the resulting target translations are scored by the language model. As the set of possible target sentences is huge, the search process is guided by many heuristics that prune the search space to highly likely left-to-right generated (prefixes of) candidate sentences.

In this paper, we present the results of our work on the development of a phrase-based statistical machine translation prototype from English to Turkish. This problem is interesting from a number of perspectives. Typologically English and Turkish are rather distant languages for which rather modest parallel text data exists. Most importantly, Turkish has complex agglutinative morphology with word structures that can correspond to complete phrases of several words in English when translated.

In our experiments:

1) we investigate how different representations of morphology on both the English and the Turkish sides impact statistical translation results;
2) to help with word alignment, we experiment with local word ordering on the English side to bring the word order of specific English prepositional phrases and verb

Manuscript received February 06, 2009; revised September 09, 2009. First published September 29, 2009; current version published July 14, 2010. This work was supported by the TÜBİTAK (Turkish Scientific and Technological Research Foundation) under Project 105E020 "Building a Statistical Machine Translation from English-to-Turkish." "Şeyma Mutlu of Sabancı University implemented the word-repair code. Some of this work was done while the second author was on sabbatical leave at the Language Technologies Institute at Carnegie Mellon University in Pittsburgh, PA. The authors are grateful for their support during this period. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dimitra Vergyri.

İ. D. El-Kahlout was with the Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı, İstanbul, 30332 Turkey. She is now with the LIMSI-CNRS, F-91403 Orsay, France (e-mail: ilknur.durgar@limsi.fr).

K. Oflazer was with the Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı, İstanbul, 30332 Turkey. He is now with Carnegie Mellon University Qatar, Doha, Qatar (e-mail: ko@cs.cmu.edu).

Digital Object Identifier 10.1109/TASL.2009.2033321

complexes in line with the morpheme order of the corresponding case marked noun forms and verbs on the Turkish side;

3) to alleviate the dearth of the parallel data available, we also augment the training data with sentences containing just the content word roots, obtained from the original training data, and with highly reliable phrase-pairs from an earlier alignment;
4) during decoding, we use a morpheme-based language model, as decoding produces morphemes or groups of morphemes as output;
5) the n-best lists produced by the decoder are then re-ranked with a word-based language model, so that longer range constraints can be incorporated, in addition to the more local morphotactic sequencing constraints provided by the morpheme-based language model;
6) lastly, we present a scheme for *repairing* the decoder output by *correcting* words which have incorrect morphological structure or which are out-of-vocabulary with respect to the training data and language model to further improve the translations.

All in all, we improve from 15.53 BLEU points for our word-based baseline model to 25.17 BLEU points for an improvement of 9.64 points or about 62% relative, using a selectively segmented morphemic representation with various additional steps. We also find that incorporating derivational morphology on the English side does not provide any improvement.

This paper is organized as follows. We first discuss the issues in English-to-Turkish statistical machine translation. We then discuss how we can exploit morphology in phrase-based statistical machine translation and present results a basic experimental setting, and after using local word reordering, English derivational morphology and word-repair. Finally, we conclude with a summary of our main results.

II. English-to-Turkish SMT

Statistical machine translation from English to Turkish poses a number of difficulties. Typologically, English and Turkish are rather distant languages: while English has very limited morphology and a rather rigid Subject-Verb-Object (SVO) constituent order, Turkish is an agglutinative language with a very rich and productive derivational and inflectional morphology, and a very flexible (but Subject-Object-Verb (SOV) dominant) constituent order.

Initial experiments [14] showed that when English–Turkish parallel data were aligned at the word level, a Turkish word would typically have to align with a complete phrase on the English side, and that sometimes these phrases on the English side needed to be discontinuous. These observations suggested that exploiting sub-lexical structure would be a fruitful avenue to pursue. For instance, the Turkish word *tatlandirabileceksek* could be translated as (and hence would have to be aligned to a potentially discontinuous phrase equivalent to) "if we were going to be able to make [something] acquire flavor." These could be aligned as follows (shown with co-indexation of Turkish surface morphemes and English words):[1]

$$(\text{tat})_1(\text{lan})_2(\text{dir})_3(\text{abil})_4(\text{ecek})_5(\text{se})_6(\text{k})_7$$
$$(\text{if})_6(\text{we are})_7(\text{going to})_5(\text{be able})_4(\text{to make})_3$$
$$[\text{something}](\text{acquire})_2(\text{flavor})_1.$$

The productive morphology of Turkish implies potentially a very large vocabulary size: noun roots have about 100 inflected forms and verbs have much more [15]. These numbers are much higher when derivations are considered; one can generate thousands of words from a single root when, say, only at most two derivations are allowed.[2] For example, the root word *aşama* (rank, phase, stage) occurs in 30 different forms with a total count of 460 in the parallel texts we experiment with. Of these 30 such forms, only 3 appear over 50 times. However, Turkish employs about 30 000 root words (about 10 000 of which are highly frequent) and about 150 distinct suffixes. Therefore, considering sublexical units in the parallel texts can alleviate the sparseness problem to some extent.

Moreover, for accurate estimation of translation model parameters, one needs large amounts of data which for the English–Turkish language pair has not been easy to acquire with no substantial improvement expected in the near future, as there are not many electronically available sources of such texts. Thus, we have to exploit our available resources maximally, instead of relying on future availability of more data.

In the context of agglutinative languages similar to Turkish (at least with respect to morphological aspects), there has been some recent work on translating to and from Finnish with the data in the Europarl corpus [16]. In this work, although the BLEU score [17] from Finnish to English was 21.8, the score in the reverse direction was reported as 13.0, which is one of the lowest scores for the languages covered in the Europarl data set. Also, the reported *from* and *to* translation scores for Finnish are the lowest on average, even with a million parallel sentences available. These may hint at the fact that standard approaches are perhaps poorly equipped to deal with translation from a morphologically poor language like English to a morphologically complex language like Finnish or Turkish.

III. Exploiting Morphology

Using morphology in statistical machine translation has been addressed by many researchers for translation from or into morphologically rich(er) languages, to have a better estimations for the parameters of the translation model and also to rely on smaller parallel texts. Niessen and Ney [18] use morphological decomposition to improve word alignment quality. Yang and Kirchhoff [19] use phrase-based backoff models to translate words that are unknown to the decoder, by morphologically decomposing the unknown source words. Corston–Oliver and Gamon [20] normalize inflectional morphology by stemming the word for German–English word alignment. Lee [21] uses

[1]Note that on the English side, the noun phrase filler for [something] would come in the middle of this phrase, while the corresponding Turkish noun phrase would have to come somewhere before the Turkish word.

[2]A recent 125 M word Turkish corpus that we have collected has about 1.5 M distinct word forms. This is almost the same number of distinct word forms in the English Gigaword Corpus which is about 15 times larger.

a morphologically analyzed and tagged parallel corpus for Arabic–English statistical machine translation. Zolmann *et al.* [22] also exploit morphology in Arabic–English statistical machine translation. Popovic and Ney [23] investigate improving translation quality from inflected languages by using stems, suffixes and part-of-speech tags. Goldwater and McClosky [24] use morphological analysis on the Czech side to get improvements in Czech-to-English statistical machine translation. Recently, Minkov *et al.* [25] have used morphological postprocessing *on the target side* using structural information and information from the source side, to improve translation quality.

Our approach in this paper is to represent Turkish words with their morphological segmentation. At the morpheme level, we split the Turkish words into their lexical morphemes, while English words with overt morphemes are stemmed, and such morphemes are marked with a tag. We do not attempt to do derivational morphology on the English side for the initial set of experiments.

The reason we use lexical morphemes instead of surface morphemes is that most surface distinctions are manifestations of word-internal phenomena such as vowel harmony, and morphotactics. Allomorphs which differ in phonological form almost always correspond to the same set of words/tags in English when translated. When surface morphemes are considered by themselves as units in alignment, statistics get fragmented. For example, Turkish has two different surface morphemes, $+ler$ and $+lar$, that mark plurality and both are translated to $+s$ in English side. With lexical morpheme representation, we can abstract away such word-internal details *and* conflate statistics for seemingly different suffixes, as at this level of representation, words that look very different on the surface, look very similar.[3] For example, although the words *evinde* (in his/her house) and *masasında* (on his/her table) look quite different, the lexical morphemes except for the root are the same:[4] $\mathrm{ev{+}sh{+}nda}$ versus $\mathrm{masa{+}sh{+}nda}$ (see Oflazer and Durgar-El Kahlout [27] for details). We could also have used a full morphological feature representation for our words where the only surface feature would be the root word. We opted not to do this mainly because our feature representation has many more additional morphosemantic features and represents default covert morphological processes overtly. Further, the lexical morpheme representation is easier to *read* when we look at the outputs. Otherwise, the representations are equivalent.

We should however note that although employing a morpheme based representation dramatically reduces the vocabulary size on the Turkish side, it also runs the risk of overloading the decoder mechanisms to account for *both* word-internal morpheme sequencing and sentence level word ordering.

We process the parallel text in the following fashion.

1) We segment the words in our Turkish corpus into lexical morphemes whereby differences in the surface representations of morphemes due to word-internal phenomena are abstracted out to improve statistics during alignment. Note that, as with many similar languages, the segmentation of a surface word is generally ambiguous. Hence, we first generate a representation using our morphological analyzer [28] that contains both the lexical segments and the morphological features encoded for all possible segmentations and interpretations of the word, and perform morphological disambiguation using morphological features [29]. Once the contextually salient morphological interpretation is selected, we remove the morphological features, leaving behind the lexical morphemes making up a word.
2) We tag the English side using TreeTagger [30], which provides a lemma and a *part-of-speech tag* for each word. We then remove any tags which do not imply an explicit morpheme or an exceptional form. For instance, if the word *book* is tagged as $+NN$, we keep *book* in the text, but remove $+NN$. For *books* tagged as $+NNS$ or *booking* tagged as $+VVG$, we keep *book* and $+NNS$, and *book* and $+VVG$. A word like *went* is replaced by $go+VVD$.[5]
3) From these morphologically segmented corpora, we also extract, for each sentence, the sequence of roots for open class content words (nouns, adjectives, adverbs, and verbs). For Turkish, this corresponds to removing *all* morphemes and any roots for closed classes. For English, this corresponds to removing all words tagged as closed class words, along with the tags such as $+VVG$ above that signal a morpheme on an open class content word. We use this to augment the training corpus and bias content root word alignments, with the hope that such roots may get a better chance to align without any additional *noise* from morphemes and other function words.

A typical sentence pair in our (fully segmented) data looks like the following, where we have highlighted the content root words with boldface font, and have co-indexed them to show their actual alignments. A copy of the sentences with these boldface tokens would comprise the content root word corpus that we use to augment the original training data. Test and reference sentences are also segmented in the same manner.[6]

$\mathbf{T}$: [$\mathbf{kat}_1$ +hl +ma] [$\mathbf{ortaklık}_2$+sh +nhn] [$\mathbf{uygula}_3$ +hn +ma +sh] [,] [$\mathbf{ortaklık}_4$] [$\mathbf{anlasma}_5$ +sh] [$\mathbf{cerceve}_6$ +sh +nda] [$\mathbf{izle}_7$ +hn +yacak+dhr] [.]

$\mathbf{E}$: the $\mathbf{implementation}_3$ of the $\mathbf{accession}_1$ $\mathbf{partnership}_2$ will be $\mathbf{monitor}_7$ +vvn in the $\mathbf{framework}_6$ of the $\mathbf{association}_4$ $\mathbf{agreement}_5$.

Note that when the morphemes/tags (tokens starting with a +) are concatenated, we get the *word-based* version of the corpus (that is, tokens in brackets above, are concatenated to make up these *words*). Since surface words are directly recoverable from the morpheme-concatenated representation and it is these words that BLEU evaluation on test would consider. We use this word-based representation also for word-based language models used for rescoring.

Table I presents statistical information about the parallel corpus. One can note there is a difference between the number of sentences in the basic training set and the content root word

[3]This is in a sense very similar to the more general problem of lexical redundancy addressed by Talbot and Osborne [26], but our approach does not require the more sophisticated solution there.

[4]Here, in the morphemes, *h* stands for high-vowels and *a* stands for low-unrounded vowels.

[5]A list of English tags used in the paper is provided in the appendix.

[6]Note that the reference sentences are only used during BLEU evaluation and then as full words where any morphemes are attached to the stem to the left.

TABLE I
STATISTICS ON TURKISH AND ENGLISH TRAINING AND TEST DATA, AND TURKISH MORPHOLOGICAL STRUCTURE.

TURKISH	Sent.	Words	Uniq. Words	Ave. Sent. Length
Train	44,709	545,952	52,338	12.21
Train-Content	56,609	437,208	13,767	7.7
Tune	200	3,258	1,442	16.29
Test	1000	11,991	4,828	11.90
ENGLISH				
Train	44,709	684,609	30,790	15.31
Train-Content	56,609	403,162	19,791	7.12
Tune	200	4,377	1,657	21.8
Test	1000	14.861	3,773	14.86

TURKISH	Morph-emes	Uniq. Morp.	Morp./ Word	Uniq. Roots	Uniq. Suff.
Train	983,303	15,014	1.43	14,892	122
Tune	6,240	859	1.42	810	49
Test	21,777	2,409	1.46	2,337	72

training set. This is due to the maximum token limit of the GIZA++ alignment tool used [31]. We removed sentences that exceed 90 *morphemes* from the basic training set in order for GIZA++ not to have any problems. However, such sentences were included in the content root word corpus, since with all the morphemes removed, their lengths did not exceed the limit. As we just kept the roots for open class content words, we observe a slight drop in the number of unique words in the content set. The final training set is the combination of basic set and the content words. One can also note that Turkish has many more distinct word forms (about twice as many as English), but has less number of distinct root words than English.

A. Experimental Framework

We employ the phrase-based statistical machine translation framework [6], and use the Moses toolkit [32], and the SRILM language modeling toolkit [33], and evaluate our decoded translations using the BLEU measure, using a *single* reference translation.

The collection of parallel texts that we have used for this work were mostly from the legal and diplomatic relations domain collected from NATO, EU, and foreign ministry sources. There is also a limited amount data parallel news corpus available from certain news sources. The parallel corpus was aligned at the sentence level using Microsoft Research Bilingual Sentence Alignment Tool.[7]

A 5-gram morpheme-based language model was constructed for Turkish (to be used by the decoder) using the Turkish side of the training data along with an additional monolingual Turkish text of about 50 K sentences (from the news domain). Training was performed with standard features and the phrase table was extracted using a maximum phrase size of 7. The test corpus was decoded using the Moses decoder with two different parameter settings: with default parameters in Moses and modified parameters *-dl*[8] -1 to allow for long distance movement and *-weight-d*[9] 0.1, to avoid penalizing any long distance movement.[10] Minimum error rate training [34], with the tune set did not provide any tangible improvements. We ran MERT on the baseline model and the morphologically segmented models forcing *-weight-d* to range around 0.1, but letting the other parameters range in their suggested ranges. Even though the procedure came back claiming that it achieved a better BLEU score on the tune set, running the new model on the test set did not show any improvement at all. This may have been due to the fact that the initial choice of *-weight-d* along with *-dl* set to -1 provides such a drastic improvement, so that perturbations in the other parameters do not have much impact, and to the fact that since the decoder produces morpheme sequences, MERT optimizes morpheme BLEU while we evaluate the final outcome with word BLEU.[11]

The decoder also produced 1000-best candidate translations. The 1000-best lists were rescored using a combination of the 4-gram *word-based* language model score and the translation score produced by the decoder weighted equally.

B. Representational Experiments

We performed four initial sets of experiments employing different morphological representations on the Turkish side and wherever needed, adjusting the English representation accordingly. For each experiment, the training and the test sentences, and additional language model training sentences for Turkish, all had the same morphological representation scheme, but in all experiments, BLEU was computed on the word-based representation.

1) **Baseline:** English and Turkish sentences are represented with *full* words. For example, the segmented form kitap+sh+nhn, representing *kitabının* (of his book) would be used on the Turkish side and book+NNS (representing *books*) on the English side.
2) **Full Morphological Segmentation:** English and Turkish sentences are represented by tokens representing root words and bound morphemes/tags. For the examples above, the three tokens kitap+sh+nhn would be used on the Turkish side and the two tokens book+NNS on the English side.
3) **Root+Morphemes Segmentation:** Turkish sentences are represented with roots and combined morphemes. For English sentences, we used the same representation in (2).

[7]Available at http://research.microsoft.com/~bobmoore/.

[8]-dl stands for distortion limit and -1 lets an unlimited reordering (default value is 6)

[9]-weight-d stands for distortion weight and penalizes long distance reordering. A high value penalizes long distance reordering more (default value is 0.3)

[10]We arrived at this combination by experimenting with the decoder to avoid the almost monotonic translation we were getting with the default parameters. These parameters boosted the BLEU scores substantially compared to default parameters used by the decoder.

[11]Some of the reviewers also suggested that this may be because of a small tuning set.

TABLE II
TURKISH MORPHEMES AND UNALIGNMENT PERCENTAGES

Morpheme	Count	Unalignment Percentage
+sh	152618	93.41
+lar,+larh	66837	20.92
+da,+nda	45620	58.26
+hl	44843	54.95
+ma	41835	86.58
+dhr (copula)	29664	75.20
+dhr (causative)	20732	83.78
+mhs	16294	19.85
+lhk	12557	77.05
+yla	8832	70.09
+dhk	7581	54.21
+ma	6148	42.20
+yacak	5105	32.67

For example for the Turkish word above, the two tokens kitap $+\mathrm{sh}+\mathrm{nhn}$ would be used.

4) **Selective Morphological Segmentation:** A systematic analysis of the alignment files produced by GIZA++ for the training sentences showed that certain morphemes on the Turkish side were almost consistently never aligned with anything on the English side, or were aligned more or less randomly. For example, the compound noun marker morpheme in Turkish $(+\mathrm{sh})$ does not have a corresponding unit on the English side, as English noun–noun compounds do not carry any overt markers. Further, since we perform derivational morphological analysis on the Turkish side but not on the English side, we also noted that most verbal nominalizations on the English side were just aligned to the verb roots on the Turkish side and the additional markers on the Turkish side indicating the nominalization, and various agreement markers, etc., were mostly unaligned.
From the word alignments, we selected unaligned morphemes with unalignment percentage over $\%80$ and attached such morphemes (and in the case of verbs, the intervening morphemes) to the root. Otherwise, we kept other morphemes, especially any case morphemes, still separate, as they almost often align with prepositions on the English side quite accurately. It should be noted that what to selectively attach to the root should be considered on a per-language basis; if Turkish were to be aligned with a language with similar morphological markers, this perhaps would not have been needed. Again one perhaps can use methods similar to those suggested by Talbot and Osborne [26]. Table II shows some highly frequent morphemes and their unalignment percentages.
Thus in this representation, the Turkish word above would be represented by the two tokens $\mathrm{kitap}+\mathrm{sh}$ $+\mathrm{nhn}$. English words are represented as in the second case above.

The results of these set of experiments are presented in Tables III and IV. The best BLEU results is obtained with selective morphological segmentation (22.81) and represents a relative improvement of 46.8%, compared to the respective baseline of 15.53. One should also note that the default decoding parameters used by the Moses decoder produces much worse results especially for the fully segmented model. Our further experiments below are performed on top of the results of the best performing representation—selective morphological segmentation with modified parameters and train-content.

TABLE III
BLEU RESULTS FOR THE BASELINE REPRESENTATION

EXPERIMENTS	BLEU	
	Default Param.	Modified Param.
Word-based	15.53	20.16
Word-based + Train-Content	15.41	19.79

TABLE IV
BLEU RESULTS FOR THE MORPHEMIC REPRESENTATIONS

SEGMENTATIONS	Default Param.		Modified Param.	
	BLEU	BLEU After Rescoring	BLEU	BLEU After Rescoring
Full morphological	14.33	14.51	19.37	19.96
Full morphological + Train-Content	14.48	14.55	19.82	20.83
Root+Morphemes	15.16	15.61	19.91	21.00
Root+Morphemes + Train-Content	15.47	16.14	20.09	21.62
Selective morphological	15.48	15.66	21.92	22.23
Selective morphological + Train-Content	15.73	16.10	22.20	**22.81**

C. Local Word Reordering

It has been observed that one gets better alignments and hence better translation results when the word orders of the source and target languages are more or less the same. When word orders are different, researchers have tried systematically reordering the tokens of source sentences to an order matching or very close to the target language word order, so that alignments could be very close to a monotonic one. Thus, instead of forcing the decoders to employ reordering schemes, the source sentences are similarly reordered and then decoded with the decoder employing simpler reordering models.

A number of previous studies have addressed the use of morpho-syntactic information in reordering schemes. Brown *et al.* [2] reorder phrases with the help of a preprocessor. Xia and McCord [35] derive reordering patterns from word alignments and use these patterns in monotonic decoding. Niessen and Ney [18] focus on reordering separated German verb prefixes and question inversion by using part-of-speech tags. Collins *et al.* [36] use handwritten rules for reordering German clauses. Popovic and Ney [37] reorder adjectives in English–Spanish SMT by using part-of-speech tags. Recently, Wang *et al.* [38] show improvement in Chinese–English translation by using

Penn Chinese Treebank phrase types. Zwarts and Dras [39] reorder source sentence words by minimizing the dependency distance between the head and the dependent.

Our goal is *not to attempt a full reordering at the sentence constituent level*. Instead, we have a more modest goal of a very local source word reordering for a certain class of phrases so that the word order in an English phrase has a more or less monotonic alignment with the morpheme order of the corresponding morphologically marked Turkish word.

To motivate such reordering, we present the following examples where we have shown morphemes on both sides by prefixing them with $+$.

- Turkish noun forms with cases other than nominative case typically correspond to (parts of) prepositional phrases in English. For example,

$$\begin{aligned} &\text{in}_1\ \text{my}_2\ \text{long}_3\ \text{story}_4{+}\text{s}_5 \leftrightarrow \text{uzun}_3 \\ &\text{hikaye}_4{+}\text{ler}_5{+}\text{im}_2{+}\text{de}_1 \end{aligned}$$

  a reordering of the function words in the English prepositional phrases leads to

$$\begin{aligned} &\text{long}_3\ \text{story}_4{+}\text{s}_5\ \text{my}_2\ \text{in}_1 \leftrightarrow \text{uzun}_3 \\ &\text{hikaye}_4{+}\text{ler}_5{+}\text{im}_2{+}\text{de}_1 \end{aligned}$$

  in which both the source (word) and the target (morpheme) tokens are monotonically aligned. The case of *of* presents special difficulties: noun phrases on both sides of *of* have to be identified and swapped, that is NP_1 *of* NP_2 is reordered to NP_2 *of* NP_1, to match the ordering on the Turkish side. Note that if the first NP_1 is part of a prepositional phrase, that would have to be reordered first.
- English auxiliary verb complexes and infinitive forms can be reordered to monotonically align to Turkish verb forms or Turkish infinitives. For example, in

$$\text{will}_1\ \text{be}_2\ \text{monitor}_3 + \text{ed}_4 \leftrightarrow \text{izle}_3 + \text{n}_4 + \text{ecek}_1 + \text{tir}_2$$

  a reordering of the auxiliary verb components leads to

$$\text{monitor}_3 + \text{ed}_4\ \text{will}_1\ \text{be}_2 \leftrightarrow \text{izle}_3 + \text{n}_4 + \text{ecek}_1 + \text{tir}_2$$

  in which again both the source and the target tokens are monotonically aligned.

To investigate the impact of such local reordering, we selected nine prepositions (*of, in, from, to, for, on, at, under, into*) occurring with high frequency on the English side of the training data and extracted prepositional phrases headed by those prepositions (of up to four tokens) and reordered these phrases. We do not actually fully parse the sentences. Our sentences are already tagged with parts-of-speech and we are essentially bracketing short PPs of up to four tokens on the English side only using part-of-speech information. The idea here is that a PP with one determiner/possessor and possibly a plural marker would most of the time have the same components of a case-marked Turkish noun with a possessor and a plural marker like the example earlier.

We extract rewrite patterns as follows. For each selected phrase type, we search the source language sentence and count the occurrences of patterns. For patterns occurring ten or more times, we start from the longest pattern, process the source language text in a left-to-right fashion and reorder phrases that match the patterns.

TABLE V
REWRITE PATTERNS AND FREQUENCIES

Before → After	**Frequency**
+in +dt +nn → +dt +nn +in	2909
+in +dt +jj +nn → +dt +jj +nn +in	1708
+in +nn → +nn +in	1465
+in +dt +nn +nn → +dt +nn +nn +in	584
+in +cd → +cd +in	491
+from +cd +nn → +cd +nn +from	71
+from +dt +np +np → +dt +np +np +from	94
+from +dt +nn+nns → +dt +nn+nns +from	53

For prepositional phrases, except the preposition of, we search patterns in the form of $preposition\ tag_1 tag_2 \cdots tag_i$ up to length 4. For nouns, the root and any plural marker are kept, any preceding possessive pronoun is placed after these two, and the preceding preposition is placed after the possessive pronoun. The case of `of` presents special difficulties: of maps to an explicit case morpheme no so frequently. For example, in NPs like `The Queen of England` the `of` does not map to a genitive morpheme on the Turkish equivalent of England. Moreover, noun phrases on both sides of `of` have to be identified and swapped, that is NP_1 of NP_2 is reordered to NP_2 of NP_1, to match the ordering on the Turkish side. Note that if the first NP_1 is part of a prepositional phrase, it has to be reordered first. The situation becomes more complicated with any errors in the bracketing of the two NPs on each side.

For preposition *of* we search patterns in the form of $of\_PP = tag_1 tag_2 \cdots tag_i\ of\ tag_1 tag_2 \cdots tag_j$ up to length 4. For preposition `of`, the first step of extraction procedure should obtain patterns also by checking preceding tags. We then swap the preceding and following tag groups.

Table V shows some of the most frequent rewrite patterns.

In addition to these local reorderings, we remove the determiner *the* from the English side as there is never a counterpart on the Turkish side except when the NP it is associated with is used as the object of a transitive verb in which case it gets an accusative case marker.[12]

As a result of these local reordering and removal of *the*, the aligned sentence pair given earlier (and with selective segmentation already applied), now looks like with aligned tokens co-indexed. For the English sentences, we have indicated with bracketing, the internally reordered phrases.

$\mathbf{T}$: $\text{kat}{+}\text{hl}{+}\ \text{ma}_1\ \text{ortaklık}{+}\text{sh}_2\ {+}\text{nhn}_3\text{uygula}{+}\text{hn}{+}\text{ma}$
${+}\text{sh}_{4,5}\ \text{ortaklık}_6\ \text{anlas}{+}\text{ma}{+}\text{sh}_7$
$\text{cerceve}{+}\text{sh}_8{+}\text{nda}_9\ \text{izle}_{10}\ {+}\text{hn}_{11}{+}\text{yacak}_{12}{+}\text{dhr}_{13\text{-}14}$
$\mathbf{E}$ $(\mathbf{Before})$:{the implementation of the accession partnership} {will be monitor + vvn} {{in the framework} of the association agreement}.

[12]On the contrary, the determiner *a* always has a counterpart.

$\mathbf{E}$ (**After**) : $\{\text{accession}_1\ \text{partnership}_2\ \text{of}_3\ \text{implementation}_4\}\ \{\text{monitor}_{10} + \text{vvn}_{11}\ \text{will}_{12}\ \text{be}_{13}\}\ \{\text{association}_6\ \text{agreement}_7\ \text{of}\ \{\text{framework}_8\ \text{in}_9\}\}$.[14]

Note that the top level phrasal constituent orders are still different (SOV versus SVO) but within each constituent, the alignments are monotonic, to the extent possible.

For experimentation, we considered different subsets of the transformations above:[13]

- in *prep1*, prepositional phrases headed all prepositions except *of*, were reordered;
- in *prep2*, prepositional phrases headed by all nine prepositions were reordered;
- in *inf*, infinitive verb constructs (headed by *to*) were reordered;
- in *the*, the determiner *the* was dropped;
- in *verb*, all auxiliary verb sequences were reordered.

The experiments of these transformations were carried on top of the setup giving the best result of 22.81 BLEU in Table IV. Table VI shows the results of experiments with various combinations of the transformations above. The best results have been obtained with the local ordering of the prepositional phrases headed by prepositions in the set prep1, the removal of the determiner *the* and reordering of the infinitive constructs.

D. Incorporating English Derivational Morphology

When processing our parallel data, we did not attempt to do derivational morphology on the English side as the tagger did not perform any further morphological decomposition other than stemming. In order to gauge if such additional information could provide any enhancement, we used the CELEX database (http://www.ru.nl/celex/) to split derivations such as *friend+ship* or *develop+ment* into root and a marker indicating the derivation (e.g., friend +NNOM to indicate a noun-to-noun derivation and develop +VNOM, to indicate a verb to noun derivation, etc.) However, we did NOT observe any improvements in the BLEU score compared to our previous best results.

E. Augmenting Training Data

In order to overcome the disadvantages of the small size of our parallel data, we experimented with ways of using portions of the phrase table that is generated by the training process, as additional training data. The Moses phrase extraction process performs English–Turkish and Turkish–English alignments using the GIZA++ tool and then combines these alignments with some additional postprocessing and extracts *phrases*, sequences of source, and target tokens.

Phrase table entries produced by Moses alignment phase, 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. So we extracted 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. At each step, augmented training data comprised of original training data and extracted phrase pairs. The BLEU score result after six iterations of this augmentation scheme (on top of the best result of 23.42) is 24.83, resulting in a 59.8% relative improvement over the 15.53 baseline, and 6.02% relative improvement over the best previous result after local reordering. The reason we believe that this improves our results is that when such phrases are included in the training set the alignments in actual longer sentences improve, since these new phrases provide additional support for mathing parts of longer sentences, forcing the other words to align better.

F. Word Repair

The detailed BLEU result of 24.83 with 1–4-gram components as 47.7/29.3/20.4/14.8, for our best performing model, indicates that only 47.7% of the words in the candidate translations are determined correctly. However, when *all words* in both the candidate and reference translations are reduced to roots and BLEU is computed again, we get the *root* BLEU results of 34.4, with 1–4-gram components 70.15/39.76/27.28/20.24. This shows that we are getting 70.15% of the roots in the translations correct but only 47.7% of the words forms are correct. Such words have correct roots, but when considered with the morphemes, do not match the reference word. They can be considered under three cases.

1) Morphologically malformed words, i.e., words with the correct root word but with morphemes that are either categorically incorrect (e.g., case morpheme on a verb), or morphotactically incorrect (e.g., morphemes in the wrong order).
2) Morphologically well-formed words which are out-of-vocabulary (OOV) relative to the training corpus and the language model corpus.[14] Interestingly, such words are synthesized by combining translations from different phrases in decoding, indicating that the phrase-based decoder can occasionally produce morphologically legitimate words which have not been seen before in training and LM corpora. See example 1 in Section III-G.
3) Morphologically well-formed words which are *not* out-of-vocabulary relative to the training corpus and the language model corpus, but do not match the reference.

Words for cases 1 and 2 can be identified easily. Words for case 1 would be rejected by our morphological analyzer, while words for case 2 would be accepted by the morphological analyzer, but would not be in the vocabulary of the training and language model corpora. However, we have no way knowing without looking at the reference, if a word falls under case 3.

The approach we have taken to deal with the words for cases 1 and 2 is as follows.

1) Using a finite-state model of lexical morpheme structure of possible Turkish words, with morphemes being the symbols, we use *morpheme-level* error-tolerant finite state recognition [40] to generate morphologically correct word

[13]The local transformations were restricted to sequences occurring more than ten times, with length up to four tokens and did not involve full NP bracketing. Implementation was done with a Perl script and was obviously not perfect.

[14]Note that since Turkish has a very large number of possible word forms, there really are no well-formed words which are OOV, though there may be well-formed words which are extremely low frequency. It is such words that we aim to identify here.

TABLE VI
BLEU RESULTS FOR VARIOUS REORDERING SCHEMES

TRANSFORMATIONS	Default Param.		Modified Param.	
	BLEU	BLEU After Rescoring	BLEU	BLEU After Rescoring
of	14.72	15.43	21.69	22.56
verb	15.29	15.71	21.82	22.58
verb+the	16.09	16.30	22.52	22.90
prep1+inf	13.90	14.71	21.01	21.83
prep1+inf+the	16.70	16.83	23.32	**23.42**
prep2+inf	14.69	15.65	21.65	22.58
prep2+inf+the	16.36	16.59	22.59	22.92
prep1+inf+verb	15.36	16.00	22.15	22.90
prep1+inf+verb+the	16.50	16.64	22.97	23.22
prep2+inf+verb	14.73	15.55	21.81	22.61
prep2+inf+verb+the	16.21	16.38	22.66	22.89

forms with the same root but with the morpheme structure up to 2 unit morpheme edit operations (add, delete, substitute, transpose morphemes) away.[15] We do this for every word in 1 and 2 in a candidate translation sentence. For instance, the word form (in lexical morpheme representation) $\mathrm{gel+da+ydh}$ is malformed and possible correct variants at (morpheme) distance 1 are $\{\mathrm{gel+yacak+ydh},\ \mathrm{gel+mhs+ydh},\ \mathrm{gel+dh+ydh},\ \mathrm{gel+sa+ydh},\ \mathrm{gel+ya+ydh}\}$. We convert the sentence to a lattice representation replacing each malformed word with morphologically correct alternatives.

2) The resulting lattice is then rescored with the word-based LM.[16]

The procedure differs slightly for cases 1 and 2. For case 2, we restrict the alternatives to the vocabulary of the training and language model corpora. Additionally, we remove morphemes from punctuations (that incorrectly had morphemes attached) and from OOV numeric tokens.

All in all, word repair provides an additional improvement of 1.3% relative improvement to 25.17 (compared to 24.83 after augmenting data) and this final BLEU score represents a relative improvement of 62% over the baseline. There may be some more improvements along these lines by applying repair to low confidence words that can be identified by a scheme suggested by Zens and Ney [41].

Finally, we provide in Table VII, a summary BLEU score results for all the steps for our best performing segmentation (all with modified parameters for the Moses decoder).

[15]The detailed analysis of the decoded output with reference translations point out that errors generally are caused by some specific morphemes such as "$+\mathrm{dhr}$," "$+\mathrm{sh}$," "$+\mathrm{nhn}$," "$+\mathrm{ya}$," "$+\mathrm{da}$", "$+\mathrm{yh}$." We restricted the possible morpheme changes (deletion, insertion, replacement) with these six morphemes and scored our lattices with both word and morpheme language models.

[16]In all sentences processed, the biggest lattice we encountered for malformed word correction had two words that needed correction and those words had a total 17 candidate replacements generated. The biggest lattice we encountered for OOV word correction had three words that needed correction, with a total 34 candidates generated.

TABLE VII
SUMMARY BLEU RESULTS FOR ALL STEPS FOR THE SELECTIVE MORPHOLOGICAL SEGMENTATION REPRESENTATION

	STEP	BLEU	% IMP
0	Word-based Baseline	15.53	
1	(1) + Decoding with Modified Parameters	20.16	29.8%
2	Decoding a with Morpheme-based LM	21.9	41%
3	(2) + Adding Train-Content	22.20	42.9%
4	(3) + Rescoring with a word-based LM	22.81	44%
5	(4) + Reordering	23.69	52.5%
6	(5) + Data Augmentation	24.83	59.8%
7	(6) + Malformed Word Repair	24.90	60.3%
8	(7) + OOV Word Repair	25.17	62%

G. Sample Translation Outputs

Below, we present translations of some sentences from the test data after the postprocessing step along with the literal English paraphrases of the translated and the reference sentences. We also provide the reordered input to the decoder. Wherever relevant, we add some remarks about the translation produced.

1) Sentence 1:

Input: promote protection of children's rights in line with eu and international standards.

Decoder Input: promote protection of child +nns +pos right +nns line in with eu and international standard +nns.

Translation:çocuk hak+lar+sh+nhn koru+hn+ma+sh+na yönelik ab ve ulus+lararasi standart+lar**+yla** uyum+lh+dhr.

Literally: it is compatible with the eu and international standards regarding the protection of children's rights.

Reference: ab ve ulus+lararasi standart+lar dogrultu+sh +nda çocuk hak+lar+sh+nhn koru+hn+ma+sh+nhn teşvik et+hl+ma+sh.

Literally: in line with eu and international standards, promote protection of children's rights.

Although the translation does not completely match the reference, it displays one interesting aspect: the morpheme in boldface is actually case marker that functions as the subcategorization marker for the following root word with which it forms a translation phrase. This case marker gets attached to the end of the previous word, so that resulting case-marked noun is well-formed and satisfies the subcategorization constraint. This is an example of the decoder creating a morphologically legitimate word by getting the root from one phrase and the morpheme from another phrase.

2) Sentence 2:

Input: as a key feature of such a strategy, an accession partnership will be drawn up on the basis of previous European council conclusions.

Decoder Input: as a key feature of such a strategy, an accession partnership will +vb draw +vvn up on basis of previous European council conclusion +nns.

Translation: böyle bir strateji+nhn kilit unsur+sh, birönce+ki avrupa konsey+sh sonuç+lar+sh temel+sh+nda bir katilim ortaklik+sh belge+sh hazirla+hn+yacak+dhr.

Literally: The lock feature of such a strategy, an accession partnership document based on the previous European council results will be prepared.

Reference: bu strateji+nhn kilit unsur+sh ol+yarak, daha önce+ki ab zirve sonuç+lar+sh+na daya+hnhl+yarak bir katilim ortaklik+sh oluş+dhr+hl+yacak+dhr.
Literally: as a lock feature of this strategy an accession partnership based on earlier EU summit resolutions will be formed.

The segment *key feature* is properly translated as *kilit unsur+sh* "lock feature," which is the contextually correct idiomatic translation.

IV. Conclusion

This paper presented the results of an English-to-Turkish phrase-based statistical machine translation study. This language pair is interesting for statistical machine translation for a number of reasons: the target language, Turkish, is morphologically very rich and essentially has infinite vocabulary while English is relatively poorer in this respect. Translation into Turkish seems to involve processes that are somewhat more complex than standard statistical translation models; for example sometimes a *single word in Turkish needs to be synthesized from the translations of two or more phrases possible distant phrases in English*. Also the dearth of available parallel texts suggests that the available data has to be exploited in various ways to make most use of it.

Major results of our work can be summarized as follows.

1) We have considered various representational schemes to take morphological structure into account and have found that employing a language-pair specific morphological representation somewhere between using full word-forms and fully morphologically segmented representations provides the best results. Contrary to our original expectations, incorporating derivational morphology does not provide any improvements. Using content word roots as additional data provides some improvement with morphologically segmented representations (by presumably biasing the root word alignments), but not with baseline word-based representation.
2) Reranking the 1000-best outputs with a word-based model after decoding with a morpheme-based language model provides some additional improvement.
3) Local reordering of most frequent English prepositional phrases and infinitive verb structures to make their order more like Turkish morpheme order, provides some additional improvement. We have only considered high-frequency but short phrase patterns here. We expect to improve on this by using a more sophisticated phrase extractor.
4) Extracting highly reliable phrase translations from the phrase table and including them in the training data seem to provide additional bias to the alignments and improves the BLEU score.
5) Postprocessing the resulting sentences to identify and fix morphologically malformed or low-frequency word-forms provides a slight improvement.

In retrospect, we believe that while the approaches presented in this paper have provided nontrivial BLEU score improvements, it is not very clear whether root words and morphemes should to be translated by the same underlying general mechanism. Morpheme ordering is a much more local and constrained process. A radically different approach for handling morphology in English-to-Turkish statistical machine translation has recently been experimented with by Yeniterzi [42]. She has looked into exploiting dependency-based syntactic analysis on the source side, in order to identify English syntactic structures that need to be realized by noun or verb morphology on the Turkish side. With this approach, she needs to only statistically translate root words and complex structural tags, and avoids morpheme translation and composing words by concatenating morphemes. The results are very promising.

Appendix
Tags Used in Morphemic Representation of English Sentences

The tag set of TreeTagger tagset is an expanded version of Penn Treebank tagset[43]. Here we provide the subset of the tags that we used in our examples for the sake of being self-contained. For the verbs *be* and *have*, the second letter is specified as B and H, respectively:

Noun, Plural: NNS;
Verb, Base form: VV, VB, VH;
Verb, Past Tense: VVD, VBD, VHD;
Verb, Gerund or present participle: VVG, VBG, VHG;
Verb, Past Participle: VVN, VBN, VHN;
Verb, third-person singular present: VVZ, VBZ, VHZ;
Verb, Non-third-person singular present: VVP, VBP, VHP.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments and suggestions.

References

[1] J. Hutchins, "Machine translation: A brief history," in *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, E. Koerner and R. E. Asher, Eds. Oxford, U.K.: Pergamon, 1995, pp. 431–445.
[2] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, J. D. Lafferty, and R. L. Mercer, "Analysis, statistical transfer, and synthesis in machine translation," in *Proc. TMI: 4th Int. Conf. Theoretical and Methodological Issues in MT*, 1992, pp. 83–100.
[3] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer, "The mathematics of statistical machine translation: Parameter estimation," *Comput. Linguist.*, vol. 19, no. 2, pp. 263–311.
[4] R. Zens, F. J. Och, and H. Ney, M. Jarke, J. Koehler, and G. Lakemeyer, Eds., "Phrase-based statistical machine translation," in *Proc. 25th German Conf. Artif. Intell. (KI2002), Vol. 2479 of Lecture Notes in Artif. Intell. (LNAI)*, Aachen, Germany, 2002, pp. 18–22.
[5] D. Marcu and W. Wong, "A phrase-based, joint probability model for statistical machine translation," in *Proc. Conf. Empirical Methods in Natural Lang. Process. (EMNLP-02)*, Philadelphia, PA, 2002.
[6] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," *Proc. HLT/NAACL*, 2003.
[7] F. J. Och and H. Ney, "The alignment template approach to statistical machine translation," *Comput. Linguist.*, vol. 30, no. 4, pp. 417–449, 2004.
[8] D. Chiang, "A hierarchical phrase-based model for statistical machine translation," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguist. (ACL'05)*, Ann Arbor, MI, Jun. 2005, pp. 263–270 [Online]. Available: http://www.aclweb.org/anthology/P/P05/P05-1033
[9] P. Koehn and H. Hoang, "Factored translation models," in *Proc. EMNLP*, 2007.
[10] K. Yamada and K. Knight, "A syntax-based statistical translation model," in *Proc. ACL*, 2001, pp. 523–530.

[11] M. Hopkins and J. Kuhn, "Machine translation as tree labeling," in *Proc. SSST, NAACL-HLT 2007/AMTA Workshop Syntax and Structure in Statist. Transl.*, Rochester, NY, Apr. 2007, pp. 41–48 [Online]. Available: http://www.aclweb.org/anthology/W/W07/W07-0406
[12] C. Cherry and D. Lin, "Inversion transduction grammar for joint phrasal translation modeling," in *Proc. SSST, NAACL-HLT 2007/AMTA Workshop on Syntax and Structure in Statistical Transl.*, Rochester, NY, Apr. 2007, pp. 17–24 [Online]. Available: http://www.aclweb.org/anthology/W/W07/W07-0403
[13] D. Wu, "Stochastic inversion transduction grammars and bilingual parsing of parallel corpora," *Comput. Linguist.*, vol. 23, no. 3, pp. 377–403, 1997.
[14] İ. D. El-Kahlout and K. Oflazer, "Initial explorations in English to Turkish statistical machine translation," in *Proc. Workshop on Statist. Mach. Translation*, New York, Jun. 2006, pp. 7–14.
[15] J. Hankamer, "Morphological parsing and the lexicon," in *Lexical Representation and Process*, W. Marslen-Wilson, Ed. Cambridge, MA: MIT Press, 1989.
[16] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proc. MT Summit X*, Phuket, Thailand, 2005.
[17] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 311–318 [Online]. Available: http://www.aclweb.org/anthology/P02-1040.pdf
[18] S. Niessen and H. Ney, "Statistical machine translation with scarce resources using morpho-syntatic information," *Comput. Linguist.*, vol. 30, no. 2, pp. 181–204, 2004.
[19] M. Yang and K. Kirchhoff, "Phrase-based backoff models for machine translation of highly inflected languages," *Proc. EACL*, pp. 41–48, 2006.
[20] S. Corston-Oliver and M. Gamon, "Normalizing German and English inflectional morphology to improve statistical word alignment," *Proc. AMTA*, pp. 48–57, 2004.
[21] Y.-S. Lee, "Morphological analysis for statistical machine translation," *Proc. HLT-NAACL 2004—Companion Vol.*, pp. 57–60, 2004.
[22] A. Zollmann, A. Venugopal, and S. Vogel, "Bridging the inflection morphology gap for Arabic statistical machine translation," in *Proc. Human Lang. Technol. Conf. NAACL, Companion Vol.: Short Papers*, New York, Jun. 2006, pp. 201–204 [Online]. Available: http://www.aclweb.org/anthology/N/N06/N06-2051
[23] M. Popovic and H. Ney, "Towards the use of word stems and suffixes for statistical machine translation," in *Proc. 4th Int. Conf. Lang. Resources Eval. (LREC)*, May 2004, pp. 1585–1588.
[24] S. Goldwater and D. McClosky, "Improving statistical MT through morphological analysis," in *Proc. Human Lang. Technol. Conf. and Conf. Empirical Methods in Natural Lang. Process.*, Vancouver, BC, Canada, Oct. 2005, pp. 676–683 [Online]. Available: http://www.aclweb.org/anthology/H/H05/H05-1085
[25] E. Minkov, K. Toutanova, and H. Suzuki, "Generating complex morphology for machine translation," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguist.*, Prague, Czech Republic, Jun. 2007, pp. 128–135 [Online]. Available: http://www.aclweb.org/anthology/P/P07/P07-1017
[26] D. Talbot and M. Osborne, "Modelling lexical redundancy for machine translation," in *Proc. 21st Int. Conf. Comput. Linguist. and 44th Annu. Meeting Assoc. Comput. Linguist.*, Sydney, Australia, Jul. 2006, pp. 969–976 [Online]. Available: http://www.aclweb.org/anthology/P/P06/P06-1122
[27] K. Oflazer and İ. D. El-Kahlout, "Exploring different representational units in English-to-Turkish statistical machine translation," in *Proc. 2nd Workshop Statist. Mach. Transl.*, Prague, Czech Republic, Jun. 2007, pp. 25–32 [Online]. Available: http://www.aclweb.org/anthology/W/W07/W07-0704
[28] K. Oflazer, "Two-level description of Turkish morphology," *Literary Linguist. Comput.*, vol. 9, no. 2, pp. 137–148, 1994.
[29] D. Yuret and F. Türe, "Learning morphological disambiguation rules for Turkish," in *Proc. Human Lang. Technol. Conf. NAACL, Main Conf.*, New York, Jun. 2006, pp. 328–334 [Online]. Available: http://www.aclweb.org/anthology/N/N06/N06-1042
[30] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *Proc. Int. Conf. New Methods in Lang. Process.*, 1994.
[31] F. J. Och and H. Ney, "A systematic comparison of various statistical alignment models," *Comput. Linguist.*, vol. 29, no. 1, pp. 19–51, 2003.
[32] 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, "Moses: Open source toolkit for statistical machine translation," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguist. (ACL'07)—Companion Vol.*, Jun. 2007.
[33] A. Stolcke, "SRILM—An extensible language modeling toolkit," in *Proc. Int. Conf. Spoken Lang. Process.*, 2002 [Online]. Available: citeseer.ist.psu.edu/stolcke02srilm.html
[34] F. J. Och, E. Hinrichs and D. Roth, Eds., "Minimum error rate training in statistical machine translation," in *Proc. 41st Annu. Meeting Assoc. Comput. Linguist.*, Sapporo, Japan, 2003, pp. 160–167 [Online]. Available: http://www.aclweb.org/anthology/P03-1021
[35] F. Xia and M. McCord, "Improving a statistical MT system with automatically learned rewrite patterns," in *Proc. 20th Int. Conf. Comput. Linguist. (COLING)*, Geneva, Switzerland, Aug. 2004, pp. 508–514.
[36] M. Collins, P. Koehn, and I. Kucerova, "Clause restructuring for statistical machine translation," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguist. (ACL'05).*, Ann Arbor, MI, Jun. 2005, pp. 531–540.
[37] M. Popovic and H. Ney, "POS-based word reorderings for statistical machine translation," in *5th Int. Conf. Lang. Res. Eval. LREC*, Genoa, Italy, May 2006, pp. 1278–1283.
[38] C. Wang, M. Collins, and P. Koehn, "Chinese syntactic reordering for statistical machine translation," in *Proc. EMNLP*, Prague, Czech Republic, Jun. 2007, pp. 737–745.
[39] S. Zwarts and M. Dras, "Syntax-based word reordering in phase-based statistical machine translation: Why does it work?," in *Proc. MT SUMMIT XI*, Copenhagen, Denmark, Sep. 2007, pp. 559–566.
[40] K. Oflazer, "Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction," *Comput. Linguist.*, vol. 22, no. 1, pp. 73–90, Mar. 1996.
[41] R. Zens and H. Ney, "N-gram posterior probabilities for statistical machine translation," in *Proc. Workshop Statist. Mach. Transl.*, New York, Jun. 2006, pp. 72–77 [Online]. Available: http://www.aclweb.org/anthology/W/W06/W06-3110
[42] R. Yeniterzi, "Syntax-to-morphology alignment and constituent reordering in factored phrase-based statistical machine translation from English to Turkish," M.S. thesis, Sabancı Univ., Istanbul, Turkey, 2009.
[43] M. Marcus, B. Santorini, and M. A. Marcinkiewitz, "Building a large annotated corpus of English: The Penn Treebank," *Comput. Linguist.*, 1993.

İlknur Durgar El-Kahlout received the B.S. degree in computer engineering from Başkent University, Ankara, Turkey, and the M.S. and Ph.D. degrees in computer science from Sabancı University, Istanbul, Turkey.

She is currently a Postdoctoral Fellow at LIMSI-CNRS, Orsay, France. Her current interests are statistical machine translation and morphologically rich languages.

Kemal Oflazer received the B.S. degree in electrical engineering and the M.S. degree in computer science from Middle East Technical University, Ankara, Turkey, and the Ph.D. degree in computer science from Carnegie Mellon University, Pittsburgh, PA.

He is currently a faculty member at Carnegie Mellon University—Qatar, on long-term leave from Sabancı University, Istanbul, Turkey. He has held visiting positions at the Computing Research Laboratory, New Mexico State University, and at the Language Technologies Institute, Carnegie Mellon University. He has served on the editorial boards of *Computational Linguistics* and *Journal of AI Research*, and currently is the Book Reviews Editor of *Natural Language Engineering* and serves on the editorial boards of *Machine Translation*, *Linguistic Issues in Language Technology*, and *Research on Language and Computation*. His current research interests are in statistical machine translation into morphologically complex languages and dependency parsing.