

Syntactic Phrase Reordering for English-to-Arabic Statistical Machine Translation

Ibrahim Badr

Rabih Zbib

James Glass

Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

{iab02, rabih, glass}@csail.mit.edu

Abstract

Syntactic Reordering of the source language to better match the phrase structure of the target language has been shown to improve the performance of phrase-based Statistical Machine Translation. This paper applies syntactic reordering to English-to-Arabic translation. It introduces reordering rules, and motivates them linguistically. It also studies the effect of combining reordering with Arabic morphological segmentation, a pre-processing technique that has been shown to improve Arabic-English and English-Arabic translation. We report on results in the news text domain, the UN text domain and in the spoken travel domain.

1 Introduction

Phrase-based Statistical Machine Translation has proven to be a robust and effective approach to machine translation, providing good performance without the need for explicit linguistic information. Phrase-based SMT systems, however, have limited capabilities in dealing with long distance phenomena, since they rely on local alignments. Automatically learned reordering models, which can be conditioned on lexical items from both the source and the target, provide some limited reordering capability when added to SMT systems.

One approach that explicitly deals with long distance reordering is to reorder the source side to better match the target side, using predefined rules. The reordered source is then used as input to the phrase-based SMT system. This approach indirectly incorporates structure information since the reordering rules are applied on the parse trees

of the source sentence. Obviously, the same reordering has to be applied to both training data and test data. Despite the added complexity of parsing the data, this technique has shown improvements, especially when good parses of the source side exist. It has been successfully applied to German-to-English and Chinese-to-English SMT (Collins et al., 2005; Wang et al., 2007).

In this paper, we propose the use of a similar approach for English-to-Arabic SMT. Unlike most other work on Arabic translation, our work is in the direction of the more morphologically complex language, which poses unique challenges. We propose a set of syntactic reordering rules on the English source to align it better to the Arabic target. The reordering rules exploit systematic differences between the syntax of Arabic and the syntax of English; they specifically address two syntactic constructs. The first is the Subject-Verb order in independent sentences, where the preferred order in written Arabic is Verb-Subject. The second is the noun phrase structure, where many differences exist between the two languages, among them the order of adjectives, compound nouns and genitive constructs, as well as the way definiteness is marked. The implementation of these rules is fairly straightforward since they are applied to the parse tree. It has been noted in previous work (Habash, 2007) that syntactic reordering does not improve translation if the parse quality is not good enough. Since in this paper our source language is English, the parses are more reliable, and result in more correct reorderings. We show that using the reordering rules results in gains in the translation scores and study the effect of the training data size on those gains.

This paper also investigates the effect of using morphological segmentation of the Arabic target

in combination with the reordering rules. Morphological segmentation has been shown to benefit Arabic-to-English (Habash and Sadat, 2006) and English-to-Arabic (Badr et al., 2008) translation, although the gains tend to decrease with increasing training data size.

Section 2 provides linguistic motivation for the paper. It describes the rich morphology of Arabic, and its implications on SMT. It also describes the syntax of the verb phrase and noun phrase in Arabic, and how they differ from their English counterparts. In Section 3, we describe some of the relevant previous work. In Section 4, we present the preprocessing techniques used in the experiments. Section 5 describes the translation system, the data used, and then presents and discusses the experimental results from three domains: news text, UN data and spoken dialogue from the travel domain. The final section provides a brief summary and conclusion.

2 Arabic Linguistic Issues

2.1 Arabic Morphology

Arabic has a complex morphology compared to English. The Arabic noun and adjective are inflected for gender and number; the verb is inflected in addition for tense, voice, mood and person. Various clitics can attach to words as well: Conjunctions, prepositions and possessive pronouns attach to nouns, and object pronouns attach to verbs. The example below shows the decomposition into stems and clitics of the Arabic verb phrase *wsyqAblhm*¹ and noun phrase *wbydh*, both of which are written as one word:

- (1) a. w+ s+ yqAbl +hm
and will meet-3SM them
and he will meet them
- b. w+ b+ yd +h
and with hand his
and with his hand

An Arabic corpus will, therefore, have more surface forms than an equivalent English corpus, and will also be sparser. In the LDC news corpora used in this paper (see Section 5.2), the average English sentence length is 33 words compared to the Arabic 25 words.

¹All examples in this paper are written in the Buckwalter Transliteration System (<http://www.qamus.org/transliteration.htm>)

Although the Arabic language family consists of many dialects, none of them has a standard orthography. This affects the consistency of the orthography of Modern Standard Arabic (MSA), the only written variety of Arabic. Certain characters are written inconsistently in different data sources: Final 'y' is sometimes written as 'Y' (*Alif maSwrp*), and initial *Alif hamza* (The Buckwalter characters '<' and '{') are written as bare *alif* (A). Arabic is usually written without the diacritics that denote short vowels. This creates an ambiguity at the word level, since a word can have more than one reading. These factors adversely affect the performance of Arabic-to-English SMT, especially in the English-to-Arabic direction.

Simple pattern matching is not enough to perform morphological analysis and decomposition, since a certain string of characters can, in principle, be either an affixed morpheme or part of the base word itself. Word-level linguistic information as well as context analysis are needed. For example the written form *wly* can mean either *ruler* or *and for me*, depending on the context. Only in the latter case should it be decomposed.

2.2 Arabic Syntax

In this section, we describe a number of syntactic facts about Arabic which are relevant to the reordering rules described in Section 4.2.

Clause Structure

In Arabic, the main sentence usually has the order Verb-Subject-Object (VSO). The order Subject-Verb-Object (SVO) also occurs, but is less frequent than VSO. The verb agrees with the subject in gender and number in the SVO order, but only in gender in the VSO order (Examples 2c and 2d).

- (2) a. Akl Alwld AltFAHp
ate-3SM the-boy the-apple
the boy ate the apple
- b. Alwld Akl AltFAHp
the-boy ate-3SM the-apple
the boy ate the apple
- c. Akl AlAwlAd AltFAHAt
ate-3SM the-boys the-apples
the boys ate the apples
- d. AlAwlAd AklwA AltFAHAt
the-boys ate-3PM the-apples
the boys ate the apples

In a dependent clause, the order must be SVO, as illustrated by the ungrammaticality of Example 3b below. As we discuss in more detail later, this distinction between dependent and independent clauses has to be taken into account when the syntactic reordering rules are applied.

- (3) a. qAl An Alwld Akl AltFAHp
said-3SM that the-boy ate the-apple
he said that the boy ate the apple
b. *qAl An Akl Alwld AltFAHp
said-3SM that ate the-boy the-apple
he said that the boy ate the apple

Another pertinent fact is that the negation particle has to always precede the verb:

- (4) lm yAkl Alwld AltFAHp
not eat-3SM the-boy the-apple
the boy did not eat the apple

Noun Phrase

The Arabic noun phrase can have constructs that are quite different from English. The adjective in Arabic follows the noun that it modifies, and it is marked with the definite article, if the head noun is definite:

- (5) AlbAb Alkbyr
the-door the-big
the big door

The Arabic equivalent of the English possessive, compound nouns and the *of*-relationship is the Arabic *idafa* construct, which compounds two or more nouns. Therefore, N_1 's N_2 and N_2 of N_1 are both translated as $N_2 N_1$ in Arabic. As Example 6b shows, this construct can also be chained recursively.

- (6) a. bAb Albyt
door the-house
the house's door
b. mftAH bAb Albyt
key door the-house
The key to the door of the house

Example 6 also shows that an *idafa* construct is made definite by adding the definite article *Al-* to the last noun in the noun phrase. Adjectives follow the *idafa* noun phrase, regardless of which noun in the chain they modify. Thus, Example 7 is ambiguous in that the adjective *kbyr* (*big*) can modify any of the preceding three nouns. The same is true for relative clauses that modify a noun.

- (7) mftAH bAb Albyt Alkbyr
key door the-house the-big

These and other differences between the Arabic and English syntax are likely to affect the quality of automatic alignments, since corresponding words will occupy positions in the sentence that are far apart, especially when the relevant words (e.g. the verb and its subject) are separated by subordinate clauses. In such cases, the lexicalized distortion models used in phrase-based SMT do not have the capability of performing reorderings correctly. This limitation adversely affects the translation quality.

3 Previous Work

Most of the work in Arabic machine translation is done in the Arabic-to-English direction. The other direction, however, is also important, since it opens the wealth of information in different domains that is available in English to the Arabic speaking world. Also, since Arabic is a morphologically richer language, translating into Arabic poses unique issues that are not present in the opposite direction. The only works on English-to-Arabic SMT that we are aware of are Badr et al. (2008), and Sarikaya and Deng (2007). Badr et al. show that using segmentation and recombination as pre- and post- processing steps leads to significant gains especially for smaller training data corpora. Sarikaya and Deng use Joint Morphological-Lexical Language Models to re-rank the output of an English-to-Arabic MT system. They use regular expression-based segmentation of the Arabic so as not to run into recombination issues on the output side.

Similarly, for Arabic-to-English, Lee (2004), and Habash and Sadat (2006) show that various segmentation schemes lead to improvements that decrease with increasing parallel corpus size. They use a trigram language model and the Arabic morphological analyzer MADA (Habash and Rambow, 2005) respectively, to segment the Arabic side of their corpora. Other work on Arabic-to-English SMT tries to address the word reordering problem. Habash (2007) automatically learns syntactic reordering rules that are then applied to the Arabic side of the parallel corpora. The words are aligned in a sentence pair, then the Arabic sentence is parsed to extract reordering rules based on how the constituents in the parse tree are reordered on the English side. No significant improvement is

shown with reordering when compared to a baseline that uses a non-lexicalized distance reordering model. This is attributed in the paper to the poor quality of parsing.

Syntax-based reordering as a preprocessing step has been applied to many language pairs other than English-Arabic. Most relevant to the approach in this paper are Collins et al. (2005) and Wang et al. (2007). Both parse the source side and then reorder the sentence based on pre-defined, linguistically motivated rules. Significant gain is reported for German-to-English and Chinese-to-English translation. Both suggest that reordering as a preprocessing step results in better alignment, and reduces the reliance on the distortion model. Popovic and Ney (2006) use similar methods to reorder German by looking at the POS tags for German-to-English and German-to-Spanish. They show significant improvements on test set sentences that do get reordered as well as those that don't, which is attributed to the improvement of the extracted phrases. (Xia and McCord, 2004) present a similar approach, with a notable difference: the re-ordering rules are automatically learned from aligning parse trees for both the source and target sentences. They report a 10% relative gain for English-to-French translation. Although target-side parsing is optional in this approach, it is needed to take full advantage of the approach. This is a bigger issue when no reliable parses are available for the target language, as is the case in this paper. More generally, the use of automatically-learned rules has the advantage of readily applicable to different language pairs. The use of deterministic, pre-defined rules, however, has the advantage of being linguistically motivated, since differences between the two languages are addressed explicitly. Moreover, the implementation of pre-defined transfer rules based on target-side parses is relatively easy and cheap to implement in different language pairs.

Generic approaches for translating from English to more morphologically complex languages have been proposed. Koehn and Hoang (2007) propose Factored Translation Models, which extend phrase-based statistical machine translation by allowing the integration of additional morphological features at the word level. They demonstrate improvements for English-to-German and English-to-Czech. Tighter integration of features is claimed to allow for better modeling of

the morphology and hence is better than using pre-processing and post-processing techniques. Avramidis and Koehn (2008) enrich the English side by adding a feature to the Factored Model that models noun case agreement and verb person conjugation, and show that it leads to a more grammatically correct output for English-to-Greek and English-to-Czech translation. Although Factored Models are well equipped for handling languages that differ in terms of morphology, they still use the same distortion reordering model as a phrase-based MT system.

4 Preprocessing Techniques

4.1 Arabic Segmentation and Recombination

It has been shown previously work (Badr et al., 2008; Habash and Sadat, 2006) that morphological segmentation of Arabic improves the translation performance for both Arabic-to-English and English-to-Arabic by addressing the problem of sparsity of the Arabic side. In this paper, we use segmented and non-segmented Arabic on the target side, and study the effect of the combination of segmentation with reordering.

As mentioned in Section 2.1, simple pattern matching is not enough to decompose Arabic words into stems and affixes. Lexical information and context are needed to perform the decomposition correctly. We use the Morphological Analyzer MADA (Habash and Rambow, 2005) to decompose the Arabic source. MADA uses SVM-based classifiers of features (such as POS, number, gender, etc.) to score the different analyses of a given word in context. We apply morphological decomposition before aligning the training data. We split the conjunction and preposition prefixes, as well as possessive and object pronoun suffixes. We then glue the split morphemes into one prefix and one suffix, such that any given word is split into at most three parts: *prefix+ stem +suffix*. Note that plural markers and subject pronouns are not split. For example, the word *wlAwlAdh* ('and for his children') is segmented into *wl+ AwlAd +P:3MS*.

Since training is done on segmented Arabic, the output of the decoder must be recombined into its original surface form. We follow the approach of Badr et. al (2008) in combining the Arabic output, which is a non-trivial task for several reasons. First, the ending of a stem sometimes changes when a suffix is attached to it. Second, word end-

ings are normalized to remove orthographic inconsistency between different sources (Section 2.1). Finally, some words can recombine into more than one grammatically correct form. To address these issues, a lookup table is derived from the training data that maps the segmented form of the word to its original form. The table is also useful in recombining words that are erroneously segmented. If a certain word does not occur in the table, we back off to a set of manually defined recombination rules. Word ambiguity is resolved by picking the more frequent surface form.

4.2 Arabic Reordering Rules

This section presents the syntax-based rules used for re-ordering the English source to better match the syntax of the Arabic target. These rules are motivated by the Arabic syntactic facts described in Section 2.2.

Much like Wang et al. (2007), we parse the English side of our corpora and reorder using predefined rules. Reordering the English can be done more reliably than other source languages, such as Arabic, Chinese and German, since the state-of-the-art English parsers are considerably better than parsers of other languages. The following rules for reordering at the sentence level and the noun phrase level are applied to the English parse tree:

1. **NP:** All nouns, adjectives and adverbs in the noun phrase are inverted. This rule is motivated by the order of the adjective with respect to its head noun, as well as the *idafa* construct (see Examples 6 and 7 in Section 2.2). As a result of applying this rule, the phrase *the blank computer screen* becomes *the screen computer blank*.
2. **PP:** All prepositional phrases of the form $N_1 \text{ of } N_2 \dots \text{ of } N_n$ are transformed to $N_1 N_2 \dots N_n$. All N_i are also made indefinite, and the definite article is added to N_n , the last noun in the chain. For example, the phrase *the general chief of staff of the armed forces* becomes *general chief staff the armed forces*. We also move all adjectives in the top noun phrase to the end of the construct. So *the real value of the Egyptian pound* becomes *value the Egyptian pound real*. This rule is motivated by the *idafa* construct and its properties (see Example 6).

3. **the:** The definite article *the* is replicated before adjectives (see Example 5 above). So *the blank computer screen* becomes *the blank the computer the screen*. This rule is applied after **NP** rule above. Note that we do not replicate *the* before proper names.
4. **VP:** This rule transforms SVO sentences to VSO. All verbs are reordered on the condition that they have their own subject noun phrase and are not in the participle form, since in these cases the Arabic subject occurs before the verb participle. We also check that the verb is not in a relative clause with a *that* complementizer (Example 3 above). The following example illustrates all these cases: *the health minister stated that 11 police officers were wounded in clashes with the demonstrators* → *stated the health minister that 11 police officers were wounded in clashes with the demonstrators*. If the verb is negated, the negative particle is moved with the verb (Example 4). Finally, if the object of the reordered verb is a pronoun, it is reordered with the verb. Example: *the authorities gave us all the necessary help* becomes *gave us the authorities all the necessary help*.

The transformation rules 1, 2 and 3 are applied in this order, since they interact although they do not conflict. So, *the real value of the Egyptian pound* → *value the Egyptian the pound the real*. The VP reordering rule is independent.

5 Experiments

5.1 System description

For the English source, we first tokenize using the Stanford Log-linear Part-of-Speech Tagger (Toutanova et al., 2003). We then proceed to split the data into smaller sentences and tag them using Ratnaparkhi’s Maximum Entropy Tagger (Ratnaparkhi, 1996). We parse the data using the Collins Parser (Collins, 1997), and then tag person, location and organization names using the Stanford Named Entity Recognizer (Finkel et al., 2005). On the Arabic side, we normalize the data by changing final ‘Y’ to ‘y’, and changing the various forms of *Alif hamza* to bare *Alif*, since these characters are written inconsistently in some Arabic sources. We then segment the data using MADA according to the scheme explained in Section 4.1.

The English source is aligned to the segmented Arabic target using the standard MOSES (MOSES, 2007) configuration of GIZA++ (Och and Ney, 2000), which is IBM Model 4, and decoding is done using the phrase-based SMT system MOSES. We use a maximum phrase length of 15 to account for the increase in length of the segmented Arabic. We also use a lexicalized bidirectional reordering model conditioned on both the source and target sides, with a distortion limit set to 6. We tune using Och’s algorithm (Och, 2003) to optimize weights for the distortion model, language model, phrase translation model and word penalty over the BLEU metric (Papineni et al., 2001). For the segmented Arabic experiments, we experiment with tuning using non-segmented Arabic as a reference. This is done by recombining the output before each tuning iteration is scored and has been shown by Badr et. al (2008) to perform better than using segmented Arabic as reference.

5.2 Data Used

We report results on three domains: newswire text, UN data and spoken dialogue from the travel domain. It is important to note that the sentences in the travel domain are much shorter than in the news domain, which simplifies the alignment as well as reordering during decoding. Also, since the travel domain contains spoken Arabic, it is more biased towards the Subject-Verb-Object sentence order than the Verb-Subject-Object order more common in the news domain. Also note that since most of our data was originally intended for Arabic-to-English translation, our test and tuning sets have only one reference, and therefore, the BLEU scores we report are lower than typical scores reported in the literature on Arabic-to-English.

The news training data consists of several LDC corpora². We construct a test set by randomly picking 2000 sentences. We pick another 2000 sentences randomly for tuning. Our final training set consists of 3 million English words. We also test on the NIST MT 05 “test set while tuning on both the NIST MT 03 and 04 test sets. We use the first English reference of the NIST test sets as the source, and the Arabic source as our reference. For

²LDC2003E05 LDC2003E09 LDC2003T18
LDC2004E07 LDC2004E08 LDC2004E11 LDC2004E72
LDC2004T18 LDC2004T17 LDC2005E46 LDC2005T05
LDC2007T24

Scheme	RandT		MT 05	
	S	NoS	S	NoS
Baseline	21.6	21.3	23.88	23.44
VP	21.9	21.5	23.98	23.58
NP	21.9	21.8		
NP+PP	21.8	21.5	23.72	23.68
NP+PP+VP	22.2	21.8	23.74	23.16
NP+PP+VP+The	21.3	21.0		

Table 1: Translation Results for the News Domain in terms of the BLEU Metric.

the language model, we use 35 million words from the LDC Arabic Gigaword corpus, plus the Arabic side of the 3 million word training corpus. Experimentation with different language model orders shows that the optimal model orders are 4-grams for the baseline system and 6-grams for the segmented Arabic. The average sentence length is 33 for English, 25 for non-segmented Arabic and 36 for segmented Arabic.

To study the effect of syntactic reordering on larger training data sizes, we use the UN English-Arabic parallel text (LDC2003T05). We experiment with two training data sizes: 30 million and 3 million words. The test and tuning sets are comprised of 1500 and 500 sentences respectively, chosen at random.

For the spoken domain, we use the BTEC 2007 Arabic-English corpus. The training set consists of 200K words, the test set has 500 sentences and the tuning set has 500 sentences. The language model consists of the Arabic side of the training data. Because of the significantly smaller data size, we use a trigram LM for the baseline, and a 4-gram for segmented Arabic. In this case, the average sentence length is 9 for English, 8 for Arabic, and 10 for segmented Arabic.

5.3 Translation Results

The translation scores for the News domain are shown in Table 1. The notation used in the table is as follows:

- **S:** Segmented Arabic
- **NoS:** Non-Segmented Arabic
- **RandT:** Scores for test set where sentences were picked at random from NEWS data
- **MT 05:** Scores for the NIST MT 05 test set

The reordering notation is explained in Section 4.2. All results are in terms of the BLEU met-

	S		NoS	
	Short	Long	Short	Long
Baseline	22.57	25.22	22.40	24.33
VP	22.95	25.05	22.95	24.02
NP+PP	22.71	24.76	23.16	24.067
NP+PP+VP	22.84	24.62	22.53	24.56

Table 2: Translation Results depending on sentence length for NIST test set.

Scheme	Score	% Oracle reord
VP	25.76	59%
NP+PP	26.07	58%
NP+PP+VP	26.17	53%

Table 3: Oracle scores for combining baseline system with other reordered systems.

ric. It is important to note that the gain that we report in terms of BLEU are more significant than comparable gains on test sets that have multiple references, since our test sets have only one reference. Any amount of gain is a result of additional n-gram precision with one reference. We note that the gain achieved from the reordering of the non-segmented and segmented systems are comparable. Replicating *the* before adjectives hurts the scores, possibly because it increases the sentence length noticeably, and thus deteriorates the alignments’ quality. We note that the gains achieved by reordering on the NIST test set are smaller than the improvements on the random test set. This is due to the fact that the sentences in the NIST test set are longer, which adversely affects the parsing quality. The average English sentence length is 33 words in the NIST test set, while the random test set has an average sentence length of 29 words. Table 2 shows the reordering gains of the non-segmented Arabic by sentence length. Short sentences are sentences that have less than 40 words of English, while long sentences have more than 40 words. Out of the 1055 sentence in the NIST test set 719 are short and 336 are long. We also report oracle scores in Table 3 for combining the baseline system with the reordering systems, as well as the percentage of oracle sentences produced by the reordered system. The oracle score is computed by starting with the reordered system’s candidate translations and iterating over all the sentences one by one: we replace each sentence with its corresponding baseline system translation then

Scheme	30M	3M
Baseline	32.17	28.42
VP	32.46	28.60
NP+PP	31.73	28.80

Table 4: Translation Results on segmented UN data in terms of the BLEU Metric.

compute the total BLEU score of the entire set. If the score improves, then the sentence in question is replaced with the baseline system’s translation, otherwise it remains unchanged and we move on to the next one.

In Table 4, we report results on the UN corpus for different training data sizes. It is important to note that although gains from VP reordering stay constant when scaled to larger training sets, gains from NP+PP reordering diminish. This is due to the fact that NP reordering tend to be more localized than VP reorderings. Hence with more training data the lexicalized reordering model becomes more effective in reordering NPs.

In Table 5, we report results on the BTEC corpus for different segmentation and reordering scheme combinations. We should first point out that all sentences in the BTEC corpus are short, simple and easy to align. Hence, the gain introduced by reordering might not be enough to offset the errors introduced by the parsing. We also note that spoken Arabic usually prefers the Subject-Verb-Object sentence order, rather than the Verb-Subject-Object sentence order of written Arabic. This explains the fact that no gain is observed when the verb phrase is reordered. Noun phrase reordering produces a significant gain with non-segmented Arabic. Replicating the definite article *the* in the noun phrase does not create alignment problems as is the case with the newswire data, since the sentences are considerably shorter, and hence the 0.74 point gain observed on the segmented Arabic system. That gain does not translate to the non-segmented Arabic system since in that case the definite article *Al* remains attached to its head word.

6 Conclusion

This paper presented linguistically motivated rules that reorder English to look like Arabic. We showed that these rules produce significant gains. We also studied the effect of the interaction between Arabic morphological segmentation and

Scheme	S	NoS
Baseline	29.06	25.4
VP	26.92	23.49
NP	27.94	26.83
NP+PP	28.59	26.42
The	29.8	25.1

Table 5: Translation Results for the Spoken Language Domain in the BLEU Metric.

syntactic reordering on translation results, as well as how they scale to bigger training data sizes.

Acknowledgments

We would like to thank Michael Collins, Ali Mohammad and Stephanie Seneff for their valuable comments.

References

- Eleftherios Avramidis, and Philipp Koehn 2008. *Enriching Morphologically Poor Languages for Statistical Machine Translation*. In Proc. of ACL/HLT.
- Ibrahim Badr, Rabih Zbib, and James Glass 2008. *Segmentation for English-to-Arabic Statistical Machine Translation*. In Proc. of ACL/HLT.
- Michael Collins 1997. *Three Generative, Lexicalized Models for Statistical Parsing*. In Proc. of ACL.
- Michael Collins, Philipp Koehn, and Ivona Kucerova 2005. *Clause Restructuring for Statistical Machine Translation*. In Proc. of ACL.
- Jenny Rose Finkel, Trond Grenager, and Christopher Manning 2005. *Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling*. In Proc. of ACL.
- Nizar Habash, 2007. *Syntactic Preprocessing for Statistical Machine Translation*. In Proc. of the Machine Translation Summit (MT-Summit).
- Nizar Habash and Owen Rambow, 2005. *Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop*. In Proc. of ACL.
- Nizar Habash and Fatiha Sadat, 2006. *Arabic Preprocessing Schemes for Statistical Machine Translation*. In Proc. of HLT.
- Philipp Koehn and Hieu Hoang, 2007. *Factored Translation Models*. In Proc. of EMNLP/CNLL.
- Young-Suk Lee, 2004. *Morphological Analysis for Statistical Machine Translation*. In Proc. of EMNLP.
- MOSES, 2007. *A Factored Phrase-based Beam-search Decoder for Machine Translation*. URL: <http://www.statmt.org/moses/>.
- Franz Och 2003. *Minimum Error Rate Training in Statistical Machine Translation*. In Proc. of ACL.
- Franz Och and Hermann Ney 2000. *Improved Statistical Alignment Models*. In Proc. of ACL.
- Kishore Papineni, Salim Roukos, Todd Ward, and Weijing Zhu 2001. *BLUE: a Method for Automatic Evaluation of Machine Translation*. In Proc. of ACL.
- Maja Popovic and Hermann Ney 2006. *POS-based Word Reordering for Statistical Machine Translation*. In Proc. of NAACL LREC.
- Adwait Ratnaparkhi 1996. *A Maximum Entropy Model for Part-of-Speech Tagging*. In Proc. of EMNLP.
- Ruhi Sarikaya and Yonggang Deng 2007. *Joint Morphological-Lexical Language Modeling for Machine Translation*. In Proc. of NAACL HLT.
- Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. *Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network*. In Proc. of NAACL HLT.
- Chao Wang, Michael Collins, and Philipp Koehn 2007. *Chinese Syntactic Reordering for Statistical Machine Translation*. In Proc. of EMNLP.
- Fei Xia and Michael McCord 2004. *Improving a Statistical MT System with Automatically Learned Rewrite Patterns*. In COLING.