

Selective Combination of Pivot and Direct Statistical Machine Translation Models

Ahmed El Kholy, Nizar Habash

Center for Computational Learning Systems, Columbia University
{akholy, habash}@ccls.columbia.edu

Gregor Leusch, Evgeny Matusov

Science Applications International Corporation
{gregor.leusch, evgeny.matusov}@saic.com

Hassan Sawaf

eBay Inc.
hsawaf@ebay.com

Abstract

In this paper, we propose a selective combination approach of pivot and direct statistical machine translation (SMT) models to improve translation quality. We work with Persian-Arabic SMT as a case study. We show positive results (from 0.4 to 3.1 BLEU on different direct training corpus sizes) in addition to a large reduction of pivot translation model size.

1 Introduction

One of the main challenges in statistical machine translation (SMT) is the scarcity of parallel data for many language pairs especially when the source and target languages are morphologically rich. Morphological richness comes with many challenges and the severity of these challenges increases when the richness and morphological complexity are expressed differently in the source and target languages.

A common SMT solution to the lack of parallel data is to pivot the translation through a third language (called pivot or bridge language) for which there exist abundant parallel corpora with the source and target languages. The literature covers many pivoting techniques. One of the best performing techniques, phrase pivoting (Utiyama and Isahara, 2007), builds an induced new phrase table between the source and target. One of the problems of this technique is that the size of the newly created pivot phrase table is very large (Utiyama and Isahara, 2007).

Given a parallel corpus between the source and target language, combining a direct model based on this parallel corpus with a pivot model could lead to better coverage and overall translation quality. However, the combination approach needs

to be optimized in order to maximize the information gain.

In this paper, we propose a selective combination approach of pivot and direct SMT models. The main idea is to select the relevant portions of the pivot model that do not interfere with the more trusted direct model. We show positive results for Persian-Arabic SMT (from 0.4 to 3.1 BLEU on different direct training corpus sizes). As a positive side effect, we achieve a large reduction of pivot translation model size.

This paper is organized as follows. Section 2 briefly discusses some related work. Section 3 presents linguistic challenges and differences between Arabic and Persian. In Section 4, we discuss our pivoting strategies. Then Section 5 discusses our approach for selective combination. In Section 6, we present our experimental results.

2 Related Work

2.1 Pivoting

Many researchers have investigated the use of pivoting (or bridging) approaches to solve the data scarcity issue (Utiyama and Isahara, 2007; Wu and Wang, 2009; Khalilov et al., 2008; Bertoldi et al., 2008; Habash and Hu, 2009). The core idea is to introduce a pivot language, for which there exist large source-pivot and pivot-target bilingual corpora. Pivoting has been explored for closely related languages (Hajič et al., 2000) as well as unrelated languages (Koehn et al., 2009; Habash and Hu, 2009). Many different pivoting strategies have been presented in the literature. The following two are perhaps the most commonly used.¹

¹Another notable strategy is to create a synthetic source-target corpus by translating the pivot side of source-pivot corpus to the target language using an existing pivot-target model (Bertoldi et al., 2008). A new source-target model is built from the new corpus.

The first strategy is sentence pivoting in which we first translate the source sentence to the pivot language, and then translate the pivot language sentence to the target language (Khalilov et al., 2008).

The second strategy is phrase pivoting (Utiyama and Isahara, 2007; Cohn and Lapata, 2007; Wu and Wang, 2009). In phrase pivoting, a new source-target phrase table (translation model) is induced from source-pivot and pivot-target phrase tables. We compute the lexical weights and translation probabilities from the two phrase tables.

In this paper, we utilize the phrase pivoting strategy as our baseline, which is shown to be better in performance compared to sentence pivoting (El Kholy et al., 2013).

2.2 Domain Adaptation

We propose a selective combination approach of pivot and direct SMT models to improve the translation quality. Our approach is similar to domain adaptation techniques where training data from many diverse sources are combined to build a single translation model which is used to translate sentences in a new domain.

Domain adaptation has been explored in the field through different methods. Some methods involve information retrieval (IR) techniques to retrieve sentence pairs related to the target domain from a training corpus (Eck et al., 2004; Hildebrand et al., 2005). Other domain adaptation methods are based on distinguishing between general and domain specific examples (Daumé III and Marcu, 2006). In a similar approach, Koehn and Schroeder (2007) use multiple alternative decoding paths to combine different translation models and the weights are set with minimum error rate training (Och and Ney, 2003).

In contrast to domain adaptation, we generate a new source-target translation model by phrase pivoting technique from two models. We then use domain adaptation approach to select relevant portions of the pivot phrase table and combine them with a direct translation model to improve the overall translation quality.

2.3 Morphologically Rich Languages

Since both Persian and Arabic are morphologically rich, we should mention that there has been a lot of work on translation to and from morphologically rich languages (Yeniterzi and Oflazer, 2010; Elming and Habash, 2009; El Kholy and Habash, 2010a; Habash and Sadat, 2006; Kathol and Zheng, 2008; Shilon et al., 2010). Most of

these efforts are focused on syntactic and morphological processing.

There have been a growing number of publications that consider translation into Arabic. Sarikaya and Deng (2007) use joint morphological-lexical language models to re-rank the output of English-dialectal Arabic MT. Other efforts report results on the value of morphological tokenization of Arabic during training and describe different techniques for detokenizing Arabic output (Badr et al., 2008; El Kholy and Habash, 2010b).

On the other hand, work on Persian SMT is limited to few studies. For example, Kathol and Zheng (2008) use unsupervised morpheme segmentation for Persian. They show that hierarchical phrase-based models can improve Persian-English translation. There are also other attempts to improve Persian-English SMT by working on syntactic reordering (Gupta et al., 2012) and rule-based post editing (Mohaghegh et al., 2012). There is also some work done on showing the effect of different orthographic and morphological processing for Persian on Persian-English translation (Rasooli et al., 2013a).

To our knowledge, there hasn't been a lot of work on Persian and Arabic as a language pair. One example is an effort based on improving the reordering models for Persian-Arabic SMT (Matusov and Köprü, 2010). Another recent effort improved the quality of Persian-Arabic by pivoting through English and adding additional features to reflect the quality of projected alignments between the source and target phrases in the pivot phrase table (El Kholy et al., 2013).

3 Arabic and Persian Linguistic Issues

In this section we present our motivation and choice for preprocessing Arabic, Persian and English data. Both Arabic and Persian are morphologically complex languages but they belong to two different language families. They both express richness and linguistic complexities in different ways (El Kholy et al., 2013).

One aspect of Arabic's complexity is its various attachable clitics and numerous morphological features (Habash, 2010) which include conjunction proclitics, e.g., $+و$ *w+* 'and', particle proclitics, e.g., $+ل$ *l+* 'to/for', the definite article $+ال$ *Al+* 'the', and the class of pronominal enclitics, e.g., $+هم$ *+hm* 'their/them'. Beyond these clitics, Arabic words inflect for person, gender, number,

aspect, mood, voice, state and case.² This morphological richness leads to thousands of inflected forms per lemma and a high degree of ambiguity: about 12 analyses per word, typically corresponding to two lemmas on average (Habash, 2010). We follow El Kholy and Habash (2010a) and use the PATB tokenization scheme (Maamouri et al., 2004) in our experiments which separates all clitics except for the determiner clitic *Al+*. We use MADA v3.1 (Habash and Rambow, 2005; Habash et al., 2009) to tokenize the Arabic text. We only evaluate on detokenized and orthographically correct (enriched) output following the work of El Kholy and Habash (2010b).

Persian on the other hand has a relatively simple nominal system. There is no case system and words do not inflect with gender except for a few animate Arabic loanwords. Unlike Arabic, Persian shows only two values for number, just singular and plural (no dual), which are usually marked by either the suffix *ها+ +hA* and sometimes *ان+ +An*, or one of the Arabic plural markers. Persian also possess a closed set of few broken plurals loaned from Arabic. Moreover, unlike Arabic which expresses definiteness, Persian expresses indefiniteness with an enclitic article *ی+ +y* ‘*a/an*’ which doesn’t have separate forms for singular and plural. When a noun is modified by one or more adjective, the indefinite article is attached to the last adjective. Persian adjectives are similar to English in expressing comparative and superlative constructions just by adding suffixes *تر+ +tar* ‘+er’ and *ترین+ +taryn* ‘+est’ respectively. Verbal morphology is very complex in Persian. Each verb has a past and present root and many verbs have attached prefix that is regarded part of the root. A verb in Persian inflects for 14 different tense, mood, aspect, person, number and voice combination values (Rasooli et al., 2013b).

We follow El Kholy et al. (2013) and tokenize Persian text using Perstem (Jadidinejad et al., 2010) which is a deterministic rule based approach for segmentation of Persian.

English, our pivot language, is quite different from both Arabic and Persian. English is poor in morphology and barely inflects for number and tense, and for person in a limited context. English preprocessing simply includes down-casing, separating punctuation and splitting off “’s”.

²We use the Habash-Soudi-Buckwalter Arabic transliteration (Habash et al., 2007) with extensions for Persian as suggested by Habash (2010).

4 Pivoting Strategies

In this section, we review the two pivoting strategies that are our baselines.

4.1 Sentence Pivoting

In sentence pivoting, English is used as an interface between two separate phrase-based MT systems; Persian-English direct system and English-Arabic direct system. Given a Persian sentence, we first translate the Persian sentence from Persian to English, and then from English to Arabic.

4.2 Phrase Pivoting

In phrase pivoting (sometimes called triangulation or phrase table multiplication), we train a Persian-to-Arabic and an English-Arabic translation models, such as those used in the sentence pivoting technique. Based on these two models, we induce a new Persian-Arabic translation model.

Since we build our models are based Moses phrase-based SMT (Koehn et al., 2007), we provide the basic set of phrase translation probability distributions.³ We follow Utiyama and Isahara (2007) in computing the probability distributions. The following are the set of equations used to compute the lexical probabilities (ϕ) and the phrase translation probabilities (p_w)

$$\begin{aligned}\phi(f|a) &= \sum_e \phi(f|e)\phi(e|a) \\ \phi(a|f) &= \sum_e \phi(a|e)\phi(e|f) \\ p_w(f|a) &= \sum_e p_w(f|e)p_w(e|a) \\ p_w(a|f) &= \sum_e p_w(a|e)p_w(e|f)\end{aligned}$$

where f is the Persian source phrase. e is the English pivot phrase that is common in both Persian-English translation model and English-Arabic translation model. a is the Arabic target phrase.

We also build a Persian-Arabic reordering table using the same technique but we compute the reordering probabilities in a similar manner to Henriquez et al. (2010).

As discussed earlier, the induced Persian-Arabic phrase and reordering tables are very large. Table 1 shows the amount of parallel corpora used to train the Persian-English and the English-Arabic and the equivalent phrase table sizes compared to the induced Persian-Arabic phrase table.⁴

³Four different phrase translation scores are computed in Moses’ phrase tables: two lexical weighting scores and two phrase translation probabilities.

⁴The size of the induced phrase table size is computed but not created.

Translation Model	Training Corpora Size	Phrase Table	
		# Phrase Pairs	Size
Persian-English	≈4M words	96,04,103	1.1GB
English-Arabic	≈60M words	111,702,225	14GB
Pivot_Persian-Arabic	N/A	39,199,269,195	≈2.5TB

Table 1: Translation Models Phrase Table comparison in terms of number of line and sizes.

We follow the work of El Kholy et al. (2013) and filter the phrase pairs used in pivoting based on log-linear scores. We present some baseline results to show the effect of filtering on the translation quality in Section 6.2.

5 Approach

In this section, we discuss our selective combination approach. We explore how to effectively combine both a pivot and a direct model built from a given parallel corpora to achieve better coverage and overall translation quality. We maximize the information gain by selecting the relevant portions of the pivot model that do not interfere with the more trusted direct model.

To achieve this goal, we investigate the idea of classifying the pivot phrase pairs into five different classes based on the existence of source and/or target phrases in the direct model. The first class contains the phrase pairs where the source and target phrases are in the direct system together. The second class is the same as the first class except that the source and target phrases exist but not together as a phrase pair in the direct system. The third, fourth and fifth classes are for the existence of source phrase only, target phrase only and neither in the direct system. Table 2 shows the different classifications of the portions extracted from the pivot phrase table with their labels which are used later in our results tables.

We use one of Moses phrase table combination techniques (Koehn and Schroeder, 2007) to combine the direct model with the different pivot portions (explained in more details in section 6.1).

6 Experiments

In this section, we present our results for the selective combination approach between direct and pivoting models.

6.1 Experimental Setup

For the direct Persian-Arabic SMT model, we use an inhouse parallel corpus of about 165k sentences and 4 million words.

Pivot phrase-pairs classification	Src exists in direct	Tgt exists in direct	Src & Tgt exist in direct
SRC : TGT	✓	✓	✓
SRC , TGT	✓	✓	×
SRC ONLY	✓	×	×
TGT ONLY	×	✓	×
NEITHER	×	×	×

Table 2: Phrase pairs classification of the portions extracted from the pivot phrase table.

In our pivoting experiments, we build two SMT models. One model to translate from Persian to English and another model to translate from English to Arabic. The English-Arabic parallel corpus is about 2.8M sentences (≈60M words) available from LDC⁵ and GALE⁶ constrained data. We use an in-house Persian-English parallel corpus of about 170K sentences and 4M words.

Word alignment is done using GIZA++ (Och and Ney, 2003). For Arabic language modeling, we use 200M words from the Arabic Gigaword Corpus (Graff, 2007) together with the Arabic side of our training data. We use 5-grams for all language models (LMs) implemented using the SRILM toolkit (Stolcke, 2002). For English language modeling, we use the English Gigaword Corpus with 5-gram LM using the KenLM toolkit (Heafield, 2011).

All experiments are conducted using Moses phrase-based SMT system (Koehn et al., 2007). We use MERT (Och, 2003) for decoding weights optimization. For Persian-English translation model, weights are optimized using a set 1000 sentences randomly sampled from the parallel corpus while the English-Arabic translation model weights are optimized using a set of 500 sentences from the 2004 NIST MT evaluation test set

⁵LDC Catalog IDs: LDC2005E83, LDC2006E24, LDC2006E34, LDC2006E85, LDC2006E92, LDC2006G05, LDC2007E06, LDC2007E101, LDC2007E103, LDC2007E46, LDC2007E86, LDC2008E40, LDC2008E56, LDC2008G05, LDC2009E16, LDC2009G01.

⁶Global Autonomous Language Exploitation, or GALE, is a DARPA-funded research project.

(MT04).

We use a maximum phrase length of size 8 across all models. We report results on an in-house Persian-Arabic evaluation set of 536 sentences with three references. We evaluate using BLEU-4 (Papineni et al., 2002).

For the combination experiments, Moses allows the use of multiple translation tables (Koehn and Schroeder, 2007). Different combination techniques are available. We use the ‘‘Either’’ combination technique where the translation options are collected from one table, and additional options are collected from the other tables. If the same translation option (identical source and target phrases) is found in multiple tables, separate translation options are created for each occurrence, but with different scores.

6.2 Baseline Evaluation

We compare the performance of sentence pivoting against phrase pivoting with different filtering thresholds. The results are presented in Table 3. In general, the phrase pivoting outperforms the sentence pivoting even when we use a small filtering threshold of size 100. Moreover, the higher the threshold the better the performance but with a diminishing gain.

Pivot Scheme	BLEU
Sentence Pivoting	19.2
Phrase_Pivot_F100	19.4
Phrase_Pivot_F500	20.1
Phrase_Pivot_F1K	20.5

Table 3: Sentence pivoting versus phrase pivoting with different filtering thresholds (100/500/1000).

We use the best performing setup across the rest of the experiments which is filtering with a threshold of 1K.

6.3 System Combinations

In this section, we investigate the selective combination approach. We start by the basic combination approach and then explore the gain/loss achieved from dividing the pivot phrase table to five different classes as discussed in Section 5.

6.3.1 Baseline Combination

Table 4 shows the results of the basic combination in comparison to the best pivot translation model and the best direct model. The results shows that combining both models leads to a gain in performance. The question is how to improve the quality

by doing a smart selection of only relevant portion of the pivot phrase table which is discussed next.

Model	BLEU%
Phrase_Pivot_F1K	20.5
Direct	23.4
Direct+Phrase_Pivot_F1K	23.7

Table 4: Baseline combination experiments between best pivot baseline and best direct model.

6.3.2 Selective Combination

In this section, we explore the idea of dividing the pivot phrase pairs into five different classes based on the existence of source and/or target phrases in the direct system as discussed in Section 5. We discuss our results and show the trade off between the quality of translation and the size of the different classes extracted from the pivot phrase table.

Table 5 shows the results of the selective combination experiments on a learning curve of 100% (4M words), 25% (1M words) and 6.25% (250K words) of the parallel Persian-Arabic corpus.

Model	Parallel data set size		
	4M	1M	250K
Direct	23.4	21.0	16.8
Phrase_Pivot_F1K	20.5		
Base Combination	23.7 *	22.1 *	21.7 *
SRC : TGT	22.9	21.2	17.3 *
SRC , TGT	23.0	21.3	18.5 *
SRC ONLY	23.5	20.1	17.5 *
TGT ONLY	23.8*	21.4 *	18.3 *
NEITHER	23.4	21.6 *	19.9 *

Table 5: Selective Combination experiments results on a learning curve. The first row shows the results of the direct system. The second row shows the result of the best pivot system. The third row shows the results of the baseline combination experiments with the whole pivot phrase table. Then the next set of rows show the results of the selective combination experiments based on the different classifications. All scores are in BLEU. (*) marks a statistically significant result against the direct baseline.

The results show that pivoting is a robust technique when there is no or small amount of parallel corpora. In our case study on Persian-Arabic SMT, the direct translation systems built from parallel corpora starts to be better than the pivot translation system when trained on 1M words or more.

The base combination between the direct translation models and the pivot translation model leads to a boost in the translation quality across the learning curve. As expected, the smaller the parallel corpus used in training the more gain we get from the combination.

The results also show that some of pivot the classes provides more information gain than others. In fact some of the classes hurt the overall quality; for example, (SRC : TGT) and (SRC , TGT) both hurt the quality of translation when combined with direct model trained on 100% of the parallel data (4M words).

An interesting observation from the results is that by building a translation system with only 6.25% of the parallel data (\approx 250K words) combined with the pivot translation model, we can achieve a better performance (21.7 BLEU) than a model trained on four times the amount of data (Size: 1M words; Score: 21.0 BLEU).

It is also shown across the learning curve that the best gains are achieved when the source phrase in the pivot phrase table doesn't exist in the direct model. This is expected due to the fact that by adding unknown source phrases, we decrease the overall OOVs.

Model	Parallel data set size		
	4M	1M	250K
SRC : TGT	0.2%	0.1%	0.1%
SRC , TGT	35.2%	29.0%	16.0%
SRC ONLY	59.9%	63.3%	64.1%
TGT ONLY	2.3%	3.4%	6.1%
NEITHER	2.3%	4.3%	13.7%

Table 6: Percentage of phrase pairs extracted from the original pivot phrase table for each pivot class across the learning curve.

Pruning the pivot phrase table is an additional benefit from the selective combination approach. Table 6 shows that percentage of phrase pairs extracted from of the original pivot phrase table for each pivot class across the learning curve. The bulk of the phrase pairs are extracted in the classes where the source phrases exist in the direct model which add the least and sometimes hurt the overall combination performance.

For the large parallel data (4M words), selective combination with (TGT ONLY) class gives a slightly better result in BLEU while hugely reducing the size of the pivot phrase table used (2.3% of the original pivot phrase table). For smaller parallel data, the advantage is reduced but here comes

the trade off between the quality of the translation and the size of the model.

7 Conclusion and Future Work

We propose a selective combination approach between pivot and direct models to improve the translation quality. We showed that the selective combination can lead to a large reduction of the pivot model without affecting the performance if not improving it. In the future, we plan to investigate classifying the pivot model based on morphological patterns extracted from the direct model instead of just the exact surface form.

Acknowledgments

The work presented in this paper was possible thanks to a generous research grant from Science Applications International Corporation (SAIC). The last author (Sawaf) contributed to the effort while he was at SAIC. We would like to thank M. Sadegh Rasooli, Jon Dehdari and Nadi Tomeh for helpful discussions and insights into Persian.

References

- Ibrahim Badr, Rabih Zbib, and James Glass. 2008. Segmentation for English-to-Arabic Statistical Machine Translation. In *Proc. of ACL'08*.
- Nicola Bertoldi, Madalina Barbaiani, Marcello Federico, and Roldano Cattoni. 2008. Phrase-based statistical machine translation with pivot languages. *Proc. of IWSLT'08*.
- Trevor Cohn and Mirella Lapata. 2007. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proc. of ACL'07*.
- Hal Daumé III and Daniel Marcu. 2006. Domain adaptation for statistical classifiers. *J. Artif. Intell. Res.(JAIR)*.
- Matthias Eck, Stephan Vogel, and Alex Waibel. 2004. Language model adaptation for statistical machine translation based on information retrieval. In *Proc. LREC'04*.
- Ahmed El Kholy and Nizar Habash. 2010a. Orthographic and Morphological Processing for English-Arabic Statistical Machine Translation. In *Proc. of TALN'10*.
- Ahmed El Kholy and Nizar Habash. 2010c. Techniques for Arabic Morphological Detokenization and Orthographic Denormalization. In *Proc. of LREC'10*.
- Ahmed El Kholy, Nizar Habash, Gregor Leusch, Evgeny Matusov, and Hassan Sawaf. 2013. Language independent connectivity strength features for phrase pivot statistical machine translation. In *Proc. of ACL'13*.

- Jakob Elming and Nizar Habash. 2009. Syntactic Reordering for English-Arabic Phrase-Based Machine Translation. In *Proc. of EACL'09*.
- David Graff. 2007. Arabic Gigaword 3, LDC Catalog No.: LDC2003T40. Linguistic Data Consortium, University of Pennsylvania.
- Rohit Gupta, Raj Nath Patel, and Ritnesh Shah. 2012. Learning improved reordering models for Urdu, Farsi and Italian using SMT. In *Proc. of WMT'12*.
- Nizar Habash and Jun Hu. 2009. Improving Arabic-Chinese Statistical Machine Translation using English as Pivot Language. In *Proc. of EACL'09*.
- Nizar Habash and Owen Rambow. 2005. Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop. In *Proc. of ACL'05*.
- Nizar Habash and Fatiha Sadat. 2006. Arabic Pre-processing Schemes for Statistical Machine Translation. In *Proc. of NAACL'06*.
- Nizar Habash, Abdelhadi Soudi, and Tim Buckwalter. 2007. On Arabic Transliteration. In A. van den Bosch and A. Soudi, editors, *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Springer.
- Nizar Habash, Owen Rambow, and Ryan Roth. 2009. MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In Khalid Choukri and Bente Maegaard, editors, *Proc. of the Second International Conference on Arabic Language Resources and Tools*.
- Nizar Habash. 2010. *Introduction to Arabic Natural Language Processing*. Morgan & Claypool.
- Jan Hajič, Jan Hric, and Vladislav Kubon. 2000. Machine Translation of Very Close Languages. In *Proc. of ANLP'00*.
- Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proc. of WMT'11*.
- Carlos Henriquez, Rafael E. Banchs, and José B. Mariño. 2010. Learning reordering models for statistical machine translation with a pivot language.
- Almut Silja Hildebrand, Matthias Eck, Stephan Vogel, and Alex Waibel. 2005. Adaptation of the translation model for statistical machine translation based on information retrieval. In *Proc. of EAMT'05*.
- Amir Hossein Jadidinejad, Fariborz Mahmoudi, and Jon Dehdari. 2010. Evaluation of PerStem: a simple and efficient stemming algorithm for Persian. In *Multilingual Information Access Evaluation I. Text Retrieval Experiments*.
- Andreas Kathol and Jing Zheng. 2008. Strategies for building a Farsi-English smt system from limited resources. In *Proc. of INTERSPEECH'08*.
- M. Khalilov, Marta R. Costa-juss, Jos A. R. Fonollosa, Rafael E. Banchs, B. Chen, M. Zhang, A. Aw, H. Li, Jos B. Mario, Adolfo Hernandez, and Carlos A. Henriquez Q. 2008. The talp & i2r smt systems for iwslt 2008. In *Proc. of IWSLT'08*.
- Philipp Koehn and Josh Schroeder. 2007. Experiments in domain adaptation for statistical machine translation. In *Proc. of WMT'07*.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Christopher Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Christopher Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. In *Proc. of ACL'07*.
- Philipp Koehn, Alexandra Birch, and Ralf Steinberger. 2009. 462 machine translation systems for europe. *Proc. of MT Summit XII*.
- Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus. In *NEMLAR'04*.
- Evgeny Matusov and Selçuk Köprü. 2010. Improving reordering in statistical machine translation from farsi. In *Proc. of AMTA'10*.
- Mahsa Mohaghegh, Abdolhossein Sarrafzadeh, and Mehdi Mohammadi. 2012. GRAFIX: Automated rule-based post editing system to improve English-Persian SMT output. In *Proc. of COLING'12*.
- Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*.
- Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proc. of ACL'03*.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of ACL'02*.
- Mohammad Sadegh Rasooli, Ahmed El Kholy, and Nizar Habash. 2013a. Orthographical and morphological processing for persian to english statistical machine translation,. In *Proc. of IJCNLP'13*.
- Mohammad Sadegh Rasooli, Manouchehr Kouhestani, and Amirsaeid Moloodi. 2013b. Development of a Persian syntactic dependency treebank. In *Proc. of NAACL'13*.
- Ruhi Sarikaya and Yonggang Deng. 2007. Joint morphological-lexical language modeling for machine translation. In *Proc. of NAACL'07*.
- Reshef Shilon, Nizar Habash, Alon Lavie, and Shuly Wintner. 2010. Machine Translation between Hebrew and Arabic: Needs, Challenges and Preliminary Solutions. In *Proc. of AMTA'10*.
- Andreas Stolcke. 2002. SRILM - an Extensible Language Modeling Toolkit. In *Proc. of ICSLP'02*.
- Masao Utiyama and Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based statistical machine translation. In *Proc. of NAACL'07*.
- Hua Wu and Haifeng Wang. 2009. Revisiting pivot language approach for machine translation. In *Proc. of ACL'09*.
- Reyyan Yeniterzi and Kemal Oflazer. 2010. Syntax-to-morphology mapping in factored phrase-based statistical machine translation from english to turkish. In *Proc. of ACL'10*.