Alignment Symmetrization Optimization Targeting Phrase Pivot Statistical Machine Translation

Ahmed El Kholy and **Nizar Habash**
Center for Computational Learning Systems, Columbia University
475 Riverside Drive New York, NY 10115
{akholy,habash}@ccls.columbia.edu

Abstract

An important step in mainstream statistical machine translation (SMT) is combining bidirectional alignments into one alignment model. This process is called symmetrization. Most of the symmetrization heuristics and models are focused on direct translation (source-to-target). In this paper, we present symmetrization heuristic relaxation to improve the quality of phrase-pivot SMT (source-[pivot]-target). We show positive results (1.2 BLEU points) on Hebrew-to-Arabic SMT pivoting on English.

1 Introduction

One of the main issues 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. 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.

Our effort in this paper is based on phrase pivoting. We focus on word alignment to improve translation quality. Word alignment is an essential step in building an SMT system. The most commonly used alignment models, such as IBM Model serial (Brown et al., 1993) and HMM (Och and Ney, 2003), all assume one-to-many alignments. However, the target is to produce a many-to-many word alignment model. A common practice solution in most state-of-the-art MT systems is to create two sets of one-to-many word alignments (bidirectional alignments), source-to-target and target-to-source, and then combine the two sets to produce the final many-to-many word alignment model. This combination process is called "Symmetrization".

In this paper, we propose a symmetrization relaxation method targeting phrase-pivot SMT. Unlike the typical symmetrization methods, the process is carried out as an optimization for phrase-pivot SMT and eventually increase the matching on the pivot phrases. We show positive results (1.2 BLEU points) on Hebrew-Arabic phrase-pivot SMT (pivoting through English).

Next, we briefly present some background information on symmetrization (Section 2) and discuss previous related work in Section 3. This is followed by our symmetrization approach in Section 4. We present our experimental results in Section 5.

2 Background

In this section, we briefly describe different symmetrization heuristics. We then explain how symmetrization affects phrase extraction and discuss the motivation for our approach.

2.1 Symmetrization Heuristics

The simplest approach is to merge the two directional alignment functions using a symmetrization heuristic to produce a many-to-many alignment matrix (Och et al., 1999; Och and Ney, 2003; Koehn et al., 2003).

One of the approaches is to take the intersection (I) of the two directional alignments. Intersection alignment matrices are very sparse and express only one-to-one relationship between words. As a result, we get a high precision in alignment due to the agreement of both models and a very low recall.

An alternative approach is to look at the two alignments as containing complementary information. Therefore, the union (U) of the two models can capture all complementary information. Unlike the intersection (I), many-to-many relationship between words are covered and the resulting matrices are dense. As a result, we get the opposite effect of intersection where we have a higher recall of alignment points but at the cost of losing in precision.

Many mid-way solutions between intersection (I) and (U) can be achieved which aim to balance between precision and recall. Some solutions start from high precision intersection points, and progressively add reliable links from the union to increase recall. Other solutions start from a high recall union points and remove unreliable links to increase precision. One of most commonly used heuristic is **Grow-diag-final-and** (GDFA) (Koehn et al., 2003).

The GDFA heuristic is composed of two steps and one constraint. The first step (**Grow-diag**) starts from the intersection of two directional alignments then gradually considers the neighborhood of each alignment point between the source and target words. The considered neighbors of an alignment point at position (i, j) span over the range of $[i - 1, i + 1]$ for source words and $[j - 1, j + 1]$ for target words. Points in this neighborhood are progressively added to the alignment if neither the source word nor the target word is already aligned and the corresponding point exists in the union (U). The second step (**-final**) adds alignment points that are not neighbor intersection alignment points. This is done for alignment points between words, of which at least one is currently unaligned and exists in the union (U). Adding the constraint (**-and**), only allows alignment points between two unaligned words to be added.

2.2 Symmetrization vs. Phrase Extraction

There is a direct relationship between the final alignment matrix after symmetrization and the phrase extraction process. One way to look at the role of alignment points in extracting phrases is that they act as constraints for which phrase pairs can be extracted. In the standard heuristic (Koehn et al., 2003) for phrase pair extraction, the extracted phrase pair should be consistent and contain at least one word-based link. Moreover, no word inside the phrase pair is aligned to a word outside it. Figure 1 shows examples of phrase pairs that obey or violate the consistency constraint.

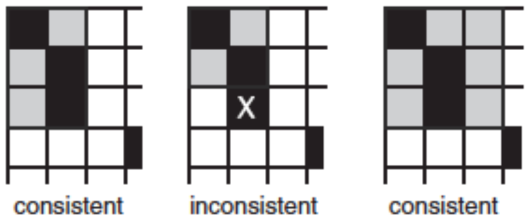

Figure 1: Phrase-pairs consistency constraints with word alignment (black squares are alignment points and the shaded area is a proposed phrase pair): The first example from the left obeys the consistency heuristic, which is violated in the second example (one alignment point in the second column is outside the phrase pair). The third example obeys the consistency heuristic despite the fact that it includes an unaligned word on the right. This diagram is taken from Koehn (2010).

The consistency constraint leads to an inverse relationship between the number of alignment points and the number of phrase pairs extracted; the fewer alignment points, the more phrase pairs can be extracted. This relationship is not valid in the extreme situation with no alignment points at all; in this extreme case, no phrase pairs are extracted.

A major issue in this heuristic is its sensitivity to word alignment errors. Since the consistency constraint is based on the alignment, an error could prevent the extraction of many good phrase pairs. In the context of phrase pivoting, this eventually leads to much less chances to pivot on potential good phrases. This problem motivates our approach to relax the symmetrization process (discussed in Section 4) and generate new pivot phrases in both systems used in pivoting. These new pivot phrases can connect potential source to target phrase pairs.

3 Related Work

Many researchers have investigated the use of pivoting (or bridging) approaches to solve the data scarcity problem (Utiyama and Isahara, 2007; Wu and Wang, 2009; Khalilov et al., 2008; Bertoldi et al., 2008; Habash and Hu, 2009). The main 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 pivot strategies have been presented in the literature. The following three are perhaps the most common.

The first strategy is the sentence pivoting technique 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 based on phrase pivoting (Utiyama and Isahara, 2007; Cohn and Lapata, 2007; Wu and Wang, 2009; El Kholy et al., a; El Kholy et al., b). In phrase pivoting, a new source-target phrase table (translation model) is induced from source-pivot and pivot-target phrase tables. Lexical weights and translation probabilities are computed from the two translation models.

The third 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).

In this paper, we build on the phrase pivoting approach, which has been shown to be the best with comparable settings (Utiyama and Isahara, 2007).

There are some recent efforts regarding alignment symmetrization or combination. In a related work to our approach but for direct SMT systems, Deng and Zhou (2009) performs alignment symmetrization as an optimization process to maximize the number of phrase translations that can be extracted within a sentence pair. There are other approaches that do not depend on heuristics. Among these are efforts that depend on unsupervised methods (Liang et al., 2006; DeNero and Macherey, 2011) where they jointly learn two directional alignment models. In another direction, Graça et al. (2007) improve bidirectional models by incorporating agreement constraints to EM training using Posterior Regularization (PR). Moreover, DeNero and Macherey (2011) proposed a model based aligner combination using dual decomposition.

Since both Hebrew 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). Most of these efforts are focused on syntactic and morphological processing to improve translation quality.

Until recently, there has not been much parallel Hebrew-English (Tsvetkov and Wintner, 2010) and Hebrew-Arabic data, and consequently little work on Hebrew-English and Hebrew-Arabic SMT. Lavie et al. (2004) built a transfer-based translation system for Hebrew-English and so did Shilon et al. (2012) for translation between Hebrew and Arabic.

Algorithm 1 Symmetrization Relaxation Algorithm (starting with union symmetrization)

```
{ generate the list of possible pivot unigram L_p }
A^U_pt = →A_pt ∪ ←A_pt
A^F_pt = A^U_pt
for (i, j) ∈ A^F_pt do
    if W_i ∉ L_p then
        A^F_pt = A^F_pt − {(i, j)}
    end if
end for
return A^F_pt
```

To our knowledge this is the first study improving phrase-pivot SMT for Hebrew-Arabic SMT. We successfully show that relaxing alignment symmetrization targeting pivoting and combining the extracted phrases with the best baseline system improve translation quality.

4 Approach

In this section, we explain our approach in relaxing the symmetrization process to improve the matching in phrase-pivot SMT. We then discuss our approach in combining the phrase pairs extracted from the basic pivot system and a pivot system using our relaxation approach which leads to our best results.

4.1 Symmetrization Relaxation

Our proposed approach is based on two parts. The first part is constructing a list of all possible pivot unigram phrases L_p that can be used in the pivoting process. This can simply be done by getting the intersection of all the pivot unigrams extracted from both the source-pivot and the pivot-target corpora.

In the second part, we start by building two directional alignment models: pivot-to-target $\overrightarrow{A_{pt}}$ and target-to-pivot $\overleftarrow{A_{pt}}$. Following Algorithm 1, we can start with union A^{U}_{pt} or grow-diag-final-and A^{GDFA}_{pt} alignment symmetrization. We then relax the symmetrization to allow the extraction of many new pivot phrases by removing a given word link that links a target word to a pivot word that is NOT in L_p. The final alignment matrix after all the deletions is A^{F}_{pt}. To remind the reader, alignment

Symm.	He-En				En-Ar				He-En-Ar	
	$\lvert A_{sp}\rvert$	$\frac{\lvert A_{sp}\rvert}{\lvert A^U_{sp}\rvert}$	$\lvert PT_{sp}\rvert$	$\frac{\lvert PT_{sp}\rvert}{\lvert PT^I_{sp}\rvert}$	$\lvert A_{pt}\rvert$	$\frac{\lvert A_{pt}\rvert}{\lvert A^U_{pt}\rvert}$	$\lvert PT_{pt}\rvert$	$\frac{\lvert PT_{pt}\rvert}{\lvert PT^I_{pt}\rvert}$	$\lvert PT_{st}\rvert$	$\frac{\lvert PT_{st}\rvert}{\lvert PT^I_{st}\rvert}$
I	0.6M	45%	15.0M	100%	0.7M	57%	11.4M	100%	1707M	100%
U	1.4M	100%	0.9M	6%	1.3M	100%	1.3M	12%	1M	0.1%
U_R	1.2M	89%	1.7M	11%	1.2M	91%	2.3M	21%	245M	14%
GDFA	1.1M	79%	3.0M	20%	1.0M	85%	3.0M	27%	267M	16%
GDFA_R	1.0M	73%	4.4M	30%	1.0M	78%	4.6M	40%	1105M	65%

Table 1: Comparison of symmetrization methods in terms of alignment set size, resulting phrase tables size (in millions) for each size SMT system used in pivoting and the final pivot phrase table.

points deletion (a.k.a alignment symmetrization relaxation) allows the extraction of more phrases.

We repeat the whole process in the other language pair of the pivoting, source-pivot, to get the final alignment set A^F_{sp}. Then, these final alignment matrices are used to extract two phrase tables PT_{sp} and PT_{pt} which are used in the phrase pivoting process to produce the final pivot phrase table PT_{st}.

Table 1 shows the impact of different word alignment symmetrization methods on phrase tables for each system used in Hebrew-Arabic phrase-pivot SMT (He-En & En-Ar) and the final phrase table (He-En-Ar).[1] We compare each method with and without our relaxation approach. The first row in the table is the intersection (I). The next two are union (U) without relaxation and then union with relaxation (U_R). The next two methods are heuristic grow-diagonal-final-and (GDFA) without relaxation and with relaxation (GDFA_R).

For each particular symmetrization method and each system used in pivoting, we compute the output alignment set size in first & fifth columns and their percentage of the union in second & sixth columns. We also compute the size of the resulting phrase tables. The numbers show the inverse relationship between the alignment set size and the phrase table sizes. The most sparse matrix in intersection leads to huge phrase tables which consequently leads a exponentially huge final pivot phrase table with potentially a lot of low quality phrase pairs. The union has an opposite effect. It has a higher recall of alignment points including some bad alignment points that can prevent the extraction of good pivoting phrase pairs.

Figure 2 illustrates how the proposed symmetrization relaxation approach can lead to good and bad English-Arabic phrase pairs.[2] The English-Arabic phrase pair (B1) is extracted into the original baseline phrase table. The word "phased" is erroneously aligned to the Arabic word وفق *wfq* 'according to/under' which prevents the extraction of smaller phrase pairs because of the consistency constraint (discussed in Section 2.2). Since the word "phased" does not appear in the English side of the Hebrew-English corpus, our relaxation method will drop all the alignment points which are connected to the word "phased". This allows the extraction of a couple of new phrase pairs (R1a & R1b). (R1a) is not a good phrase pair since it includes an extra word ("phased") in the English side that is absent in the Arabic. That said, it will not be used in the pivoting.(R1b), on the other hand, is a good phrase pair that could lead to a pivot match.

The lower half of Figure 2 illustrates how symmetrization relaxation does not always lead to good phrase pairs. The English-Arabic phrase pair (B2), which appears in the original baseline phrase table, is a perfectly good phrase pair. However, since the word "Saloniki" doesn't appear in the English side of the Hebrew-English corpora, deleting it leads to the creation of two bad phrase pairs (R2a & R2b) where the English and Arabic side do not have the same meaning.

4.2 Model Combination

The alignment symmetrization relaxation explained in Section 4.1 leads to an increase in the number of phrase pairs extracted in the translation model. Some of these phrase pairs would be useful but many others are of low quality which affects the translation choices during decoding and the overall translation quality as shown in Figure 2.

As a solution, we construct a combined phrase table using phrase pairs from the best baseline pivoting system without relaxation and then add any

[1] The experimental setup is discussed in details in Section 5.1

[2] We use the Habash-Soudi-Buckwalter Arabic transliteration (Habash et al., 2007).

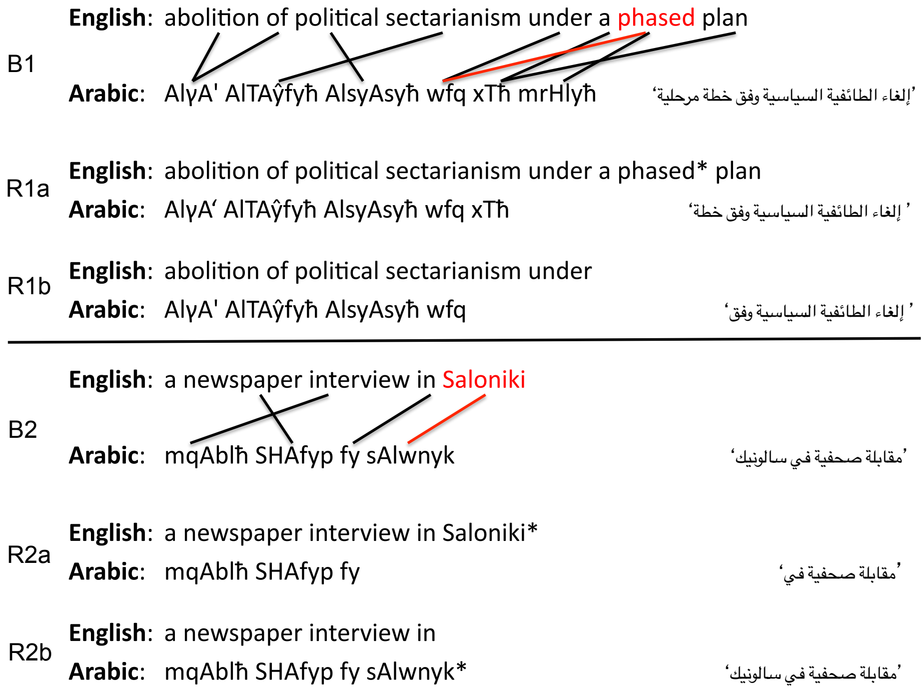

Figure 2: Two examples of baseline (GDFA) phrase pairs (B1 & B2) together with two pairs of phrases that are generated after symmetrization relaxation (R1a, R1b, R2a &R2b). The alignment links that are deleted as part of symmetrization relaxation are colored in red. The words marked with an asterisk do not have an equivalent in the opposite language in the phrase pair they appear in. The examples are discussed in detail in Section 4.1.

additional phrase pairs extracted after relaxation. We add a binary feature $f_{\mathbf{s},\mathbf{t}}$ to the log linear space of features in order to mark the source of the pivot phrase pairs as follows:[3]

$$f_{(\mathbf{s},\mathbf{t})} = \begin{cases} 2.718 & \text{if } (\mathbf{s},\mathbf{t}) \text{ from the baseline system} \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

The aim from the added binary feature is to bias the translation model after tuning to favor phrase pairs from the baseline system over the complementary phrase pairs from the relaxed model.

5 Experiments

Next, we present a set of experiments on symmetrization relaxation for phrase-pivot SMT and on model combination.

5.1 Experimental Setup

In our pivoting experiments, we build two SMT models; one model to translate from Hebrew to English and another model to translate from English to Arabic. For both models, we use the same size of parallel corpus($\approx$ 1M words) despite the fact that more English-Arabic data are available. The English-Arabic parallel corpus is a subset of available data from LDC.[4] The Hebrew-English corpus is available from sentence-aligned corpus produced by Tsvetkov and Wintner (2010).

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).

All experiments are conducted using the Moses phrase-based SMT system (Koehn et al., 2007). We use MERT (Och, 2003) for decoding weight optimization. Weights are optimized using a set of 517 sentences (single reference) developed by Shilon et al. (2010).

We use a maximum phrase length of size 8 across all models. We report results on a Hebrew-Arabic evaluation set of 300 sentences with three references developed by Shilon et al. (2010). We evaluate using BLEU-4 (Papineni et al., 2002),

[3]The log values of 2.718 and 1 will lead to a binary representation in the log linear space.

[4]LDC Catalog IDs: LDC2004T17, LDC2004E72, LDC2005E46, LDC2004T18

METEOR v1.4 (Lavie and Agarwal, 2007) and TER (Snover et al., 2006).

5.2 Linguistic Preprocessing

In this section we present our motivation and choice for preprocessing Arabic, Hebrew and English data. Both Arabic and Hebrew are morphologically complex languages (Fabri et al., 2014). 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. 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).

Similar to Arabic, Hebrew poses computational processing challenges typical of Semitic languages (Itai and Wintner, 2008; Shilon et al., 2012; Habash, 2010). Hebrew inflects for gender, number, person, state, tense and definiteness. Furthermore, Hebrew has a set of attachable clitics that are typically separate words in English, e.g., conjunctions (such as +ו *w*+ 'and'),[5] prepositions (such as +ב *b*+ 'in'), the definite article (+ה *h*+ 'the'), or pronouns (such as הם+ +*hm* 'their'). These issues contribute to a high degree of ambiguity that is a challenge to translation from Hebrew to English or to any other language. We use the best preprocessing scheme for Hebrew (HTAG) identified by Singh and Habash (2012) .

English, our pivot language, is quite different from both Arabic and Hebrew. English is morphologically poor and barely inflects for number, person and tense. English preprocessing simply includes down-casing, separating punctuation and splitting off "'s".

[5]The following Hebrew 1-to-1 transliteration is used (in Hebrew lexicographic order): *abgdhwzxTiklmns'pcqršt*. All examples are undiacritized and final forms are not distinguished from non-final forms.

Symm.	BLEU	METEOR	TER
GDFA	20.4	33.4	62.7
GDFA_R	**20.8**	**34.0**	**62.4**
U	20.1	33.5	62.7
U_R	20.7	34.0	62.5
I	20.8	34.0	63.6

Table 2: Symmetrization relaxation results for different symmetrization methods. The best performer is the relaxed grow-diag-final-and (GDFA_R). (GDFA_R) BLEU score is statistically significant over the baseline (GDFA) with p-value $= 0.12$. All other results are not statistically significant.

5.3 Symmetrization Relaxation

We compare the performance of symmetrization relaxation in contrast with different symmetrization methods. The results are presented in Table 2. In general, as expected grow-diag-final-and (GDFA) outperforms all other symmetrization methods and it is considered our baseline. Moreover, the performance improves with the symmetrization relaxation for both union (U_R) and grow-diag-final-and (GDFA_R) and the best performer is the relaxed grow-diag-final-and (GDFA_R). While (I) leads to comparable results to (GDFA_R), BLEU score against the baseline (GDFA) is not statistically significant and TER is the worst across all methods.[6]

Since (GDFA_R) is the best performing model, we use (GDFA) and (GDFA_R) in our model combination experiments, next.

5.4 Model Combination

We test the performance of combining the baseline (GDFA) phrase table with the relaxed (GDFA_R) phrase table as explained in Section 4.2.

The results in Table 3 show that we get a nice improvement of 1.2/1/0.8 (BLEU/METEOR/TER) points by combing the two models (GDFA) and (GDFA_R). The difference in BLEU score is statistically significant with p-value < 0.01. This re-

[6]Statistical significance is done using MultEval (https://github.com/jhclark/multeval) which implements statistical significance testing between systems based on multiple optimizer runs and approximate randomization (Resampling, 1989; Clark et al., 2011)

Symm.	BLEU	METEOR	TER
GDFA	20.4	33.4	62.7
GDFA_R	20.8	34.0	62.4
GDFA+GDFA_R	**21.6**	**34.4**	**61.6**

Table 3: Model combination experiment result. (GDFA+GDFA_R) shows a big improvement in BLEU score which is statistically significant with p-value < 0.01.

sult shows that our relaxation approach helps in combination with a baseline system to improve the overall translation quality. Moreover, since (GDFA_R) is a proper super-set of (GDFA) by design then the big jump in performance is due to the additional binary feature added to the log linear model. As we hoped, the binary feature biases the combined model towards the more trusted phrase pairs from (GDFA) and complement the translation model with the additional phrase pairs from symmetrization relaxation.

6 Conclusion and Future Work

We presented a symmetrization relaxation method targeting phrase-pivot SMT. The symmetrization is carried out as an optimization process to increase the matching on the pivot phrases. We show positive results (1.2 BLEU points) on Hebrew-Arabic phrase-pivot SMT. In the future, we plan to work on symmetrization based on morpho-syntactic information between Hebrew and Arabic. We expect an improvement in quality since both languages come from the same Semitic family. We also plan to work on techniques to determine the quality of pivot phrase pairs using alignment information and relationships between the three languages used in pivoting.

Acknowledgments

The work presented in this paper was funded by a Google research award. We would like to thank Reshef Shilon for helpful feedback and discussions. We also thank the anonymous reviewers for their insightful comments.

References

Bertoldi, Nicola, Madalina Barbaiani, Marcello Federico, and Roldano Cattoni. 2008. Phrase-based statistical machine translation with pivot languages. *Proc. of International Workshop on Spoken Language Translation (IWSLT'08)*.

Brown, P. F., S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*.

Clark, Jonathan H, Chris Dyer, Alon Lavie, and Noah A Smith. 2011. Better hypothesis testing for statistical machine translation: Controlling for optimizer instability. In *Proc. of Association for Computational Linguistics (ACL'11)*.

Cohn, Trevor and Mirella Lapata. 2007. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proc. of ACL'07*.

DeNero, John and Klaus Macherey. 2011. Model-based aligner combination using dual decomposition. In *Proc. of ACL'11*.

Deng, Yonggang and Bowen Zhou. 2009. Optimizing word alignment combination for phrase table training. In *Proc. of the International Joint Conference on Natural Language Processing (IJCNLP'09)*, Suntec,Singapore.

El Kholy, Ahmed and Nizar Habash. 2010a. Orthographic and Morphological Processing for English-Arabic Statistical Machine Translation. In *Proc. of Traitement Automatique du Langage Naturel (TALN'10)*. Montréal, Canada.

El Kholy, Ahmed and Nizar Habash. 2010b. Techniques for Arabic Morphological Detokenization and Orthographic Denormalization. In *Proc. of the International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

El Kholy, Ahmed, Nizar Habash, Gregor Leusch, Evgeny Matusov, and Hassan Sawaf. Language independent connectivity strength features for phrase pivot statistical machine translation. In *Proc. ACL'13*, Sofia, Bulgaria.

El Kholy, Ahmed, Nizar Habash, Gregor Leusch, Evgeny Matusov, and Hassan Sawaf. Selective combination of pivot and direct statistical machine translation models. In *Proc. IJCNLP'13*, Nagoya, Japan.

Elming, Jakob and Nizar Habash. 2009. Syntactic Reordering for English-Arabic Phrase-Based Machine Translation. In *Proc. of the EACL'09*, Athens, Greece.

Fabri, Ray, Michael Gasser, Nizar Habash, George Kiraz, and Shuly Wintner. 2014. Linguistic introduction: The orthography, morphology and syntax of semitic languages. In *Natural Language Processing of Semitic Languages*, pages 3–41. Springer.

Graça, Joao, Kuzman Ganchev, and Ben Taskar. 2007. Expectation maximization and posterior constraints. In *Proc. of Conference on Neural Information Processing Systems (NIPS'07)*.

Graff, David. 2007. Arabic Gigaword 3, LDC Catalog No.: LDC2003T40. Linguistic Data Consortium (LDC), University of Pennsylvania.

Habash, Nizar and Jun Hu. 2009. Improving Arabic-Chinese Statistical Machine Translation using English as Pivot Language. In *Proc. of the Workshop on Statistical Machine Translation (WMT'09)*, Athens, Greece, March.

Habash, Nizar and Owen Rambow. 2005. Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop. In *Proc. of ACL'05*, Ann Arbor, Michigan.

Habash, Nizar and Fatiha Sadat. 2006. Arabic Preprocessing Schemes for Statistical Machine Translation. In *Proc. of NAACL'06*, New York City, USA.

Habash, Nizar, Abdelhadi Soudi, and Tim Buckwalter. 2007. On Arabic Transliteration. In *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Springer.

Habash, Nizar, Owen Rambow, and Ryan Roth. 2009. MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In *Proc. of the Second International Conference on Arabic Language Resources and Tools*. The MEDAR Consortium.

Habash, Nizar. 2010. *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers.

Hajič, Jan, Jan Hric, and Vladislav Kubon. 2000. Machine Translation of Very Close Languages. In *Proc. of the Applied Natural Language Processing Conference (ANLP'00)*, Seattle.

Itai, Alon and Shuly Wintner. 2008. Language resources for Hebrew. In *Proc. of LREC'08*.

Kathol, Andreas and Jing Zheng. 2008. Strategies for building a Farsi-English smt system from limited resources. In *Proc. of the Annual Conference of the International Speech Communication Association (INTERSPEECH'08)*, Brisbane, Australia.

Khalilov, M., Marta R. Costa-juss, Jos A. R. Fonollosa, Rafael E. Banchs, B. Chen, M. Zhang, A. Aw, H. Li, Jos B. Mario, Adolfo Hernndez, and Carlos A. Henrquez Q. 2008. The talp & i2r smt systems for iwslt 2008. In *Proc. of IWSLT'08*

Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proc. NAACL'03*.

Koehn, Philipp, 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*, Prague, Czech Republic.

Koehn, Philipp, Alexandra Birch, and Ralf Steinberger. 2009. 462 machine translation systems for europe. *Proc. of MT Summit XII*.

Koehn, Philipp. 2010. *Statistical machine translation*. Cambridge University Press.

Lavie, Alon and Abhaya Agarwal. 2007. Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments. In *Proc. of WMT'07*, Prague, Czech Republic.

Lavie, Alon, Katharina Probst, Erik Peterson, Stephan Vogel, Lori Levin, Ariadna Font-Llitjos, and Jaime Carbonell. 2004. A trainable transfer-based machine translation approach for languages with limited resources. In *Proc. of European Association for Machine Translation (EAMT'04)*, Malta.

Liang, Percy, Ben Taskar, and Dan Klein. 2006. Alignment by agreement. In *Proc. of ACL'06*.

Maamouri, Mohamed, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus. In *NEMLAR Conference on Arabic Language Resources and Tools*, Cairo, Egypt.

Och, Franz Josef and Herman Ney. 2003a. A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics.

Och, Franz Josef, Christoph Tillmann, and Hermann Ney. 1999. Improved alignment models for statistical machine translation. In *Proc. EMNLP'99*.

Och, Franz Josef. 2003. Minimum error rate training in statistical machine translation. In *Proc. of ACL'03*.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of ACL'02*, Philadelphia, PA.

Resampling, Bootstrap. 1989. Computer-intensive methods for testing hypotheses: an introduction. *Computer*.

Shilon, Reshef, Nizar Habash, Alon Lavie, and Shuly Wintner. 2010. Machine translation between Hebrew and Arabic: Needs, challenges and preliminary solutions. In *Proc. of the Association for Machine Translation in the Americas (AMTA'10)*.

Shilon, Reshef, Nizar Habash, Alon Lavie, and Shuly Wintner. 2012. Machine translation between Hebrew and Arabic. *Machine Translation*.

Singh, Nimesh and Nizar Habash. 2012. Hebrew morphological preprocessing for statistical machine translation. In *Proc. of The European Association for Machine Translation (EAMT'12)*.

Snover, Matthew, Bonnie J. Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proc. of AMTA'06*, Boston, Massachusetts.

Stolcke, Andreas. 2002. SRILM - an Extensible Language Modeling Toolkit. In *Proc. of the International Conference on Spoken Language Processing (ICSLP'02)*, Denver, CO.

Tsvetkov, Y. and S. Wintner. 2010. Automatic acquisition of parallel corpora from websites with dynamic content. In *Proc. of LREC'10*.

Utiyama, Masao and Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based statistical machine translation. In *Proc. of ACL'07*, Rochester, New York.

Wu, Hua and Haifeng Wang. 2009. Revisiting pivot language approach for machine translation. In *Proc. of ACL-AFNLP'09*, Suntec, Singapore.

Yeniterzi, Reyyan and Kemal Oflazer. 2010. Syntax-to-morphology mapping in factored phrase-based statistical machine translation from english to Turkish. In *Proc. of ACL'10*, Uppsala, Sweden.