Arabic-English Semantic Class Alignment to Improve Statistical Machine Translation

Ines Turki Khemakhem
MIRACL Laboratory,
University of Sfax-TUNISIA
ines_turki@yahoo.fr

Salma Jamoussi
MIRACL Laboratory,
University of Sfax-TUNISIA
salma.jamoussi@isimsf.rnu.tn

Abdelmajid Ben Hamadou
MIRACL Laboratory,
University of Sfax-TUNISIA
abdelmajid.benhamadou@isimsf.rnu.tn

Abstract

Clustering words is a widely used technique in statistical natural language processing. It requires syntactic, semantic, and contextual features. Especially, semantic clustering is gaining a lot of interest. It consists in grouping a set of words expressing the same idea or sharing the same semantic properties.

In this paper, we present a new method to integrate semantic classes in a Statistical Machine Translation (SMT) context to improve the Arabic-English translation quality.

In our method, we first apply a semantic word clustering algorithm for English. We then project the obtained semantic word classes from the English side to the Arabic side. This projection is based on available word alignments provided by the alignment step using GIZA++ tool. Finally, we apply a new process to incorporate semantic classes in order to improve the SMT quality. The experimental results show that introducing semantic word classes achieves 4 % of relative improvement on the BLEU score for the Arabic → English translation task.

1 Introduction

In the past decade, statistical machine translation (SMT) has been advanced from word based SMT to phrase and syntax based SMT. Although this advancement produces major improvements in BLEU scores, important meaning errors still harm the quality of SMT translations.

More recently, research in statistical machine translation has witnessed many attempts to integrate semantic feature into SMT models, to generate not only grammatical but also meaning preserved translations.

Integrating semantic features into SMT tasks aims at improving translation adequacy. In a bilingual corpus, different senses of words in the source language can have different translations in the target language, as the context in which they appear.

This motivates the introduction of semantic word classes in statistical machine translation.

A semantic word class is represented by a set of words expressing the same idea and sharing the same semantic properties. For example, the words plane, train, boat, bus can all correspond to the semantic class "transport".

Semantic word clustering is a technique for partitioning sets of words into subsets of semantically similar words. It is increasingly becoming a major technique used in SMT task.

Furthermore, most of the SMT system well suited for processing English and other languages with a relatively rigid word order, while languages with complicated morphological paradigms still pose difficulties as Arabic.

In this paper, we present a new method to integrate the underlined semantic classes in a SMT context to improve the Arabic-English translation quality.

We first describe the semantic word clustering algorithm for English and we proceed to directly project the obtained semantic word classes from English side into Arabic side. This projection is based on available word-alignments provided by

Proceedings of Recent Advances in Natural Language Processing, pages 663–671,
Hissar, Bulgaria, Sep 7–9 2015.

the alignment step using GIZA++ tool. The rest of the paper is organized as follows.
Section 2 presents an overview of some recent approaches attempting to introduce semantic features into the statistical machine translation framework. In Section 3, we describe our method to improve the Arabic-English translation quality. In this section, we first give an overview of the baseline SMT. Then, we present the semantic word clustering algorithm for English and we proceed to directly project the obtained semantic word classes from English side into Arabic side. Finally, we introduce the proposed method to incorporate semantic word classes in SMT. Section 4 describes the experimental settings and results, which are discussed in the remainder of this Section. Finally, section 5 presents the most relevant conclusions of this work and suggest possible directions for future work.

2 Related Work

Several attempts to integrate semantic features into the statistical machine translation framework have been reported in the majority of previous works (Kevin and Smith, 2008). We provide a brief overview of some of the most recent work within this area which are relevant to the phrase based statistical machine translation approach.
Vickrey et al. (2005) build word sense disambiguation inspired classifiers to fill in blanks in partially completed translations.
Stroppa et al. (2007) add source-side contextual features into a phrase based SMT system by integrating context dependent phrasal translation probabilities learned using a decision-tree classifier. Authors obtain significant improvements on Italian-to-English and Chinese-to-English IWSLT tasks.
In Carpuat et Wu (2007), word sense disambiguation techniques are introduced into statistical machine translation; and in Carpuat et Wu (2008), authors show that dynamically-built context-dependant phrasal translation lexicons are more useful resources for phrase-based machine translation than conventional static phrasal translation lexicons, which ignore all contextual information.
Some work has been reported to improve translation quality with word classes, by using syntactic and semantic information for the SMT decoding in Baker et al. (2010).
In a previous work (Turki Khemakhem I. et al, 2010), a solution for disambiguation of the output of the Arabic morphological analyzer was presented. This method was used to help in selecting the proper word tags for translation purposes via word-aligned bitext.
In Banchs et Costa-jussà (2011), a semantic feature for statistical machine translation, based on Latent Semantic Indexing, is proposed and evaluated. The objective of this feature is to account for the degree of similarity between a given input sentence and each individual sentence in the training dataset. This similarity is computed in a reduced vector-space constructed by means of the Latent Semantic Indexing decomposition. The computed similarity values are used as an additional feature in the log-linear model combination approach to statistical machine translation. Authors obtain significant improvements on a Spanish-to-English translation task.
The system, presented in Costa-jussà et al. (2014), is Moses-based with an additional feature function based on deep learning. This feature function introduces source-context information in the standard Moses system by adding the information of how similar is the input sentence to the different training sentences. Significant improvements are reported in the task from English to Hindi.

On the other hand, there are approaches which use machine learning techniques. In Haque et al. (2009), authors have proposed syntactic and lexical context features, for integrating information about the neighboring words into a phrase-based SMT system ; and in España-Bonet et al.(2009), authors implements a standard Phrase-Based SMT architecture, extended by incorporating a local discriminative phrase selection model to address the semantic ambiguity of Arabic. Local classifiers are trained, using linguistic and context information, to translate a phrase.

3 Proposed Method

3.1 Phrase-Based Machine Translation

SMT methods have evolved from using the simple word based models (Brown et al,1993) to phrase based models (Marcu and Wong, 2002; Koehn P, 2004; Och and Ney, 2004). It has been formulated as a noisy channel model in which the target language sentence, *s* is seen as distorted by the channel into the foreign language *t*. In that, we try to find the sentence *t* which maximizes the *P(t/s)* probability:

$$argmax_t P(t|s) = argmax_t P(s|t)P(t) \quad (1)$$

here *P(t)* is the language model and *P(s|t)* is the translation model. We can get the language model from a monolingual corpus (in the target language). The translation model is obtained by using an aligned bilingual corpus.
The translation model is combined together with the following six additional feature models: the target language model, the word and the phrase penalty and the source-to-target and target-to-source lexicon model and the reordering model. These models are optimized by the decoder[1]. In our case, we use the open source Moses decoder described in (Koehn et al, 2007).

3.2 Pre-processing Step

Arabic is a morphologically complex language. In Arabic, various clitics such as pronouns, conjunctions and articles appear concatenated to content words such as nouns and verbs (Example: the Arabic word "أتتذكّروننا" corresponds in English to the sentence: "Do you remember us"). This can cause data sparseness issues. Thus clitics are typically segmented in a preprocessing step.

The aim of a preprocessing step is to recognize word composition and to provide specific morphological information about it. For Example: the word "سيخبرهم" (in English: he will notify them) is the result of the concatenation of the proclitic "س" indicating the future, enclitic "هم" (for the masculine plural possession pronoun) and the rest of the word "يخبر" (verb).

In our proposed method, each Arabic word, from the target Arabic training data, is replaced by the reduced word (obtained by removing its clitics), where clitic are featured with their morphological classes (e.g. proclitic and prefix). For example, the verbal form "سيخبرهم" can be decomposed as follows:

"enclitic _هم يخبر proclitic _س"

3.3 Extraction of English Concepts Using Clustering Methods

Our aim is to cluster input set of words W = {w_1 , . . . , w_n } into disjoint groups containing words sharing similar meaning C = {C_1 , . . . , C_k } (C forms a partition of W).
In the context of this work it is assumed that there is a semantic affinity between two words if they are topically related. For example C_i = {w_1, w_2, w_3, w_4, w_5} = {baseball, game, football, pitch, hit} would be a cluster of semantically related words.
The aim of this step is to identify the semantic concepts of the English side of the parallel corpus. The manual determination of these concepts is a very heavy task, so we should find an automatic method to achieve such a work.
To build up the appropriate concepts, the corpus words have to be gathered in several classes.
To reach our goal we used an unsupervised classification technique proposed in (Jamoussi et al., 2009). In the later, a new method to automatically extract semantic concepts for automatic speech understanding was suggested. This method gives good results. In (Jamoussi et al., 2009), authors use the average mutual information measure to compute similarities between words. They then associate to each word a vector with M elements, where M is the size of the lexicon. The j^{th} element of this vector represents the average mutual information between the word j of the lexicon and the word to be represented.

$$w_i = \begin{bmatrix} I(w_1 : w_i) \\ I(w_2 : w_i) \\ \vdots \\ I(w_j : w_i) \\ \vdots \\ I(w_M : w_i) \end{bmatrix}$$

Where

$$I(w_i : w_j) = P(w_i, w_j) log \frac{P(w_i|w_j)}{P(w_i)P(w_j)} + P(\overline{w_i}, w_j) log \frac{P(\overline{w_i}|w_j)}{P(\overline{w_i})P(w_j)} +$$

$$P(w_i, \overline{w_j}) log \frac{P(w_i|\overline{w_j})}{P(w_i)P(\overline{w_j})} + P(\overline{w_i}, \overline{w_j}) log \frac{P(\overline{w_i}|\overline{w_j})}{P(\overline{w_i})P(\overline{w_j})}$$

$P(w_i,w_j)$ is the probability to find w_i and w_j in the same sentence, $P(w_i|w_j)$ is the probability to find w_i knowing that we already met w_j, $P(w_i)$ is the probability of w_i and $P(\overline{w_i})$ is the probability of any other word except w_i .
To combine context and mutual information vector, (Jamoussi et al., 2009) represent each word by a matrix M×3 of average mutual information measures. The first column of this matrix corresponds to a vector of average mutual information, the second column represents the average mutual information measures between the vocabulary words and the left context of the represented word. The third column is determined in the same manner but it concerns

[1] http://www.statmt.org/moses/

the right context. The j^{th} value of the second column is the weighted average mutual information between the j^{th} word of the vocabulary and the vector constituting the left context of the word W_i. It is calculated as follows:

$$IMM_j(C_l^i) = \frac{\sum_{w_l \in L_{w_i}} I(w_j : w_l) * K_{wl}}{\sum_{w_l \in L_{w_i}} K_{wl}}$$

Where $IMM_j(C_l^i)$ is the average mutual information between the word w_j of the lexicon and the left context of the word W_i. L_{Wi} is a set of words belonging to the left context of W_i. $I(w_j:w_l)$ represents the average mutual information between the word j of the lexicon and the word w_l belonging to the left context of W_i. K_{wl} is the occurrence number of the word w_l found in the left context of W_i. The word W_i is thus represented by the matrix shown in the figure 1.

$$w_i = \begin{bmatrix} I(w_1 : w_i) & IMM_1(C_l^i) & IMM_1(C_r^i) \\ I(w_2 : w_i) & IMM_2(C_l^i) & IMM_2(C_r^i) \\ \vdots & \vdots & \vdots \\ I(w_j : w_i) & IMM_j(C_l^i) & IMM_j(C_r^i) \\ \vdots & \vdots & \vdots \\ I(w_M : w_i) & IMM_M(C_r^i) & IMM_M(C_l^i) \end{bmatrix}$$

Figure 1: The matrix representation of the word W_i

The matrix representation of words as described previously, exploits a maximum of information related to a given word. It considers its context and its similarity to all the other words in the corpus. We use then the PAM method, proposed by (Kaufman and Rousseeuw, 1990), for classification of words in the corpus. We obtain a coherent list of concepts that will be used in our statistical translation system.

3.4 Projection of English Concepts to Arabic

After extracting English concepts, we proceed to directly project those concepts from English side into Arabic side. This projection is based on available word-alignments provided by the alignment step using GIZA++ tool. This projection is performed in three main steps:

- Each English word of the parallel corpus is combined with its respective semantic class. In the other side, Arabic words are kept unchanged.
- This obtained bilingual corpus is automatically word aligned by the alignment toolkit.

Arabic-English sentence alignment is illustrated in Figure 2, where each Arabic morpheme is aligned to one or zero English word and its semantic classes.

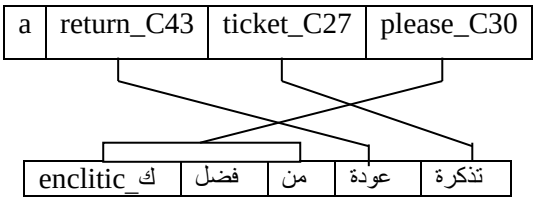

Figure 2. An example of word alignment

The alignment model was trained with the popular toolkit GIZA++ (Och and Ney, 2003), which implements the most typical IBM and HMM alignment models for translation. The alignment models used in our case are IBM-1, HMM, IBM-3 and IBM-4.

After this alignment step we obtain one model table containing English words and its respective semantic classes, aligned with Arabic words with an alignment probability.

- The obtained table is sorted and the probability that correspond to the same Arabic word and the same semantic class is added. Then the resulting probabilities are sorted, and the semantic class that corresponds to the maximum probability is selected.

Finally a matching table is got, where each line from this table refers to the corresponding Arabic word in the training corpus and its semantic class projected from the English word.

your_C9 0,0124172 امتعة baggages_C27 0,5000000 امتعة baggage_C27 0,2914286 امتعة luggage_C27 0,4179104 امتعة my_C32 0,0282752 امتعة out_C8 0,0024038 امتعة things_C49 0,0208333 امتعة	1. English words and its respective semantic classes, aligned with Arabic words with an alignment probability
C9 0,0124172 أمتعة C27 1,209339 أمتعة C32 0,0282752 أمتعة C8 0,0024038 أمتعة C49 0,0208333 أمتعة	2. Probabilities corresponding to the same Arabic word and the same semantic class are added
C27 1,209339 أمتعة C32 0,0282752 أمتعة C49 0,0208333 أمتعة C9 0,0124172 أمتعة C8 0,0024038 أمتعة	3. Semantic class which correspond to the maximum probability is selected.

Figure 3. An example of projection of English concepts to Arabic

3.5 SMT Using Semantic Word Classes

The translation model of most phrase-based SMT systems is parameterized by two phrasal and two lexical channel models (Koehn et al., 2003) which are estimated as relative frequencies. Their counts are extracted heuristically from a word aligned bilingual training corpus.

Our phrase-based baseline system is built upon the open-source MT toolkit Moses (Koehn et al, 2007). Phrase pairs are extracted from word alignments generated by GIZA++ (Och and Ney, 2003). The phrase-based translation model provides direct and inverted frequency-based and lexical-based probabilities for each phrase pair. The English side of the training corpora was used to generate 3-gram target language model for the translation task. For this purpose, the SRI language modeling toolkit (Stolcke, 2002) was used.

To incorporate semantic word classes in our SMT process, we first proceed to run the semantic word clustering algorithm for English side of the bilingual training data, as already described in section 3.3, to cluster the vocabulary into semantic classes. The obtained classes are directly projected from English side into Arabic side.
Then, we replace each word on both source and target side of the training data with their respective semantic word classes.

By considering the same training procedure as usual, we can easily train the standard models conditioned on word classes.
We obtain finally two phrase tables, the first one with word identities and the second with semantic word classes.

By considering both sorted tables simultaneously, we can select the translation for Arabic word in input test. However, each Arabic word (s_i) in the test corpus is mapped to a single semantic class c_i. We can first uses the phrase table based on word classes to select the translation for this semantic class (c_i'). The translation of the source word (s_i) is among the words of the class (c_i'). Then, to generate the target word e_i (translation of s_i), we uses the generated phrase table based on word identities. Our approach is shown in Figure 4.

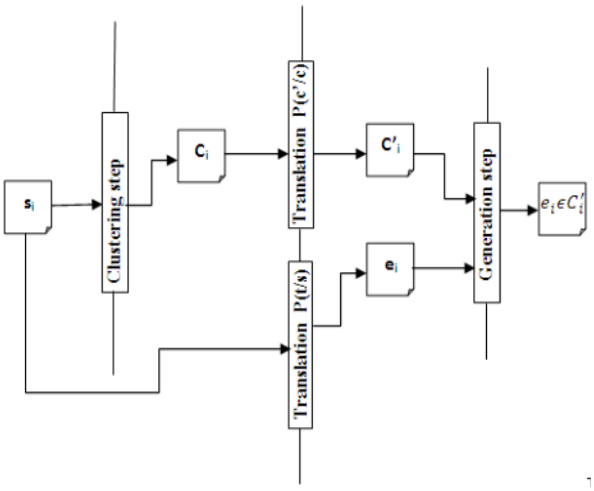

Figure4. The proposed approach: SMT using semantic word classes

Our approach to integrate semantic word classes in SMT process is performed in four main steps:

- Clustering step: (Input: word s_i , output: semantic class c_i)
Each word on source side of the test corpus s_i is replaced by their respective semantic word classes c_i.

- Translation P(t/s): (Input : word s_i ,output: E_i: list of translation of s_i)
The phrase table based on word identities is used to select the list of the translation of the word s_i .

- Translation P(c'/c): (Input : class c_i ,output semantic class c_i')
The phrase table based on word classes is used to select the translation for the semantic class c_i (c_i')

- Generation step: (Input: E_i: list of translation of s_i, semantic class c_i'; output: e_i (translation of s_i) $\in$ c_i')
The target word e_i (translation of s_i), witch is among the words of the class (c_i'), is generated.

4 Experiments

This section describes the experimental work conducted to evaluate the incidence of the proposed method to integrate semantic classes in a SMT context on translation quality. First, subsection 4.1 describes the used dataset. Then, subsection 4.2 presents and discusses the results.

4.1 Used Resources

Our experiments are performed on an Arabic → English task. We train the system on the data provided for the evaluation campaign of the International Workshop on Spoken Language Translation (IWSLT 2010) task[2].

The test set is made up of 507 sentences, which corresponds to the IWSLT08 data (there were 16 English reference translations for each Arabic sentence).

To confirm our results we also run experiments on the Arabic → English task of the IWSLT 2014 evaluation campaign[3].
Table 1 presents the main statistics related to the used data.

		Arabic	English
Train (IWSLT 2010)	sentences	19972	
	words	18149	7296
Test (IWSLT 2008)	sentences	507	
	words	459	184
Train (IWSLT 2014)	sentences	155047	
	words	162148	65774
Test (tst 2010)	sentences	3138	
	words	8101	5733

Table 1: Corpus description of the Arabic→English translation tasks.

4.2 Experimental Results

The proposed method is evaluated on the Arabic-to-English translation task, using the MOSES framework as baseline phrase-based statistical machine translation system (Koehn et al., 2007). The performances reported in this paper were measured using the BLEU score (Papineni et al., 2002).

a- Pre-processing Step:

The Arabic part of the bitext was systematically segmented to train the phrase tables.
Thus each Arabic word of the training corpus is replaced by its segmentation according to the "proclitic stem enclitic" form, as described in section 3.2.

To perform morphological decomposition of the Arabic source, we use the morphological analyzer MADA (Habash et al, 2009).

MADA is a system for Morphological Analysis and Disambiguation for Arabic. MADA produces for each input word a list of analyses specifying every possible morphological interpretation of that word, covering all morphological features of the word (diacritization, POS, lemma, and 13 inflectional and clitic features). MADA then uses SVM-based classifiers for features (such as POS, number and gender, etc.) to choose among the different analyses of a given word in context.

The resulting corpus was paired with the word-based English corpus to train the translation model. The translation table was trained using the so obtained parallel data (no change was made on the English side). In decoding, the same segmentation form was also applied to the test input.

b- SMT Using Semantic Word Classes:

In this section, we investigated to incorporate semantic word classes in Arabic-English SMT task.
We first proceed by running the semantic word clustering algorithm for English side of the bilingual training data to cluster the vocabulary into 100 classes each. The obtained classes are directly projected from English side into Arabic side.

We train the models conditioned on word classes as described above. We also train the models based on word identity, by using the same training data.
Table 2 presents the score BLEU, measured over the test sets, for three different Arabic → English SMT systems : the baseline system, a second system using the pre-processing step (pp-SMT), and a third system integrating the semantic word class in the SMT process (swc-SMT)

System	Test (IWSLT 2008)	Test 2010
Baseline	40.69	21.42
pp-SMT	42.15	22.59
Swc-SMT	43.75	23.77

Table 2: Comparison of the Arabic-English translation systems

[2] Basic Travel Expression Corpus (BTEC) 2010
[3] Basic Travel Expression Corpus (BTEC) 2014

As seen from the table, the system implementing the semantic word classes outperforms the pp-SMT system by almost 1.4 absolute BLEU point.

To confirm our results we also run experiments on the English → Arabic task of the IWSLT evaluation campaign. In this case, both training and decoding phases use Arabic segmented words. The final output of the decoder will be also composed of segmented words. Therefore these words must be recombined into their surface forms. Therefore we apply reconstruction of the Arabic segmented words just by agglutinating the morphological segments in the following order:

Proclitic + stem + enclitic.

For example: in the segmented words:

"سلمت ك ذلك ال كتاب"

The clitic "ك" can be recombined with the previous word ("ك": enclitic).

So the segmented words "سلمت ك ذلك ال كتاب" cans be recombined to "سلمتك ذلك ال كتاب", in English: "I gave this book". The clitic "ك" can be recombined also with the following word ("ك": proclitic), in this case, the segmented words "سلمت ك ذلك ال كتاب" can be recombined to "سلمت كذلك ال كتاب", in English: "I also gave the book".

Those two sentences have the same segmented form, but they have different meanings. By introducing morphological features (e.g. proclitic and enclitic) for each segment, we may remove this ambiguity.

The English-Arabic translation performance of this English-Arabic SMT system is reported in table 3. We show that the swc-SMT yields 0.8% BLEU.

System	Test (IWSLT 2008)	Test 2010
Baseline	12.86	9.3
pp-SMT	13.14	10.1
Swc-SMT	14.07	10.91

Table 3: Comparison of the English-Arabic translation systems

4.3 Discussion

Experimental results on an Arabic-to-English translation task on the corpus showed significant improvements. In this work, we integrate semantic word classes in Arabic to English SMT context for improving machine translation quality. With this, we expect to reduce the noise resulting from data sparseness problems.

To better illustrate the concepts discussed here, let us consider the Arabic word "أم" and the corresponding English translations for its two senses: "mother" and "or". Both translations can be automatically inferred from training data; and Table 4 illustrates the resulting probability values derived for both senses of the Arabic word "أم" from the actual training dataset used in this work.

phrase	φ(f|e)	lex(f|e)	φ(e|f)	lex(e|f)
{أم|||or}	0.5652096	0.720501	0.284662	0.318320
{أم|||mother}	0.264679	0.120287	0.407367	0.435377

Table4. Actual probability values for the two possible translations of the Arabic word "أم".

Notice from the table, how in general the most probable sense of "أم" in our considered dataset is "or". This actually happens because the English word "or" is always related to the Arabic word "أم". Whereas by integrating semantic word classes in the SMT system, the English word "mother" can refer to the Arabic word "أم".

5 Conclusion

We have presented a method to integrate semantic word classes in a Arabic to English SMT context for improving machine translation quality. In our method, we first have applied a semantic word clustering algorithm for English. Then, we have projected the obtained semantic word classes from the English side to the Arabic side. This projection is based on available word alignments provided by the alignment step using GIZA++ tool. Finally, we have applied a new process to incorporate semantic classes in order to improve the SMT quality.

We have shown the efficiency of this method on Arabic to English translation tasks. To confirm our results we have also run experiments on the English → Arabic task.
In our experiments, the baseline is improved by 1.4% BLEU on the Arabic → English task and by 0.3% BLEU on the English → Arabic task.
In future work we plan to apply our method to a wider range of languages.

References

Baker K., Bethard S., Blodgood M., Brown R., Callison-Burch C., Copper-smith G., Dorr B., Filardo W., Giles K. (2009). *Semantically Informed Machine Translation.* Final report of the 2010 Summer Camp for Advanced Language Exploration (SCALE).

Banchs R. and Costa-jussà M. (2011). *A semantic feature for statistical machine translation*. In proceedings of SSST-5, Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation, ACL HLT 2011, Portland, Oregon, USA, 126-134.

Brown. P., Della Pietra V., Della Pietra S., and Mercer R. 1993. *The mathematics of statistical machine translation: parameter estimation*, Computational Linguistics, 19(1): 263–311.

Carpuat, M., Wu, D. (2007) *How Phrase Sense Disambiguation Outperforms Word Sense Disambiguation for Statistical Machine Translation.* In: 11th International Conference on Theoretical and Methodological Issues in Machine Translation. Skovde

Carpuat, M., Wu, D. (2008). *Evaluation of Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation*. In: 6th International Conference on Language Resources and Evaluation (LREC). Marrakech.

Costa-jussà M., Gupta P., Banchs R. and Rosso P. (2014). *English-to-Hindi system description for WMT 2014: Deep Source-Context Features for Moses.* In proceedings of the Ninth Workshop on Statistical Machine Translation, Baltimore, Maryland USA, 79–83.

España-Bonet C., Gimenez J., Marquez L. (2009). *Discriminative Phrase-Based Models for Arabic Machine Translation*. ACM Transactions on Asian Language Information Processing Journal (Special Issue on Arabic Natural Language Processing)

Habash, N., Rambow, O., and Roth, R. 2009. *MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization*, Proceedings of the Second International Conference on Arabic Language Resources and Tools.

Haque R., Naskar S. K., Ma Y., Way A. (2009). *Using Supertags as Source Language Context in SMT*. In 13th Annual Conference of the European Association for Machine Translation, pp. 234--241. Barcelona

Jamoussi S. (2009). *New Word Vector Representation for Semantic Clustering*. TAL 50(3): 23-57.

Kaufman L. et Rousseeuw P. J. (1990). *Finding groups in data : An introduction tocluster analysis*. John Wiley & Sons (New York), 19-20.

Kevin G. and Smith N. A. (2008). *Rich Source-Side Context for Statistical Machine Translation.* In Proceedings of the Third Workshop on Statistical Machine Translation

Koehn P. 2004. *Pharaoh: A Beam Search Decoder for phrase-based Statistical Machine Translation Models*. In R. Frederking & K. Taylor (eds.) Machine Translation: From Real Users to Research; 6th Conference of the Association for Machine Translation in the Americas, AMTA, Berlin/Heidelberg, Germany: Springer Verlag, 115–124.

Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowa B., Shen W., Moran C., Zens R., Dyer C., Bojar O., Constantin A., and Herbst E., 2007. *Moses: Open source toolkit for statistical machine translation.* In Proceedings of the ACL Demoand Poster Sessions, Prague, Czeck Republic, 177–180.

Marcu D. and Wong W. 2002. *A Phrase-Based, Joint Probability Model for Statistical Machine Translation.* Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA, 133-139.

Och F. J., and Ney H., 2003. *A Systematic comparison of various statistical alignment models.* Computational Linguistics, 29(1): 19-51.

Och F. J., Ney H. 2004. *The alignment template approach to statistical machine translation*. Computational Linguistics, 30(4): 417-449.

Papineni K. A., Roukos S., Ward T., and Zhu W.J., 2002. *Bleu: a method for automatic evaluation of machine translation*. The Proc. of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, 311–318.

Stolcke A., 2002. *SRILM an Extensible Language Modeling Toolkit.* The Proc. of the Intl. Conf. on Spoken Language Processing, Denver, CO, USA, 901–904.

Stroppa N., van den Bosch A., and Way A. (2007). *Exploiting source similarity for SMT using context-informed features*. In Proc. of TMI.

Turki Khemakhem I., Jamoussi S., and Ben Hamadou A. (2010). *Arabic morpho-syntactic feature*

disambiguation in a translation context. In Proceedings of SSST-4, Fourth Workshop on Syntax and Structure in Statistical Translation, COLING, Beijing, 61-65.

Vickrey D. ,Biewald L., Teyssier M. and Koller D. (2005). *Word-sense disambiguation for machine translation*. In Proc. of HLT-EMNLP.