

LIMSI @ IWSLT 2010

Alexandre Allauzen, Josep M. Crego, İlknur Durgar El-Kahlout, Le Hai-Son,
Guillaume Wisniewski and François Yvon

LIMSI/CNRS and Université Paris-Sud 11, France
BP 133, 91403 Orsay Cédex

{allauzen, jmcrego, ilknur, lehaison, wisniewski, yvon}@limsi.fr

Abstract

This paper describes LIMSI's Statistical Machine Translation systems (SMT) for the IWSLT evaluation, where we participated in two tasks (*Talk* for English to French and *BTEC* for Turkish to English). For the *Talk* task, we studied an extension of our in-house n -code SMT system (the integration of a bilingual reordering model over *generalized translation units*), as well as the use of training data extracted from Wikipedia in order to adapt the target language model. For the *BTEC* task, we concentrated on pre-processing schemes on the Turkish side in order to reduce the morphological discrepancies with the English side. We also evaluated the use of two different continuous space language models for such a small size of training data.

1. Introduction

LIMSI took part in the IWSLT 2010 evaluation for two different tasks: *Talk* and *BTEC*. The goal of the new *Talk* task is to translate public speeches on a variety of topics, from English to French. Since the allowed training data includes the parallel corpora distributed by the ACL 2010 Workshop on Statistical Machine Translation (WMT), our starting system is the one submitted to the evaluation campaign [1]. We enhanced our in-house n -code SMT system with an additional reordering model which is estimated as a standard n -gram language model over *generalized translation units* (part-of-speech in the described experiments). In order to add more closely related training data, the use of Wikipedia as an additional source of monolingual text for the target language model was also evaluated.

For the *BTEC* task, the LIMSI participated in the Turkish to English translation track with a system based on the open source Moses system [2]. The linguistic discrepancies between these two languages appear both

at the syntactic and at the morphological level. For instance, the morphology of Turkish is productive as well as agglutinative and this yields to a large number of different word forms. To counteract this effect, we experimented different pre-processing schemes for Turkish, and we also evaluated the use of continuous space language models which might lessen the sparsity issues.

As our submissions to the two tasks differ greatly in terms of language pairs, systems and corpora, the rest of this paper is organized as follow: the system developed for the *Talk* task is first described in Section 2, while Section 3 reports our work on Turkish pre-processing and on the use of continuous space language models.

2. TALK task

2.1. n -code SMT system

Our in-house n -code SMT system implements the bilingual n -gram approach to statistical Machine Translation [3]. A translation hypothesis t given a source sentence s is defined as the sentence which maximizes a linear combination of feature functions:

$$\hat{t}_1^I = \arg \max_{t_1^I} \left\{ \sum_{m=1}^M \lambda_m h_m(s_1^J, t_1^I) \right\}, \quad (1)$$

where s_1^J and t_1^I respectively denote the source and the target sentences, and λ_m is the weight associated with the feature function h_m . The most important feature is the log-score of the translation model based on bilingual units called *tuples*. The probability assigned to a sentence pair by the translation model is estimated by using the n -gram assumption:

$$p(s_1^J, t_1^I) = \prod_{k=1}^K p((s, t)_k | (s, t)_{k-1} \dots (s, t)_{k-n+1})$$

where s refers to a source symbol (t for target) and $(s, t)_k$ to the k^{th} tuple of the given bilingual sentence pairs. It is worth noticing that, since both languages are linked up in tuples, the context information provided by this translation model is bilingual. As for any standard n -gram language model, our translation model is estimated over a training corpus composed of parallel sentence pairs. Tuples represent the core elements of our statistical machine translation (SMT) system. They are extracted from a word aligned corpus (using GIZA++ with default settings) in such a way that a unique segmentation of the bilingual corpus is achieved, allowing to estimate the n -gram model. Figure 1 presents a simple example illustrating the unique tuple segmentation for a given word-aligned pair of sentences (top).

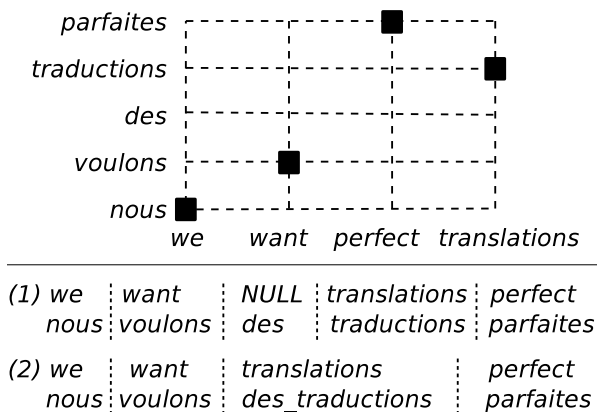


Figure 1: *Tuple extraction from a sentence pair.*

The resulting sequence of tuples (1) is further refined to avoid *NULL* words in source side of the tuples (2). Once the whole bilingual training data is segmented into tuples, n -gram language model probabilities can be estimated. In this example, note that the English source words *perfect* and *translations* have been reordered in the final tuple segmentation, while the French target words are kept in their original order.

In addition to the translation model, *eleven* feature functions are optimally combined using a discriminative training framework [4]: a *target-language model*; four *lexicon models*; two *lexicalized reordering models* [5] aiming at predicting the orientation of the next translation unit; a 'weak' distance-based *distortion model*; and finally a *word-bonus model* and a *tuple-bonus model* which compensate for the system preference for short translations. The four *lexicon models* are similar to the ones use in a standard phrase based system: two scores correspond to the relative frequencies of

the tuples and two lexical weights estimated from the IBM 4 word alignments.

During decoding, the source sentences are encoded in the form of word lattices containing the most promising reordering hypotheses, so as to reproduce the word order modifications introduced during the tuple extraction process. Hence, at decoding time, only those reordering hypotheses encoded in the word lattice are translated. Reordering hypotheses are introduced following a set of reordering rules automatically learned from the bi-text corpus word alignments.

Following on the previous example, the rule [*perfect translations* \rightsquigarrow *translations perfect*] produces the swap of the English words that is observed for the French and English pair. Typically, part-of-speech (POS) information is used to increase the generalization power of such rules. Hence, rewriting rules are built using POS rather than just surface word forms. See [6] for details on tuple extraction and on reordering rules.

2.2. Bilingual Data

All the available textual corpora are processed and normalized using in-house tools. Previous experiments revealed that using better normalization tools provides a significant reward in BLEU . The downside is the need to post-process our translation hypotheses so as to 'detokenize' them for scoring purpose, a process that is not entirely error prone. Based again on previous experiments, our systems are built in 'true case': the first letter of each sentence is lowercased when it should be, and the remaining tokens are left as is.

Finally note that the n -code systems require the source to be morpho-syntactically analysed; the same holds for the POS-based bilingual n -gram model presented in section 2.4. In all cases, POS tagging is performed by the TREETAGGER¹. Table 1 reports the number of sentences in the bitexts used for this evaluation.

Table 1: *Number of parallel sentences.*

Bitext	Sentences
GigaFrEn	20,864,682
UN data	7,078,557
Europarl-v5	1,666,183
News Co.'2010	84,251
TED-1.1	86,225
Total	29,779,898

¹<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger>

2.3. Language Models

Since the constrained task includes the same monolingual training data as the evaluation of the *Workshop on statistical Machine Translation 2010*, the target language model (LM) for French is very similar to the one described in [1]. As an overview, we use all the authorized news corpora. The overall training corpus contains 1.4 billion tokens and is divided into several sets based on dates and genres (9 sets). On each set, a standard 4-gram LM is estimated with in-house tools using the standard modified Kneser-Ney smoothing. The resulting LMs are then linearly combined using interpolation coefficients chosen so as to minimize perplexity of the development set described in section 2.5.

In order to add more closely related training data, a 4-gram LM language model is estimated on data extracted from the French Wikipedia. For this purpose, all the texts are roughly extracted, filtered and tokenized from the current version of Wikipedia. After filtering and preprocessing, the resulting corpus contains 40 million tokens.

2.4. A bilingual n -gram reordering model

An additional reordering model is experimented, which is estimated as a standard n -gram language model, over *generalized translation units*. In the experiments reported below, we generalize the tuples using POS tags instead of surface forms. Figure 2 displays the same sequence of tuples displayed in Figure 1 but built from POS tags.

PP VBP NNS JJ
 $PRO:PER$ $VER:pres$ $PRP:det_NOM$ ADJ

Figure 2: Sequence of tuples built from POS-tags.

Generalizing units greatly reduces the number of symbols in the model and enables to take larger n -gram contexts into account: in the experiments reported below, we used a context of up to 5 units (6-grams). This new model is thus helping to capture the mid-range syntactic reordering rules that are observed in the training corpus. This model can also be seen as a translation model of the sentence structure. It models the adequacy of translating sequences of source POS tags into target POS tags. Additional details on these new reordering models are given in [7].

2.5. Results

n -code systems are tuned using the implementation of minimum error rate training (MERT) [4] distributed with the Moses decoder. This implementation of MERT is slightly modified to match the requirements of our decoder. The BLEU score is used as objective function for MERT and to evaluate the test performance. The interpolation weights for language models are tuned on a held-out subset of the training TED-1.1 corpus. MERT optimization is carried out over our talk-tune (the first 862 lines) and tested over our talk-test (the last 862 lines) extracted from the official TALK development data set distributed for this year's evaluation.

Table 2 reports translation accuracy results for several configurations of our n -code system. We also evaluate the impact of using the monolingual French texts extracted from Wikipedia integrated into our target n -gram language model.

Table 2: Translation accuracy results in terms of BLEU .

Configuration	talk-tune	talk-test
base	35.82	35.35
base+bil6g	35.60	35.22
base-wikipedia	35.40	35.32

We can observe that all configurations achieve very similar accuracies. Moreover the bilingual reordering model, as well as the use of Wikipedia do not yield to a significant BLEU improvement.

3. Turkish-English BTEC Track

State-of-the-art SMT systems rely on word forms to estimate their translation models. When the parallel data is limited and the involved languages are linguistically different, applying preprocessing is crucial for improving translation accuracy.

Turkish to English machine translation is an interesting problem for several reasons. The productive morphology of Turkish implies a large vocabulary size, hence data sparsity issues. Turkish has also a very complex agglutinative morphology, where a single word may correspond to a complete phrase of several words in English. On the other hand, the word orders of these languages differ and induce long distance reordering patterns: English has subject-verb-object (SVO) word order while Turkish has a more flexible word order, but

mainly subject-object-verb (SOV). These linguistic discrepancies strongly affect the reliability of the standard methods used in SMT, especially the word alignment process and the phrase extraction step. Moreover, the limited amount of the parallel data worsens all of the problems stated above.

3.1. Previous Work

In the last few years, statistical machine translation of English to Turkish has been addressed by many researchers. Durgar El-kahlout and/or Oflazer [8, 9, 10, 11] use morphological analysis to separate some Turkish inflectional morphemes that have counterparts on the English side. Recently, Yeniterzi and Oflazer [12] applied syntactic transformations such as joining function words on the English side to the related Turkish content words. Also, an important amount of effort was spent on Turkish to English SMT in the last year's IWSLT BTEC task by several research groups [13, 14, 15, 16, 17, 18].

Using morphology in statistical machine translation with morphologically complex languages has also been addressed for several languages and especially for German: Niessen and Ney [19] use morphological decomposition with base forms and POS tags to introduce a hierarchical lexicon model; Corston-Oliver and Gamon [20] and Koehn [21] normalize inflectional variants by replacing word forms with stems on both sides; Yang and Kirchhoff [22] discuss the use of phrase-based back-off models at the test time to translate words that are unknown to the decoder, by morphologically decomposing the unknown source words.

Other studies have also been carried out for other languages that can be of interest for Turkish: for Arabic, Lee [23] and Sadat and Habash [24] use morphology to analyze and/or tag parallel corpus for translation; for Spanish, Catalan and Serbian, Popovic and Ney [25] investigate improving translation quality from inflected languages by using stems, suffixes and part-of-speech tags; for Czech, Goldwater and McClosky [26] use morphological analysis on the Czech side to introduce lemmas and pseudo words; Avramidis and Koehn [27] use syntax tree to annotate English for English to Greek and Czech translation; Minkov et al. [28] perform morphological postprocessing *on the target side* using structural information and information from the source side, to improve translation quality of translation into Russian and Arabic.

3.2. Turkish Preprocessing

We exploit the available resources by preprocessing the Turkish texts. Preprocessing scheme (mostly for Turkish) is divided in the following steps:

- **Tokenization:** For both languages, we used in-house tokenizers. For Turkish, the tokenization process was trivial. We only separated punctuations.
- **Morphological analysis and disambiguation:** Turkish texts are morphologically analyzed [29] and disambiguated [30]. For all experiments, Turkish words are represented with stems and lexical morphemes. For example, Turkish word *evin* (your house) is represented as *ev+hn*. After disambiguation, Turkish texts are lowercased but we kept English texts in true-case.
- **Frequency based segmentation:** The agglutinative structure of Turkish causes a Turkish word to be typically aligned with a complete phrase on the English side. This problem can be solved by applying segmentation to Turkish word to split them into smaller units. As the state-of-the-art translation systems are capable of learning phrase translations, there is no need to segment more than the phrase extraction can learn.

For this evaluation, we introduce a frequency based segmentation of Turkish words. Words are segmented iteratively only if their frequency falls under the specified threshold. At each iteration, only the last morpheme is separated and the frequency is checked for the remaining prefix until this frequency exceeds the given threshold. Otherwise, the word is separated into all its components. Let us consider for instance, the Turkish word *ver+ma+dh+m* (I did not give): if its count is lower than the threshold, the last morpheme (marked by +) is split. Then if the frequency of the remaining part (*ver+ma+dh*) is above the threshold, we use *ver+ma+dh+m* as the segmentation of this word. Otherwise, the word will be split one more time as *ver+ma+dh+m* and so on. We carried out different experiments, with threshold values 5, 10 and 15. During our development, the best BLEU improvement was achieved with a threshold of 10. Low threshold values tend to split less words into mor-

phemes. On the other hand, higher threshold values segments more words. For example, the morphemes in the following Turkish words *dolar+sh*, *ol+ma+dh*(the dollar, did not happen) are kept when the threshold selected as 5 but splitted into morphemes as *dolar +sh*, *ol +ma +dh* when the threshold is 15.

- **Question inversion:** It has been observed that one gets better alignments and hence better translation results when the word orders of the source and target languages are more or less the same. The word order for interrogative sentences differ greatly in Turkish and English, though these sentences can be easily reordered to get a more monotonic alignment. Turkish introduces the functional word *mu* (generally at the end of the sentence) to indicate the interrogative sentence while English just places the auxiliary verb at the beginning of the sentence. Question reordering is mentioned by Niessen and Ney [31] to modify interrogative sentences. We followed a different way, instead of inverting question words, we tend to move question tags in order to get more similar word order between two languages. We thus move the words that have the *+Ques* tag in their morphological analysis to the beginning of the sentence. For instance, in the question *japon+ca konus+yabil+hyor mu+shn*, the word *mu+shn* is moved to the beginning of sentence as **mu+shn japon+ca konus+yabil+hyor ?** to get more monotonic alignment with English sentence (**do you speak Japanese ?**).
- **Short distance morpheme reordering:** We applied very local source word reordering for a certain class of morphemes so that the word/morpheme order in an Turkish word has a more or less monotonic alignment with the word order of the corresponding English word/phrase. We moved the case morphemes *dative*, *ablative*, *genitive*, *locative* and *accusative* (whenever they are separated by the segmentation) in front of the word and removed the (*+sh*) morpheme as this morpheme does not have a real counterpart in English. For example, the Turkish word *ev+lar +sh +nda* (in the houses) became *+nda ev+lar* by moving locative morpheme to the beginning of sentence. This allows monotonic alignment within phrase pairs.

- **Augmentation of training data with open-class words:** From the morphologically segmented Turkish corpora, we also extract the sequence of roots for open class content words (nouns, adjectives, adverbs, and verbs) for each sentence. For Turkish, this corresponds to removing all morphemes and any roots for closed classes.

We tag the English side using TreeTagger, which provides a lemma and a part-of-speech tag for each word. We removed all words that are tagged as closed class words, along with tags such as *+VVG* (verb), which signal morpheme from an open class content word. We use this approach to augment the training corpus and bias content root word alignments, so as to obtain better alignments without any additional noise from morphemes and other function words.

- **Out-of-vocabulary word treatment:** Morphological productivity also means more out-of-vocabulary (OOV) words at test time. To decrease the number of OOV words, we also processed these words before tuning and decoding. Similar to the segmentation, we split morpheme by morpheme to get a "known" word from the OOV word. Only when the root word is OOV do we remove the whole word with all its morphemes. There are only a few exceptions such as words that are composed of a bare root word. These words are assumed to be proper nouns and kept in the test data.

3.3. System

We used all official IWSLT Turkish-English data which includes 20k training sentences, and about 500 sentences for development and test. We used IWSLT10 devset1(CSTAR03) (with 16 references on English side) as our development set and IWSLT10 devset2(IWSLT04) (with 16 references on English side) as our internal test set.

For each system, all available parallel corpora distributed for this evaluation were aligned with *GIZA++* for word-to-word alignments with *grow-diag-final-and* and default settings. We built Moses-based systems (with default settings as well) and with morphological preprocessing on the Turkish side. Systems were tuned using the implementation of minimum error rate training (MERT) [4] distributed with Moses.

3.4. English language model

The provided training data for language model are significantly smaller than the usual (less than 400k tokens to be compared with billions). The training data are thus very sparse and we propose to overcome this issue by using continuous space language models as described in [32]. Several different models were experimented and we selected two of them based on their impact on the BLEU score: a standard 7-gram neural network language model [33], and a 4-gram log-bilinear model [34]. We also trained standard n -gram back-off language models (n varying from 3 to 5).

The integration of a continuous space language model in such a system is far from easy, given the computational cost of computing word probabilities, a task that is performed repeatedly during the search of the best translation. We then had to resort to a two pass decoding approach: the first pass uses a conventional back-off language model to produce a 1000-best list; in the second pass, the probability of the neural language model is computed for each hypothesis and the 1000-best list is accordingly reordered to produce the final translations.

3.5. Experiments and Results

In order to evaluate the different pre-processing schemes and the different LMs, Table 3 shows each set of experiments with associated BLEU score measured on our internal test set, where systems marked with + is built on top of the previous one. For these experiments only standard back-off LM were used. First, we can observe that using a 5-gram LM improves the BLEU score despite the small amount of training texts, thus this last model was used for the following experiments. For the frequency based segmentation, we tried several threshold values. Even the BLEU scores for most threshold values are more or less close to each other (e.g. 45.90 for threshold 5 and 45.74 for threshold 15), there is a significant reward in BLEU with the frequency based segmentation with a threshold of 10 (*Segmentation-t10*), whereas the question inversion and local ordering do not seem to help much. Moreover, the specific processing of OOV words *OOV* and to a lesser extent the introduction of open-class words *Content Words* yield an additional BLEU improvement. Based on these results, our system uses the frequency based segmentation, open class words and the pre-processing of OOV words, along with a 5-gram tar-

get LM. The use of neural network language model provided us with an addition gain of 1.5 BLEU score.

System	BLEU
Baseline-3g-lm	37.15
Baseline-4g-lm	37.21
Baseline-5g-lm	38.37
Segmentation-t10	50.06
+Question Inversion	48.85
+Local Ordering	49.74
+Content Words	51.72
+OOV	57.25

Table 3: *Internal test set (BLEU) results for different configurations.*

In the table, even question inversion seems to decrease the BLEU scores, we observed that it helps improving the overall system. We designed a system with all other experiments except the question inversion and got an BLEU score 55.68.

3.6. Final System

For the final system, we used internal test set as an additional data resource. We generated a new 8k parallel data by duplicating each Turkish sentence for each 16 references. The final training data statistics are shown in Table 4.

System	Sentences	Total Words	Unique Words
Turkish	55,729	332,995	7,560
English	55,729	364,653	8,737

Table 4: *Statistics for training data*

The table 5 summarizes our official results with the final system on both the 2009 and 2010 test sets. This system includes the frequency based segmentation, the open-class words, the splitting of OOV words and the use of the 7-gram standard neural-network language model in a two pass decoding approach.

System	case+punc. BLEU	no case+punc. BLEU
iwslt09	52.97	50.75
iwslt10	48.42	46.00

Table 5: *Final system (BLEU) results.*

4. Conclusion

In this paper, we presented our statistical machine translation systems developed for the IWLST'10 evaluation. For the *Talk* task, we studied an extension of our in-house n -code SMT system (the integration of a bilingual reordering model *generalized translation units*), as well as the use of training data extracted from wikipedia in order to adapt the target language model. Experimental results did not show any improvement with these novelties.

For the *BTEC* task, most of our efforts were concentrated in the design of an appropriated pre-processing schemes for Turkish in order to lessen the morphological discrepancies between Turkish and English. Our results showed significant BLEU improvements using a frequency based automatic segmentation and a specific splitting method for OOV words. Moreover, the use of neural-network language models increased the BLEU score of 1.5 points.

5. Acknowledgements

This work was partly realized as part of the Quaero Programme, funded by OSEO, French State agency for innovation. The authors also wish to thank Kemal Oflazer for providing the morphologically analyzed data.

6. References

- [1] A. Allauzen, J. M. Crego, I. Durgar El-Kahlout, and F. Yvon, "LIMSI's statistical translation systems for WMT'10," in *Proc. of the Joint Workshop on Statistical Machine Translation and Metrics-MATR*, Uppsala, Sweden, 2010, pp. 54–59.
- [2] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*, Prague, Czech Republic, 2007.
- [3] J. B. Mariño, R. Banchs, J. M. Crego, A. de Gispert, P. Lambert, J. Fonollosa, and M. R. Costajussà, "N-gram-based machine translation," *Computational Linguistics*, vol. 32, no. 4, 2006.
- [4] F. J. Och, "Minimum error rate training in statistical machine translation," in *ACL '03: Proc. of the 41st Annual Meeting on Association for Computational Linguistics*, 2003, pp. 160–167.
- [5] C. Tillmann, "A unigram orientation model for statistical machine translation," in *HLT-NAACL '04: Proceedings of HLT-NAACL 2004: Short Papers on XX*. Association for Computational Linguistics, 2004, pp. 101–104.
- [6] J. M. Crego and J. B. Mariño, "Improving statistical MT by coupling reordering and decoding," *Machine Translation*, vol. 20, no. 3, pp. 199–215, 2007.
- [7] J. M. Crego and F. Yvon, "Improving reordering with linguistically informed bilingual n-grams," in *Proc. of the 23rd International Conference on Computational Linguistics (Coling 2010: Posters)*, Beijing, China, 2010, pp. 197–205.
- [8] İlknur Durgar El-Kahlout and K. Oflazer, "Exploiting morphology and local word reordering in english to turkish phrase-based statistical machine translation," *To be Appear, IEEE Transactions on Audio, Speech, and Language Processing*, 2010.
- [9] K. Oflazer, "Statistical machine translation into a morphologically complex language," in *In Proc. of the Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, Haifa, Israel, 2008, pp. 376–387.
- [10] K. Oflazer and İlknur Durgar El-Kahlout, "Exploring different representational units in english-to-turkish statistical machine translation," in *In Proc. of Statistical Machine Translation Workshop at the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 2007, pp. 25–32.
- [11] İ. D. El-Kahlout and K. Oflazer, "Initial explorations in English to Turkish statistical machine translation," in *Proc. on the Workshop on Statistical Machine Translation*. New York City: Association for Computational Linguistics, June 2006, pp. 7–14.
- [12] R. Yeniterzi and K. Oflazer, "Syntax-to-morphology mapping in factored phrase-based statistical machine translation from english to turkish," Uppsala, Sweden, pp. 454–464, July 2010.

- [13] S. Köprü, “AppTek turkish-english machine translation system description for IWSLT 2009,” Tokyo, Japan, pp. 19–23, 2009.
- [14] Y. Ma, T. Okita, O. Çetinoğlu, J. Du, and A. Way, “Low-resource machine translation using Ma-TrEx: The DCU machine translation system for IWSLT 2009,” in *Proc. of IWSLT*, Tokyo, Japan, 2009, pp. 29–36.
- [15] N. Bertoldi, A. Bisazza, M. Cettolo, G. Sanchis-Trilles, and M. Federico, “FBK @ IWSLT-2009,” in *Proc. of IWSLT*, Tokyo, Japan, 2009, pp. 37–44.
- [16] Y. Lepage, A. Lardilleux, and J. Gosme, “The GREYC translation memory for the IWSLT 2009 evaluation campaign: one step beyond translation memory,” in *Proc. of IWSLT*, Tokyo, Japan, 2009, pp. 45–49.
- [17] W. Shen, B. Delaney, A. Aminzadeh, T. Anderson, and R. Slyh, “The MIT-LL/AFRL IWSLT-2009 system,” in *Proc. of IWSLT*, Tokyo, Japan, 2009, pp. 29–36.
- [18] C. Mermer, H. Kaya, and M. U. Doğan, “The TÜBITAK-UEKAE SMT system for IWSLT,” in *Proc. of IWSLT*, Tokyo, Japan, 2009, pp. 113–117.
- [19] S. Niessen and H. Ney, “Statistical machine translation with scarce resources using morpho-syntactic information,” *Computational Linguistics*, vol. 30, no. 2, pp. 181–204, 2004.
- [20] S. Corston-oliver and M. Gamon, “Normalizing german and english inflectional morphology to improve statistical word alignment,” in *Proc. of the Conference of the Association for Machine Translation in the Americas*, DC, USA, 2004, pp. 48–57.
- [21] P. Koehn, “Europarl: A parallel corpus for statistical machine translation,” in *MT Summit X*, Phuket, Thailand, 2005, pp. 79–86.
- [22] M. Yang and K. Kirchhoff, “Phrase-based backoff models for machine translation of highly inflected languages,” in *Proc. of 11th Conference of EAACL*, Trento, Italy, 2006, pp. 41–48.
- [23] Y.-S. Lee, “Morphological analysis for statistical machine translation,” in *Proc. HLT-NAACL, Companion Volume*, Boston, USA, 2004, pp. 57–60.
- [24] F. Sadat and N. Habash, “Combination of arabic preprocessing schemes for statistical machine translation,” in *Proc. of the 21st COLING/ACL*, Sydney, Australia, 2006, pp. 1–8.
- [25] M. Popovic and H. Ney, “Towards the use of word stems and suffixes for statistical machine translation,” in *Proc. of 4th LREC*, Lisbon, Portugal, 2004, pp. 1585–1588.
- [26] S. Goldwater and D. McClosky, “Improving statistical MT through morphological analysis,” in *Proc. of HLT/EMNLP*, Vancouver, British Columbia, Canada, 2005, pp. 676–683.
- [27] E. Avramidis and P. Koehn, “Enriching morphologically poor languages for statistical machine translation,” in *In Proc. of ACL-08/HLT*, Columbus, Ohio, 2008, pp. 763–770.
- [28] E. Minkov, K. Toutanova, and H. Suzuki, “Generating complex morphology for machine translation,” in *Proc. of the 45th Annual Meeting of ACL*, Prague, Czech Republic, June 2007, pp. 128–135.
- [29] K. Oflazer, “Two-level description of Turkish morphology,” *Literary and Linguistic Computing*, vol. 9, no. 2, pp. 137–148, 1994.
- [30] H. Sak, T. Güngör, and M. Saraçlar, “Turkish language resources: Morphological parser, morphological disambiguator and web corpus,” in *GoTAL 2008*, ser. LNCS, vol. 5221, 2008, pp. 417–427.
- [31] S. Niessen and H. Ney, “Morpho-syntactic analysis for reordering in statistical machine translation,” in *Proc. of MT Summit VIII*, Santiago de Compostela, Spain, 2001, pp. 247–252.
- [32] H. S. Le, A. Allauzen, G. Wisniewski, and F. Yvon, “Training continuous space language models: Some practical issues,” in *Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, 2010, pp. 778–788.
- [33] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, “A neural probabilistic language model,” *JMLR*, vol. 3, pp. 1137–1155, 2003.
- [34] A. Mnih and G. Hinton, “Three new graphical models for statistical language modeling,” in *Proc. of ICML’07*, 2007, pp. 641–648.