Domain Adaptation for Statistical Machine Translation with Domain Dictionary and Monolingual Corpora

Hua Wu, Haifeng Wang
Toshiba (China) R&D Center
Beijing, 100738, China
wuhua@rdc.toshiba.com.cn
wanghaifeng@rdc.toshiba.com.cn

Chengqing Zong
NLPR, Institute of Automation
Chinese Academy of Sciences
Beijing 100080, China
cqzong@nlpr.ia.ac.cn

Abstract

Statistical machine translation systems are usually trained on large amounts of bilingual text and monolingual text. In this paper, we propose a method to perform domain adaptation for statistical machine translation, where in-domain bilingual corpora do not exist. This method first uses out-of-domain corpora to train a baseline system and then uses in-domain translation dictionaries and in-domain monolingual corpora to improve the in-domain performance. We propose an algorithm to combine these different resources in a unified framework. Experimental results indicate that our method achieves absolute improvements of 8.16 and 3.36 BLEU scores on Chinese to English translation and English to French translation respectively, as compared with the baselines using only out-of-domain corpora.

1 Introduction

In statistical machine translation (SMT), the translation process is modeled to obtain the translation $\mathbf{e}_{\text{best}}$ of the source sentence $\mathbf{f}$ by maximizing the following posterior probability (Brown et al., 1993).

$$\begin{aligned} \mathbf{e}_{\text{best}} &= \arg\max_{\mathbf{e}} p(\mathbf{e} \mid \mathbf{f}) \\ &= \arg\max_{\mathbf{e}} p(\mathbf{f} \mid \mathbf{e}) p_{\text{LM}}(\mathbf{e}) \end{aligned} \qquad (1)$$

State-of-the-art SMT systems are trained on large collections of bilingual corpora for the

translation model $p(\mathbf{f} \mid \mathbf{e})$ and monolingual target language corpora for the language model (LM) $p_{\text{LM}}(\mathbf{e})$. The trained SMT systems are suitable for translating texts in the same domain as the training corpus. However, for some specific domains, it is difficult to obtain a bilingual corpus. In this case, the performance of SMT systems will be degraded.

Generally, it is easier to obtain in-domain monolingual corpora in either source or target language. Moreover, in some specific domains, although in-domain bilingual corpora do not exist, in-domain translation dictionaries, which usually contain domain-specific terms and their translations, are available. And even if such dictionaries are not available, it is easier to compile one than to build a bilingual corpus. Thus, in this paper, we address the problem of domain-specific SMT, where only domain-specific dictionaries and/or monolingual corpora exist. In a specific domain, there are two kinds of words: common words, which also frequently occur in out-of-domain corpora, and domain-specific words, which only occur in the specific domain. Thus, we can combine the out-of-domain bilingual corpus, the in-domain translation dictionary, and monolingual corpora for in-domain translation.

If an in-domain translation dictionary is available, we combine it with the out-of-domain translation model to improve translation quality. If an in-domain target language corpus (TLC) is available, we use it to build an in-domain language model, which can be combined with the out-of-domain language model to further improve translation quality. Moreover, if an in-domain source language corpus (SLC) is available, we automatically translate it and obtain a synthetic in-domain bilingual corpus. By adding this synthetic bilingual corpus to the training data, we rebuild the translation model to improve

translation quality. We can repeatedly translate the in-domain source language corpus with the improved model until no more improvement can be made. This is similar to transductive learning described in (Ueffing et al., 2007).

We perform domain adaptation experiments on two tasks: one is the Chinese to English translation, using the test set released by the International Workshop on Spoken Language Translation 2006 (IWSLT 2006), and the other is the English to French translation, using the data released by the Second Workshop on Statistical Machine Translation (WMT 2007) (Callison-Burch et al., 2007).

Experimental results indicate that our method achieves absolute improvements of 8.16 and 3.36 BLEU scores on Chinese to English translation and English to French translation respectively, as compared with the baselines only using the out-of-domain corpora. The results on both translation tasks also show that the translation quality achieved by our methods is comparable to that of the method using both in-domain and out-of-domain bilingual corpora. Moreover, even if in-domain and out-of-domain bilingual corpora are available, adding an in-domain dictionary also helps to improve the translation quality.

The remainder of the paper is organized as follows. In section 2, we describe the related work. Section 3 briefly introduces the baseline MT system used in our experiments. Section 4 describes our domain adaptation method of using in-domain dictionary and monolingual corpora. And then we present the experimental results in sections 5. In the last section, we conclude this paper.

2   Related Work

Translation model and language model adaptation are usually used in domain adaptation for SMT. Language model adaptation has been widely used in speech recognition (Bacchiani and Roark, 2003). In recent years, language model adaptation has also been studied for SMT (Bulyko et al., 2007). They explored discriminative estimation of language model weights by directly optimizing machine translation performances such as BLEU score (Papineni et al., 2002). Their experiments indicated about 0.4 BLEU score improvement.

A shared task is organized as part of the Second Workshop on Statistical Machine Translation. A part of this shared task focused on domain adaptation for machine translation among European languages. Several studies investigated mixture model adaptation for both translation model and language model in SMT (Civera and Juan, 2007; Foster and Kuhn, 2007). Koehn and Schroeder (2007) investigated different adaptation methods for SMT. Their experiments indicate an absolute improvement of more than 1 BLEU score.

To enlarge the in-domain bilingual corpus, Munteanu and Marcu (2005) automatically extracted in-domain bilingual sentence pairs from comparable corpora. Adding the extracted bilingual corpus to the training data improved the performance of the MT system. In addition, Ueffing et al. (2007) explored transductive learning for SMT, where source language corpora are used to train the models. They repeatedly translated source sentences from the development set and test set. Then the generated translations are used to improve the performance of the SMT system. This kind of transductive learning can be seen as a means to adapt the SMT system to a new type of texts.

In this paper, we use an in-domain translation dictionary and/or in-domain monolingual corpora (in both source language and target language) to improve the performance of a SMT system trained on the out-of-domain corpora. Thus, our method uses these resources, instead of an in-domain bilingual corpus, to adapt a baseline system trained on the out-of-domain corpora to in-domain texts.

3   Baseline MT System

The phrase-based SMT system used in our experiments is Moses (Koehn et al., 2007). In Moses, phrase translation probabilities, reordering probabilities, and language model probabilities are combined in the log-linear model to obtain the best translation $\mathbf{e}_{\text{best}}$ of the source sentence $\mathbf{f}$:

$$
\begin{aligned}
\mathbf{e}_{\text{best}} &= \arg \max_{\mathbf{e}} \; p(\mathbf{e} \mid \mathbf{f}) \\
&\approx \arg \max_{\mathbf{e}} \sum_{m=1}^{M} \lambda_m h_m(\mathbf{e}, \mathbf{f})
\end{aligned} \tag{2}
$$

The weights are set by a discriminative training method using a held-out data set as described in (Och, 2003). The models or features which are employed by the decoder are (a) one or several phrases tables, (b) one or more language models trained with SRILM toolkit (Stolcke, 2002), (c) distance-based and lexicalized distortion models, (d) word penalty, (e) phrase penalty.

Input Out-of-domain training data L_O
In-domain translation dictionary D_I
In-domain target language corpus T_I (optional)
In-domain source language corpus S_I (optional)
Begin Assign translation probabilities to D_I
If T_I is available
Training step: $\pi = \textbf{Estimate}\,(L_O, D_I, T_I)$, where π represents the general model.
Else
Training step: $\pi = \textbf{Estimate}\,(L_O, D_I)$
End if
If S_I is available
$\pi^{(0)} = \pi$, $i = 0$
Repeat
$i = i + 1$
Labeling step: Translate S_I with $\pi^{(i-1)}$ to get a synthetic bilingual corpus L_I
Training step: $\pi^{(i)} = \textbf{Re-estimate}\,(L_O, D_I, L_I)$
Until no more improvement can be achieved
$\pi = \pi^{(i)}$
End if
End
Output Model π for in-domain translation

Figure 1.The domain adaptation algorithm

4 The Framework

4.1 The Algorithm

The detailed information about the algorithm is shown in Figure 1. In our algorithm, a phrase table and a language model are first constructed based on the out-of-domain corpus L_O. Then probabilities are automatically assigned to the entries in the in-domain translation dictionary D_I, from which another phrase table is constructed. At last, the two phrase tables are combined. This is the procedure of the training step $\pi = \textbf{Estimate}\,(L_O, D_I)$.

If an in-domain target language corpus is available, we train an in-domain LM, which is combined with the out-of-domain LM. The built phrase tables and LMs are integrated in the log-linear model as described in section 3. This is the procedure of $\pi = \textbf{Estimate}\,(L_O, D_I, T_I)$.

Moreover, if an in-domain source language corpus is available, we use the built log-linear model to translate the in-domain source texts and obtain a synthetic bilingual corpus. And then we add the synthetic bilingual corpus into the training data to improve the current log-linear model

($\pi^{(i)} = \textbf{Re-estimate}\,(L_O, D_I, L_I)$). With the improved model, we repeatedly translate the in-domain source texts until no more improvement on a development set can be achieved.

4.2 Dictionary Probabilities

In general, there is no translation probability in a manually-made translation dictionary. In order to construct a phrase table, we have to assign probabilities to the entries in the dictionary.

Uniform Translation Probability: Since we have no parallel corpus to estimate the translation probabilities, we simply assign uniform probabilities to the entries in the dictionary. With this method, if a source word has n translations, then we assign $1/n$ to each translation of this phrase for all the four scores of the phrase pair.

Constant Translation Probability: For each entry in the dictionary, we assign a fixed score. In this case, the sum of the translation probability is not necessarily equal to 1.

Corpus Translation Probability: If an in-domain monolingual source corpus exists, we translate it with the method as described in Figure 1, and then estimate the translation probabilities for the entries in the dictionary. And for the

entries whose translation probabilities are not estimated, we assign average probabilities that are calculated from the entries that have obtained probabilities.

4.3 Combining Phrase Tables

In the algorithm, there are two kinds of phrase tables. We need to combine them to translate the in-domain texts.

Mixture Model: The most commonly-used method is linear interpolation.

$$p(\mathbf{e}\,|\,\mathbf{f}) = \lambda p_I(\mathbf{e}\,|\,\mathbf{f}) + (1-\lambda)\,p_o(\mathbf{e}\,|\,\mathbf{f}) \qquad (3)$$

Where $p_I(\mathbf{e}\,|\,\mathbf{f})$ and $p_o(\mathbf{e}\,|\,\mathbf{f})$ are the in-domain and out-of-domain translation probabilities. λ is the interpolation weight.

Discriminative Model: An alternative is to combine the two tables in the log-linear model. During translation with Moses, for each phrase in the sentence, the decoder obtains all of its translations in both phrase tables, and then uses them for translation expansion.

4.4 Combining Language Models

If an in-domain target language corpus exists, we use it to construct an in-domain language model, which is combined with the out-of-domain language model. In this paper, we investigate two combination methods: linear interpolation and log-linear interpolation.

5 Experiments

5.1 Setting

We ran experiments on two different tasks: one is the Chinese to English translation in IWSLT 2006 evaluation, and the other is the English to French domain adaptation translation in the WMT 2007 shared task.

For the Chinese to English translation task, we use the Chinese-English bilingual corpus provided by the Chinese Linguistic Data Consortium (CLDC)[2] as the out-of-domain corpus. It contains 156,840 sentence pairs, with about 3 million English words and about 5 million Chinese characters. In addition, we use the Basic Traveling Expression Corpus (BTEC) released by IWSLT 2006 (Paul, 2006) to construct an in-domain phrase table, as a comparison with that one constructed with the in-domain dictionary.

Corpora	Sentences	OOV
CLDC	156,840	89 (6.31%)
BTEC	39,953	179 (12.69%)
IWSLT06-dev4	489	NA
IWSLT06-test	500	NA

Table 1. Chinese-English corpora

Dictionaries	Entries	OOV
LDC	82,090	228 (16.16%)
in-domain	32,821	121 (8.57%)

Table 2. Chinese-English dictionaries

Corpora	Sentences	OOV
Europarl	949,410	412 (5.90%)
NC	43,060	599 (8.58%)
WMT07 dev	1,057	NA
WMT07 test	2,007	NA

Table 3. English-French corpora

Its source part and target part are separately used as the in-domain monolingual corpora in our experiments. The target parts of both CLDC and BTEC are used for language model construction (see Table 1). From the IWSLT 2006 evaluation, we choose the devset4 as our development data. Evaluation was performed on IWSLT 2006 test set. The references for the test set contain lower-case words and punctuations. The detailed information is shown in Table 1.

We use two kinds of manually-made dictionaries for comparison: one is the LDC Chinese-English Translation Lexicon Version 3.0 (LDC2002L27), and the other is the in-domain spoken language dictionary made by ourselves, which contains in-domain Chinese words and their English translations. The dictionary is manually constructed. Some entries of the dictionary are collected from phrase books. Some of them are collected from the general-domain dictionaries. And then, the entries are filtered and modified by a Chinese native speaker specialized in English. The detailed information is shown in Table 2. If a source word has two translations, it is counted as two entries. The OOV rates of the test set uncovered by the LDC dictionary and the in-domain dictionary are 16.16% and 8.57%, respectively.

For the English to French translation task, the out-of-domain corpus is the Europarl corpus distributed for the shared task of WMT 2007 (Callison-Burch et al., 2007)[3]. We filter the sentence pairs whose lengths are above 40 words. For the

in-domain corpus, we use the News Commentary (NC) corpus distributed in WMT 2007. We also use the same development set and test set in the domain adaptation shared task (see Table 3). We manually built an in-domain English-French dictionary according to the in-domain bilingual corpus, which includes 26,821 entries. It contains in-domain English words and their French translations. The OOV rate of the test set uncovered by this dictionary is 22.34%.

5.2 Evaluation Measures

To perform phrase-based SMT, we use the Moses decoder and its support training scripts. We run the decoder with its default settings and then use Moses' implementation of minimum error rate training (Och, 2003) to tune the feature weights on the development set. Translation quality was evaluated using BLEU score (Papineni et al., 2002).

5.3 Results on Chinese-English Translation

Translation Dictionary

With the out-of-domain bilingual corpus, we train an out-of-domain phrase table. With the in-domain translation dictionary, we construct in-domain phrase tables by assigning different translation probabilities with two different methods: uniform and constant. For the constant translation probability, we set the score using the development set. In our experiments, we set it to 1. We use the target part of the out-of-domain corpus to train a language model[4].

With two phrase tables, we combine them in a linear or log-linear method as described in section 4.3. In our experimental results, log-linear translation models outperform the linear models (16.38 vs. 15.12), where the entries of the dictionary are assigned with the constant translation probabilities. Thus, we will use log-linear models for phrase table combination in the following experiments.

Another method to combine the out-of-domain corpus and the translation dictionary is to add the in-domain dictionary into the training corpus. In this case, only one phrase table is trained.

Table 4 describes the results using the out-of-domain corpus and the in-domain dictionary. The baseline method only uses the out-of-domain corpus to train a phrase table. Then we use both

Methods	Resources Used	BLEU(%)
baseline	out-of-domain corpus	13.59
baseline + dictionary	+dictionary as corpus	15.52
	+uniform prob.	16.00
	+constant prob.	16.38
	+corpus prob.	16.72

Table 4. Translation results of using out-of-domain corpus and in-domain dictionary

the out-of-domain corpus and the in-domain dictionary. The results indicate that adding an in-domain dictionary significantly improves the translation quality by 2.79 BLEU score[5]. The methods using the dictionary as a phrase table outperform the method adding it to the training corpus. And the method using constant translation probabilities significantly outperforms that using the uniform translation probabilities.

For comparison, we also assign corpus probabilities to the entries in the dictionary by translating the source part of the BTEC corpus with the method described in Section 4.2. This also improves the translation quality.

In-Domain Monolingual Corpora

We use the target part of the BTEC corpus to train an in-domain LM. We combine it with the out-of-domain LM in two methods: linear interpolation and log-linear interpolation. The experimental results indicate that linear interpolation outperforms log-linear interpolation (17.16 vs. 16.20). Thus, we will use linear interpolation for LMs in all of the following experiments.

Table 5 describes the results of using the interpolated language model. As compared with the results in Table 4, it can be seen that adding the in-domain language model greatly improves the translation quality. It achieves an absolute improvement of 3.57 BLEU score as compared with the baseline model. If the in-domain translation dictionary is used, the translation quality is further improved by 4 BLEU score.

If the in-domain source language data is available, we translate it and obtain a synthetic bilingual corpus. Then we perform transductive learning as described in Figure 1. The difference between our method and that in (Ueffing et al., 2007) is that we translate a larger in-domain source corpus, and we use 1-best translation

---

[4] We also used LDC English Gigaword to train a large language model. However, this language model did not improve the translation quality.

[5] We use the method described in (Koehn and Monz, 2006) for significance test. In this paper, significant improvement means method A outperforms method B on a significance level of no less than 95%.

Methods	Models	Resources used	BLEU(%)
baseline	Model 1	out-of-domain corpus	13.59
baseline + TLC	Model 2	+ in-domain TLC	17.16
baseline + TLC + dictionary	Model 3	+ in-domain TLC + dictionary (uniform prob.)	20.83
	Model 4	+ in-domain TLC + dictionary (constant prob.)	21.16
transductive learning	Model 5	+ in-domain SLC	15.98
	Model 6	+ in-domain SLC and TLC	18.19
	Model 7	+ in-domain SLC and TLC + dictionary (corpus prob.)	21.75

Table 5. Translation results of using in-domain resources

Dictionary types		Entries	OOV	BLEU(%)
general domain	LDC	82,090	228 (16.16%)	19.11
in-domain	manual	32,821	121 (8.57%)	21.16
	extracted	11,765	330 (23.39%)	19.88
combined	LDC + manual	106,572	45 (3.19%)	21.34
	LDC + extracted	95,660	202 (14.31%)	20.49

Table 6. Comparison of different dictionaries

result with full re-training. The results indicate that transductive learning improves translation quality in all cases. For example, Model 5 achieves an absolute improvement of 2.39 BLEU score over Model 1, and Model 6 achieves 1.03 BLEU score improvement over Model 2. Model 7 uses the in-domain dictionary with corpus translation probabilities, which are obtained from the phrase table trained with the synthetic bilingual corpus. The results indicate that Model 7 outperforms Model 4, with a significant improvement of 0.59 BLEU score.

The results also indicate that when only the in-domain monolingual corpus is used, the translation quality is improved by 4.6 BLEU score (Model 6 vs. Model 1). By adding the in-domain dictionary, the translation quality is further improved, achieving an absolute improvement of 8.16 BLEU score (Model 7 vs. Model 1).

Comparison of Different Dictionaries

We compare the effects of different dictionaries with concern to the translation quality. Besides the manually-made in-domain dictionary, we use other two dictionaries: the LDC dictionary and an automatically built dictionary, which is extracted from the BTEC corpus. This extracted dictionary only contains Chinese words and their translations. The extraction method is as follows:

- Build a phrase table with the in-domain bilingual corpus.
- Filter those phrase pairs whose values are below a threshold as described in (Wu and Wang, 2007).

- From the filtered phrase table, extract the Chinese words and their translations.
- Assign constant translation probabilities to the entries of the extracted dictionary.

Table 6 shows the translation results. All of the methods use the out-of-domain corpus, the in-domain target language corpus, and the corresponding translation dictionaries with constant translation probabilities. The results indicate that using the general-domain dictionary also improves translation quality, achieving an improvement of about 2 BLEU score as compared with Model 2 in Table 5. It can also be seen that the in-domain dictionaries significantly outperform the LDC dictionary although the extracted dictionary has a higher OOV rate than the LDC dictionary. Further analysis shows that the LDC dictionary does not contain the in-domain translations of some words. Results also indicate that combining the two kinds of dictionaries helps to slightly improve translation quality since the OOV rates are reduced.

Comparison with In-domain Bilingual Corpus

The aim of this section is to investigate whether the in-domain dictionary helps to improve translation quality when an in-domain bilingual corpus is available. And we will also compare the translation results with those of the methods only using in-domain dictionaries and monolingual corpora.

To train the in-domain translation model, we use the BTEC corpus. The translation results are
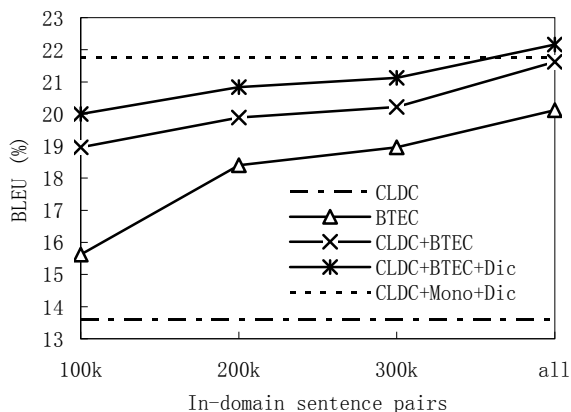
Figure 2. Comparison of different methods using different resources.

shown in Figure 2. CLDC and BTEC represent the methods that only use the out-of-domain and the in-domain corpus, respectively. The method "CLDC+BTEC" uses linear interpolation of the phrase tables and LMs trained with CLDC and BTEC. "Dic" means using the in-domain dictionary, and "Mono" means using in-domain source and target language corpora.

From the results, it can be seen that (a) even if an in-domain bilingual corpus exists, the in-domain dictionary also helps to improve the translation quality, as "CLDC+BTEC+Dic" achieves an improvement of about 1 BLEU score in comparison with "CLDC+BTEC"; (b) the method "CLDC+Mono+Dic", which uses both the in-domain monolingual corpora and the in-domain dictionary, achieves high translation quality. It achieves slightly higher translation quality than "CLDC+BTEC" that uses the in-domain bilingual corpus (21.75 vs. 21.62) and achieves slightly lower translation quality than "CLDC+BTEC+Dic" (21.75 vs. 22.05). But the differences are not significant. This indicates that our method using an in-domain dictionary and in-domain monolingual corpora is effective for domain adaptation.

5.4 Results on English-French Translation

We perform the same experiments for English to French translation. Table 7 describes the domain adaptation results. "Interpolated LM" means that we use the target part of the NC corpus to train an in-domain LM, and then linearly interpolate it with the out-of-domain LM trained with the Europarl corpus. The results indicate that using an in-domain target corpus significantly improves the translation quality, achieving an improvement of 2.19 BLEU score (from 25.44 to 27.63).

Methods	Out-of-domain LM	Interpolated LM
Europarl	25.44	27.63
Europarl+Dic	26.24	28.22
transductive learning	-	28.80
Europarl+NC	-	29.19
Europarl+NC+Dic	-	29.41

Table 7. Translation results of using in-domain dictionary and monolingual corpora

Using the in-domain translation dictionary improves translation quality in all cases, even when the in-domain bilingual corpus is used (from 29.19 to 29.41).

We also perform transductive learning with the source part of the NC corpus. The model used to translate the corpus is that one created by "Europarl+Dic" in Table 7. The results indicate that transductive learning significantly improves translation quality, achieving an absolute improvement of 0.58 BLEU score (from 28.22 to 28.80).

In summary, using an in-domain dictionary and in-domain monolingual corpora improves the translation quality by 3.36 BLEU score (from 25.44 to 28.80). Although the translation quality is slightly lower than that method of using both in-domain and out-of-domain bilingual corpora, the difference is not statistically significant.

6 Conclusion

This paper proposed a domain adaptation approach for statistical machine translation. This approach first used an out-of-domain corpus to build a baseline system, and then used an in-domain translation dictionary and in-domain monolingual corpora to adapt it to the in-domain texts. The contribution of this paper lies in the following points:

- We proposed a method to integrate a domain-specific translation dictionary into a phrase-based SMT system for domain adaptation.

- We investigated the way of using in-domain monolingual corpora in either source or target language, together with the in-domain translation dictionary, to improve the translation quality of a baseline system.

We performed experiments on both Chinese to English and English to French translation. Experimental results on Chinese to English translation indicate that all of the in-domain resources are useful to improve in-domain translation quality, with an overall improvement of 8.16 BLEU score as compared with the baseline trained with out-of-domain corpora. Results on English to French translation also show that using in-domain translation dictionaries and in-domain monolingual corpora is effective for domain adaptation, achieving an absolute improvement of 3.36 BLEU score. And the results on both translation tasks indicate that the translation quality achieved by our methods is comparable with that of the method using both in-domain and out-of-domain bilingual corpora. Moreover, even if in-domain and out-of-domain bilingual corpora are available, adding an in-domain dictionary also helps to improve the translation quality.

In the future work, we will investigate to assign translation probabilities to the dictionaries using comparable in-domain corpora and examine its effect on the MT performance. And we will also examine the effect of an in-domain dictionary on transductive learning in more details.

References

Bacchiani, Michiel and Brian Roark. 2003. Unsupervised Language Model Adaptation. In *Proc. of the 28th International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2003)*, pages 224-227.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2): 263-311.

Bulyko, Ivan, Spyros Matsoukas, Richard Schwartz, Long Nguyen, and John Makhoul. 2007. Language Model Adaptation in Machine Translation from Speech. In *Proc. of the 32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2007)*, pages 117-120.

Callison-Burch, Chris, Cameron Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. 2007. (Meta-) Evaluation of Machine Translation. In *Proc. of the Second Workshop on Statistical Machine Translation*, pages 136-158.

Civera, Jorge and Alfons Juan. 2007. Domain Adaptation in Statistical Machine Translation with Mixture Modelling. In *Proc. of the Second Workshop on Statistical Machine Translation*, pages 177-180.

Foster, George and Roland Kuhn. 2007. Mixture-Model Adaptation for SMT. In *Proc. of the Second Workshop on Statistical Machine Translation*, pages 128-135.

Koehn, Philipp and Christof Monz. 2006. Manual and Automatic Evaluation of Machine Translation between European Languages. In *Proc. of the HLT-NAACL 2006 Workshop on Statistical Machine Translation*, pages 102-121.

Koehn, Philipp, Hieu Hoang, Alexanda Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proc. of the 45th Annual Meeting of the Association for Computational Linguistics (ACL-2007), demonstration session*, pages 177-180.

Koehn, Philipp and Josh Schroeder. 2007. Experiments in Domain Adaptation for Statistical Machine Translation. In *Proc. of the Second Workshop on Statistical Machine Translation*, pages 224-227.

Munteanu, Dragos Stefan, and Daniel Marcu. 2005. Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. *Computational Linguistics*, 31(4): 477-504.

Och, Franz Josef. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-2003)*, pages 160-167.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of the 40th Annual Meeting of the Association of Computational Linguistics (ACL-2002)*, pages 311-318.

Paul, Michael. 2006. Overview of the IWSLT 2006 Evaluation Campaign. In *Proc. of the International Workshop on Spoken Language Translation (IWSLT-2006)*, pages 1-15.

Stolcke, Andrea. 2002. SRILM -- an Extensible Language Modeling Toolkit. In *Proc. of International Conference on Spoken Language Processing (ICSLP-2002)*, pages 901-904.

Ueffing, Nicola, Gholamreza Haffari, and Anoop Sarkar. 2007. Transductive Learning for Statistical Machine Translation. In *Proc. of 45th Annual Meeting of the Association of Computational Linguistics (ACL-2007)*, pages 25-32.

Wu, Hua and Haifeng Wang. 2007. Comparative Study of Word Alignment Heuristics and Phrase-Based SMT. In *Proc. of Machine Translation Summit XI*, pages 507-514.