Translation model adaptation for Statistical Machine Translation with domain classifier

Luan Nghia Pham
Haiphong University
Haiphong, Vietnam
luanpn@dhhp.edu.vn

Vinh Van Nguyen
University of Engineering and Technology
Vietnam National University
Hanoi, Vietnam
vinhnv@vnu.edu.vn

Huy Quang Nguyen
IT Department
Vocational College of Vinhphuc
Vinhphuc, Vietnam
quanghuyinfor@gmail.com

Abstract

In this paper, we propose a new domain adaptation method for SMT systems. Specifically, our method only using monolingual data in a target language to adapting translation model, our system brings improvements over the SMT baseline system. We use two methods to improve the quality of SMT system: (i) classify phrases on the phrase-table of the SMT baseline system, and (ii) adapt to translation model through updating the direct translation probability of phrases.

Our experiments are applied to the English-Vietnamese language pair and use of law domain *(in-domain)* and general domain *(out-domain)* data sets. The English-Vietnamese parallel corpus is provided by the IWSLT 2015 organizers and the experimental results showed that our method significantly outperformed the baseline system. Our system improved on the quality of machine translation in the law domain up to 0.9 BLEU score over baseline trained on the English-Vietnamese data set.

1 Introduction

Statistical Machine Translation (SMT) systems (Koehn et al., 2003) are usually trained on large amounts of bilingual data and monolingual target language data. In general, these corpora may include quite heterogeneous topics and these topics usually define a set of terminological lexicons. Terminologies need to be translated taking into account the semantic context in which they appear.

The Neural Machine Translation (NMT) approach (Wu et al., 2016) has recently been proposed for machine translation. However, the NMT method has some limitations that NMT system is too computationally costly and resource, system NMT also requires much more training time than compared to SMT system (Bentivogli et al., 2016). Therefore, we are still researching on SMT.

Monolingual data are usually available in large amounts, bilingual data are a sparse resource for most language pairs. Collecting sufficiently large high-quality bilingual data is hard, especially on domain-specific data. For this reason, most of the world languages are resource-poor for statistical machine translation, including the English-Vietnamese language pair.

When SMT system is trained on the small amount of domain specific data leading to narrow lexical coverage which again results in a low translation quality. On the other hand, SMT systems are trained, tuned on specific-domain data will perform well on the corresponding domains, but performance deteriorates for out-domain sentences (Haddow and Koehn, 2012). Therefore, SMT systems often suffer from domain adaptation problem during practical applications. When the test data and the training data come from the same domains, SMT systems can achieve good quality. Otherwise, the translation quality degrades dramatically. Therefore, domain adaptation is of significant importance to developing translation systems which can be effectively transferred from one domain to another.

In recent years, domain adaptation problem in SMT becomes more important (Banerjee et al.,

2010) and is an active field of research in SMT with more and more techniques being proposed and put into practice (Foster and Kuhn, 2007); (Banerjee et al., 2010); (Foster et al., 2013); (Dahlmeier et al., 2013); (Chen et al., 2013); (Hasler et al., 2014); (Masumura et al., 2015); (Onrust et al., 2016); (Cuong et al., 2016). The common techniques used to adapt two main components of contemporary state-of-the-art SMT systems: the language model and the translation model. In addition, there are also some proposals for adapting the Neural Machine Translation (NMT) system to a new domain (Freitag and Al-Onaizan, 2016); (Chu et al., 2017). Although the NMT system has begun to be studied more, domain adaptation for the SMT system still plays an important role, especially for machine translation systems in the English-Vietnamese language pair.

In this paper, we propose a new method to adapt translation model of SMT system for the English-Vietnamese language pair, our method uses monolingual data in the target language and based on phrases classification of phrase-table into a specific domain. This is the first adaptive method for the translation model of SMT system in the English-Vietnamese language pair. Experimental results showed that our method significantly outperforms the baseline system. Our system improved the translation quality of machine translation system in the in-domain data *(law domain)* by up to 0.9 BLEU points over baseline.

The paper is organized as follows. In the next section, we present related works on the problem of adaptation in SMT; Section 3 introduces our method; Section 4 describes and discusses the experimental results. Finally, we end with a conclusion and the future works in Section 5.

2 Related works

Domain adaptation for machine translation is known to be a challenging research problem that has substantial real-world application and this has been one of the topics of increasing interest for the recent years. Recent work on machine translation domain adaptation has focused on either the language model component or the translation model component of an SMT system.

Some authors used monolingual in-domain data and adapted the language model. The main advantage of language model adaptation in contrast to translation model adaptation is that monolingual in-domain data is needed.

(Xu et al., 2007) used a source classification document to classify an input document into a domain. This work makes the translation model shared across different domains.

For many language pairs and domains, no new-domain parallel training data is available. (Wu et al., 2008) machine translate new-domain source language monolingual corpora and use the synthetic parallel corpus as additional training data by using handmade dictionaries and monolingual source and target language text.

(Banerjee et al., 2010) build several domain specific translation systems, and trained a classifier to assign each incoming sentence to a domain and use the domain specific system to translate the corresponding sentence. They assume that each sentence in test set belongs to one of the already existing domains.

(Cuong et al., 2016) build MT systems for different domains, it trains, tunes and deploys a single translation system that is capable of producing adapted domain translations and preserving the original generic accuracy at the same time. The approach unifies automatic domain detection and domain model parameterization into one system.

Above related works automatically detected the domain and the classifier works as a “switch” between two independent MT decoding runs.

To adapt a translation model trained from the data in one domain to another, previous works paid more attention to the studies of parallel corpus while ignoring the in-domain monolingual corpora which can be obtained more easily.

Our method have some differences from above methods. Firstly, we use the Maximum Entropy (ME) classification model to estimate the probability of target phrases in-domain. Secondly, our method uses monolingual data in the target language and we study adaptation for the English-Vietnamese language pair.

3 Our method

In the phrase-based SMT, a system is trained on a large parallel corpus of "general" (out-domain) training data, of which some subset (unknown domain) is "in-domain". A phrase in the source language may have many translation hypotheses for translation into the target language with different probability, the hypothesis is selected to translate into target language often has the higher probability than other hypotheses, that selection is defined by formula 1.

Monolingual in-domain data are usually available in large amounts. Thus, two basic adaptation approaches are usually pursued:

- Generating synthetic paralell corpus by the SMT system and using this corpus to adapt its translation and reordering models;
- Using synthetic or provided target texts to adapt its language model.

Most of the related works in section 2 use monolingual data to adapt language model, or to synthesize bilingual data, or to classify source data into specifying domains then training in-domain translation models. The Vietnamese language has resource-poor parallel corpus for SMT, so we propose a new method which only uses monolingual in-domain data to adapt the translation model by classifying phrases in the phrase-table and to update the phrase's direct translation probability.

3.1 Overview of Phrase-Based Statistical Machine Translation

Our baseline is a standard phrase-based SMT system (Koehn et al., 2003). The statistical machine translation approach is based on the noisy-channel model. The best translation for a foreign sentence f is:

$$e = arg\max_{e} p(e)p(e|f) \tag{1}$$

The model consists of two components: a language model assigning a probability p(e) for any target sentence e, and a translation model that assigns a conditional probability p(e|f). The language model is learned using a monolingual corpus in the target language. The parameters of the translation model are estimated from a parallel corpus, i.e. the set of foreign sentences and their corresponding translations into the target language.

MT systems are phrase-based, parallel data is used to derive a phrase-based lexicon (Koehn et al., 2003). The resulting lexicon consists of a list of pairs $(seq_e; seq_f)$ where seq_e is a sequence of one or more foreign words, seq_f is a predicted translation. Each pair comes with an associated score. At decoding time, all phrases from sentence *e* are collected with their corresponding translations observed in training.

These are scored together with the language modeling scores and may include other features. The phrase-based approach by (Koehn et al., 2003) uses a log-linear model (Och and Ney, 2002), and the best correction maximizes the following:

$$e = arg\max_{e} p(e|f) \tag{2}$$

$$= arg\max_{e} \sum_{m=1}^{M} \lambda_m h_m(e, f) \tag{3}$$

where h_m is a feature function, such as language model score and translation scores, and λ_m corresponds to a feature weight.

Figure 1 presents the architecture of Phrase-based Statistical Machine Translation system. There is some translation knowledge that can be used as language models, translation models, etc. The combination of component models (language model, translation model, word sense disambiguation, reordering model...).

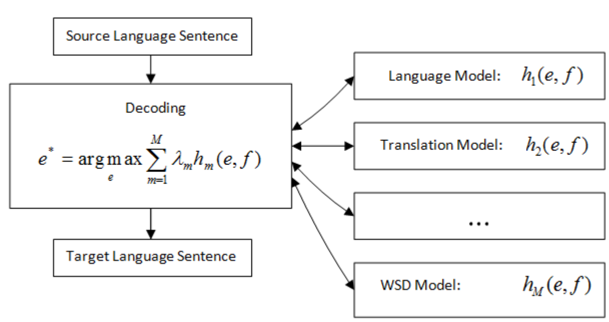

Figure 1: Architecture of Phrase-based Statistical Machine Translation

3.2 Translation model adaptation based on phrase classification

In this section, we describe the domain adaptation techniques that we applied to our experiments for translation model adaptation.

State-of-the-art SMT systems use a log-linear combination of models to decide the best-scoring target sentence given a source sentence. Among these models, the basic ones are a translation model P(e|f) and a target language model P(e).

The translation model is based on phrases; we have a table of the probabilities of translating a specified source phrase f into a specified target phrase e, including phrase translation probabilities in both translation directions. In this paper we conduct the experiments with translation from English to Vietnamese, we consided the direct phrase translation probability ϕ(e|f).

To build a classification model, we use the Stanford Classifier toolkit[1] with standard configurations. This toolkit uses a maximum entropy classifier with character n-grams features,.... The Maximum Entropy classifier is a probabilistic classifier which belongs to the class of exponential models. The Maximum Entropy is based on the principle of Maximum Entropy and from all the models that fit training data, selects the one which has the largest estimate probability. The Maximum Entropy Classifier is used to classify effectively text. Maximum Entropy models can be shown to have the following formula:

$$p(y|x) = \frac{\exp(\sum\limits_{k} \lambda_k f_k(x,y))}{\sum \exp(\sum\limits_{k} \lambda_k f_k(x,z))} \tag{4}$$

where λ_k are model parameters and f_k are features of the model (Berger et al., 1996).

We trained classification model with 2 classes, Law and General classes.

Our method to translation model adaptation can be summarised by the following general algorithm and our system can be described in Figure 2.

1. Build a phrase classification model in the target language on the in-domain data.

[1] https://nlp.stanford.edu/software/classifier.html

2. Train basic component models of SMT, include translation model and language model.

3. Apply classification model for phrases of the phrase-table.

4. Update the phrase's direct translation probability if that phrase is within the specific in-domain class.

Algorithm: Update the phrase's direct translation probability

```
Input:  all phrases of phrase-table
Output:translation model is adapted

for all phrase in phrase-table do
  classify phrases
  if phrase has law lable then
      update the phrase's direct
      translation probability
   end if
end for
return translation model is adapted
```

4 Experiments and Results

In this section, we describe the domain adaptation technique that we applied to our experiments for translation model adaptation.

4.1 Data sets

We conduct our experiments on the data sets of the English-Vietnamese language pair. We consider two very different domains that are law domain and general domain. Detailed statistics for the data sets are given in table 1.

In-domain data: We use monolingual law data in the Vietnamese language collected from documents, dictionaries of the legal domain. This data set consists of 2238 phrases, manually labeled, including 526 in-domain phrases *(in law domain and labeled Law)* and 1712 out-domain phrases *(in general domain and labeled General)*. This data set is used to train the classification model with 2 classes, Law and General class.

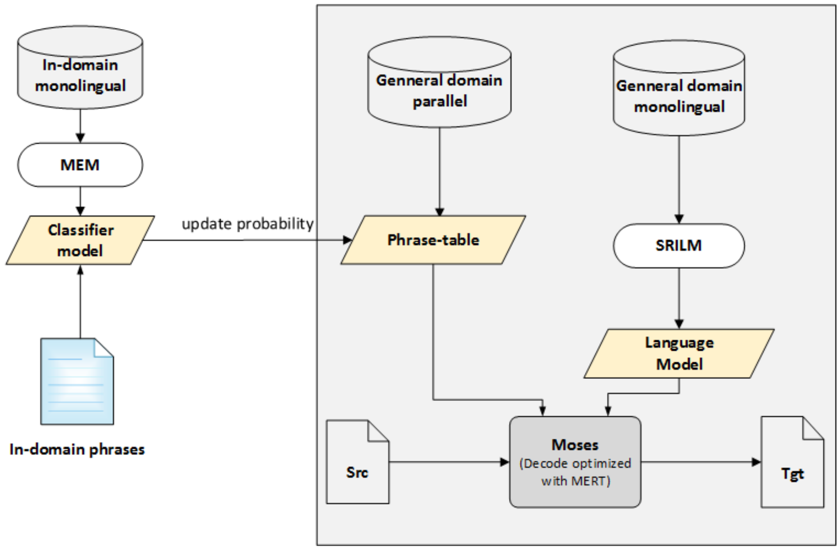

Figure 2: Architecture of the our translation model adaptation system

Additionally, 500 parallel sentences in law domain are used for test set.

Out-domain data: We use all the general parallel training data which are provided by the IWSLT 2015 organisers for the English-Vietnamese translation task.

Additionally, 1000 parallel sentences are used for development set and another 500 parallel sentences are used for test set.

Preprocessing: Data pre-processing plays a very important role in any data-driven method. We carried out preprocessing in two steps:

- ***Cleaning Data:*** We performed cleaning in two phases:
  - ***Phase-1:*** following the cleaning process described in (Pal et al., 2015)
  - ***Phase-2:*** using the corpus cleaning scripts in Moses toolkit (Koehn et al., 2007) with minimum and maximum number of tokens set to 1 and 80 respectively.
- ***Word Segmentation:***
  - *Morphology in Vietnamese:* Vietnamese does not have morphology (Thompson, 1963) and (Aronoff and Fudeman, 2011). In Vietnamese, whitespaces are not used to separate words. The smallest meaningful part of Vietnamese orthography is a syllable (Ngo., 2001). Some examples of Vietnamese words are shown as follows:

    + Single words: "nhà" - house, "nhặt" - pick up, "mua" - buy and "bán" - sell.

    + Compound words: "mua bán" - buy and sell, "bàn ghế" - table and chair, “cây cối”- trees, “đường xá”- street, "hành chính"- administration.

    Thus, a word in Vietnamese may consist of several syllables separted by whitespaces.
  - *Morphology in English:* In English, whitespaces are used to separate words (Aronoff and Fudeman, 2011).

  We used vnTokenizer (Phuong et al., 2008) script to segment for Vietnamese data sets, this is the quite popular toolkit for Vietnamese segmentation and we used tokenizer script in Moses to segment for English data sets.

4.2 Experiments

We performed experiments on the MT1 and MT2 system:

- The MT1 is a SMT baseline system. This system is the phrase-based statistical machine translation with standard settings in the Moses toolkit (Koehn et al., 2007). The MT1 is trained on the Genneral domain *(out-domain)* data set

Data Sets		Language	
		English	**Vietnamese**
Training	Sentences	**122132**	
	Average Length	15.93	15.58
	Words	1946397	1903504
	Vocabulary	40568	28414
Dev	Sentences	**745**	
	Average Length	16.61	15.97
	Words	12397	11921
	Vocabulary	2230	1986
G_test	Sentences	**1046**	
	Average Length	16.25	15.97
	Words	17023	16889
	Vocabulary	2701	2759
L_test	Sentences	**500**	
	Average Length	15.21	15.48
	Words	7605	7740
	Vocabulary	1530	1429

Table 1: The Summary statistical of data sets: English-Vietnamese

and the MT1 is evaluated sequentially on the G_test and L_test data sets.

- The MT2 is the MT1 baseline system after being adapted to the translation model by applying classification model for phrases of the phrase-table. Then update the phrase's direct translation probability if that phrase is within the specific in-domain class and the MT2 is evaluated on the L_test data set.

We use the BLEU score (Papineni et al., 2002) to evaluate translations quality of the MT1 and MT2 systems.

4.2.1 Baseline System

Our MT1 baseline system is a standard phrase-based SMT system based on the Moses SMT toolkit[2](Koehn et al., 2007), this is a state-of-the-art open-source phrase-based SMT system. It uses fourteen features functions in a standard configuration, each system scores translation candidates using standard features: 4 phrase-table features, including inverse phrase translation probability ϕ(f|e), inverse lexical weighting lex(f|e), direct phrase translation probability ϕ(e|f), direct lexical weighting lex(e|f) and phrase translation probabilities; a 7-feature lexicalized re-ordering model; one language model and previously, there was phrase penalty *(always exp(1) = 2.718)*.

The translation and the re-ordering model relied on "grow-diag-final" symmetrized word-to-word alignments built using GIZA++ (Koehn et al., 2003) and the training script of Moses. In our systems, feature weights were optimized using MERT (Och, 2003).

We train a language model with 4-gram and Kneser-Ney smoothing was used in all the experiments. We used SRILM[3] (Stolcke, 2002) as the language model toolkit.

4.2.2 Results

The MT1 baseline system is trained on the general domain data set and the MT1 is evaluated sequentially on G_test and L_test data sets. The MT2

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

[3]http://www.speech.sri.com/projects/srilm/

Source sentences *(Test on Law domain)*	Target sentences		Reference sentences
	MT1 system *(Baseline)*	MT2 system *(Our system)*	
the working party ***took note*** of this commitment .	bữa tiệc làm_việc ***nốt_nhạc*** đã cam_kết này	nhóm làm_việc ***ghi_nhận*** cam_kết này.	ban công tác đã ***ghi_nhận*** cam_kết này .
according to the ***general statistical office*** , services had accounted for 37.98 percent of vietnam 's gdp in 2004	theo ***tổng_quát văn_phòng thống_kê*** , dịch_vụ đã chiếm 37.98% của việt_nam là gdp vào năm 2004 .	theo ***tổng_cục thống_kê*** , dịch_vụ đã chiếm 37.98% của việt_nam là gdp vào năm 2004 .	theo ***tổng_cục thống_kê*** , dịch_vụ chiếm 37,98% gdp năm 2004 của việt_nam .
renewable certificates valid for five years were granted by the construction departments of cities and provinces .	giấy chứng_nhận ***tái_tạo*** có giá_trị trong 5 năm qua là cấu_trúc của các thành_phố và đại_lục .	giấy_phép ***gia_hạn*** có giá_trị trong 5 năm bởi cấu_trúc của các thành_phố và đại_lục .	sở xây dựng các tỉnh và thành_phố cấp giấy_phép hành_nghề có hiệu_lực 5 năm và các giấy_phép này có_thể được ***gia_hạn*** .
the economic police received specialized training on intellectual property ***enforcement*** .	cảnh_sát kinh_tế được đào_tạo chuyên về ***cơ_quan*** sở_hữu_trí_tuệ .	cảnh_sát kinh_tế được đào_tạo chuyên về ***thực_thi*** quyền_sở_hữu_trí_tuệ .	cảnh_sát kinh_tế được đào_tạo chuyên_sâu về ***thực_thi*** quyền_sở_hữu trí_tuệ .
administrative measures only applied to ***acts*** of low ***gravity*** .	***đo hành_chính*** chỉ áp_dụng cho ***hành_động*** của ***trọng_lực*** thấp .	***các biện_pháp hành_chính*** chỉ áp dụng cho ***các hành_vi nghiêm_trọng*** thấp.	***các biện_pháp hành_chính*** chỉ áp_dụng với ***những hành_vi*** có tính ***nghiêm_trọng*** thấp .
evidence collected during an ***administrative procedure*** could be used by the civil court if necessary in accordance with ***civil procedure code*** of 2004 .	bằng_chứng thu_thập được trong một ***ca hành_chính*** có_thể được sử_dụng bởi những toà dân_sự nếu cần_thiết theo ***thủ_tục dân_sự mã*** của năm 2004 .	bằng_chứng thu_thập được trong ***thủ_tục hành_chính*** có_thể được sử_dụng bởi toà dân_sự nếu cần_thiết theo ***bộ_luật thủ_tục dân_sự*** của năm 2004	chứng_cứ thu được trong quá_trình ***xử_lý hành_chính*** sẽ được sử_dụng tại tòa dân_sự nếu thấy cần_thiết theo ***bộ_luật tố_tụng dân_sự*** năm 2004 .

Table 2: Some examples in our experiments

system *(after being adapted)* is evaluated on the data set L_test. Experimental results of MT1 and MT2 systems show that our system has better result than baseline, results are shown in table 3 and some examples in our experiments are shown in table 2.

The phrase ***"took note"*** in context *"the working party* ***took note*** *of this commitment" (source sentence in the table 2)* should been translated into ***"ghi_nhận"*** as reference sentence. The MT1 baseline system translated the phrase ***"took note"*** into ***"nốt_nhạc"*** and our MT2 system translated the phrase ***"took note"*** into ***"ghi_nhận"*** as reference sentence. Some other phrases in the source sentence of table 2 show translation quality on our MT2 system is better than on the MT1 baseline system.

SYSTEM	BLEU SCORE
MT1 *(with G_test)*	**31.3**
MT1 *(with L_test)*	**28.8**
MT2 *(with L_test)*	**29.7**

Table 3: The experiment results of MT1 and MT2 systems

Table 3 shows that the MT1 baseline system is trained on the general domain data set, if the test data *(here is the G_test data set)* is in the same domain as the training data, the BLEU score will be 31.3. If the test data is on the law domain *(here is the L_test data set)*, the BLEU score will be 28.8. And if the MT2 system *(the MT1 baseline system after being adapted to the translation model)* is experimented on the law domain *(here is the L_test data set)*, the BLEU score will be 29.7.

The experiment results in table 3 showed that the SMT system is trained on the general domain if the test domain is different with the training domain, the quality of translation will be down. In this experiments, the BLEU score was reduced 2.5 points from 31.3 to 28.8. The MT1 system is adapted by our technique will improve the quality of the translation system though we are still experimenting with the L_test data set in law domain. In this experiments, the BLEU score is improved to 0.9 points from 28.8 up to 29.7.

5 Conclusions and Future Works

In this paper, we presented a new method for domain adaptation in SMT. Our method only uses monolingual in-domain data to adapt the translation model by classifying phrases in phrase-table and to update the phrase's direct translation probability. Our system obtained a improved on the quality of machine translation in the law domain up to 0.9 BLEU points over baseline. Experimental results show that our method is effective in improving the translation accuracy.

In future work, we intend to study the problem in the other multi-domain and integrate automatically our technique to decode of SMT system.

Acknowledgement

This work described in this paper has been partially funded by Hanoi National University (QG.15.23 project)

References

Mark Aronoff and Kirsten Fudeman. 2011. What is morphology, vol. 8. john wiley and sons.

Pratyush Banerjee, Jinhua Du, Baoli Li, Sudip Kr. Naskar, Andy Way, and Josef van Genabith. 2010. Combining multi-domain statistical machine translation models using automatic classifiers. In Proceedings of AMTA 2010.

Luisa Bentivogli, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. 2016. Neural versus phrase-based machine translation quality: a case study.

Adam Berger, Stephen Della Pietra, and Vincent Della Pietra. 1996. A maximum entropy approach to natural language processing. Computational Linguistics, 22.

Boxing Chen, Roland Kuhn, and George Foster. 2013. Vector space model for adaptation in statistical machine translation. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1285–1293.

Chenhui Chu, Raj Dabre, and Sadao Kurohashi. 2017. An empirical comparison of simple domain adaptation methods for neural machine translation.

Hoang Cuong, Khalil Sima'an, and Ivan Titov. 2016. Adapting to all domains at once: Rewarding domain invariance in smt. Proceedings of the Transactions of the Association for Computational Linguistics (TACL).

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu4. 2013. Building a large annotated corpus of learner english: The nus corpus of learner english. In Proceedings of the NAACL Workshop on Innovative Use of NLP for Building Educational Appli- cations.

George Foster and Roland Kuhn. 2007. Mixture-model adaptation for smt. Proceedings of the Second Workshop on Statistical Machine Translation, pages 128–135, Prague. Association for Computational Linguistics.

George Foster, Boxing Chen, and Roland Kuhn. 2013. Simulating discriminative training for linear mixture adaptation in statistical machine translation. Proceedings of the MT Summit.

Markus Freitag and Yaser Al-Onaizan. 2016. Fast domain adaptation for neural machine translation.

Barry Haddow and Philipp Koehn. 2012. Analysing the effect of out-of-domain data on smt systems. In Proceedings of the Seventh Workshop on Statistical Machine Translation, pages 422–432.

Eva Hasler, Phil Blunsom, Philipp Koehn, and Barry Haddow. 2014. Dynamic topic adaptation for phrase-based mt. In Proceedings of the 14th Conference of the European Chapter of The Association for Computational Linguistics, pages 328–337.

Philipp Koehn, Franz Josef Och, , and Daniel Marcu. 2003. Statistical phrase-based translation. In Proceedings of HLT-NAACL 2003, pages 127–133. Edmonton, Canada.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. In ACL-2007: Proceedings of demo and poster sessions, Prague, Czech Republic, pp.177-180.

Ryo Masumura, Taichi Asam, Takanobu Oba, Hirokazu Masataki, Sumitaka Sakauchi, and Akinori Ito. 2015. Hierarchical latent words language models for robust modeling to out-of domain tasks. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1896–1901.

Binh N. Ngo. 2001. The vietnamese language learning framework. Journal of Southeast Asian Language Teaching 10. Pages 1–24.

Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In Proceedings of ACL 2003, pages 160-167.

Louis Onrust, Antal van den Bosch, and Hugo Van hamme. 2016. Improving cross-domain n-gram language modelling with skipgrams. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 137–142, Berlin, Germany.

Santanu Pal, Sudip Naskar, and Josef van Genabith. 2015. Uds-sant: English–german hybrid machine translation system. In Proceedings of the Tenth Workshop on Statistical Machine Translation, pages 152–157, Lisbon, Portugal, September. Association for Computational Linguistics.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. ACL.

Le Hong Phuong, Nguyen Thi Minh Huyen, Azim Roussanaly, and Ho Tuong Vinh. 2008. A hybrid approach to word segmentation of vietnamese texts.

Andreas Stolcke. 2002. “srilm - an extensible language modeling toolkit,” in proceedings of international conference on spoken lan-guage processing.

Laurence C. Thompson. 1963. The problem of the word in vietnamese. Word journal of the International Linguistic Association, 19(1):39–52.

Hua Wu, Haifeng Wang, and Chengqing Zong. 2008. Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, pp.993–1000.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.

Jia Xu, Yonggang Deng, Yuqing Gao, and Hermann Ney. 2007. Domain dependent statistical machine translation. In Proceedings of the MT Summit XI, pages 515–520.