Adapting Neural Machine Translation for English-Vietnamese using Google Translate system for Back-translation

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

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

Abstract

Monolingual data have been demonstrated to be helpful in improving translation quality of both statistical machine translation (SMT) systems and neural machine translation (NMT) systems, especially in resource-poor language or domain adaptation tasks where parallel data are not rich enough. Google Translate is a well-known machine translation system. It has implemented the Google Neural Machine Translation (GNMT) over many language pairs and English-Vietnamese language pair is one of them.

In this paper, we propose a method to better leveraging monolingual data by exploiting the advantages of GNMT system. Our method for adapting a general neural machine translation system to a specific domain, by exploiting Back-translation technique using target-side monolingual data. This solution requires no changes to the model architecture from a standard NMT system. Experiment results show that our method can improve translation quality, results significantly outperforming strong baseline systems, our method improves translation quality in legal domain up to 13.65 BLEU points over the baseline system for English-Vietnamese pair language.

1 Introduction

Machine translation relies on the statistics of a large parallel corpus, datasets of paired sentences in both sides the source and target language. Monolingual data has been traditionally used to train language models which improved the fluency of statistical machine translation (Koehn2010). Neural machine translation (NMT) systems require a very large amount of training data to make generalizations, both on the source side and on the target side. This data typically comes in the form of a parallel corpus, in which each sentence in the source language is matched to a translation in the target language. Unlike parallel corpus, monolingual data are usually much easier to collect and more diverse and have been attractive resources for improving machine translation models since the 1990s when data-driven machine translation systems were first built. Adding monolingual data to NMT is important because sufficient parallel data is unavailable for all but a few popular language pairs and domains.

From the machine translation perspective, there are two main problems when translating English to Vietnamese: First, the own characteristics of an analytic language like Vietnamese make the translation harder. Second, the lack of Vietnamese-related resources as well as good linguistic processing tools for Vietnamese also affects to the translation quality. In the linguistic aspect, we might consider Vietnamese is a source-poor language, especially parallel corpus in many specific domains, for example, mechanical domain, legal domain, medical domain, etc.

Google Translate is a well-known machine translation system. It has implemented the Google Neural Machine Translation (GNMT) over many language pairs and English-Vietnamese language pair is one of them. The translation quality is good for the general domain of this language pair. So we want to leverage advantages of GNMT system *(resources, techniques,...)* to build a domain translation system for this language pair, then we can improve the quality of translation by integrating more features of Vietnamese.

Language is very complicated and ambiguous. Many words have several meanings that change according to the context of the sentence. The accuracy of the machine translation depends on the topic thats being translated. If the content translated includes a lot of technical or specialized things, its unlikely that Google Translate will work. If the text includes jargon, slang and colloquial words this can be almost impossible for Google Translate to identify. If the tool is not trained to understand these linguistic irregularities, the translation will come out literal and (most likely) incorrect.

This paper presents a new method to adapt the general neural machine translation system to a different domain. Our experiments were conducted for the English-Vietnamese language pair in the direction from English to Vietnamese. We use domain-specific corpora comprising of two specific domains: legal domain and general domain. The data has been collected from documents, dictionaries and the IWSLT2015 workshop for the English-Vietnamese translation task.

This paper is structured as follows. Section 2 summarizes the related works. Our method is described in Section 3. Section 4 presents the experiments and results. Analysis and discussions are presented in Section 5. Finally, conclusions and future works are presented in Section 6.

2 Related works

In statistical machine translation, the synthetic parallel corpus has been primarily proposed as a means to exploit monolingual data. By applying a self-training scheme, the pseudo parallel corpus was obtained by automatically translating the source-side monolingual data (Nicola Ueffing2007; Hua Wu and Zong2008). In a similar but reverse way, the target-side monolingual data were also employed to build the synthetic parallel corpus (Bertoldi and Federico2009; Patrik Lambert2011). The primary goal of these works was to adapt trained SMT models to other domains using relatively abundant in-domain monolingual data.

In (Bojar and Tamchyna2011a), synthetic parallel corpus by Back-translation has been applied successfully in phrase-based SMT. The method in this paper used back-translated data to optimize the translation model of a phrase-based SMT system and show improvements in the overall translation quality for 8 language pairs.

Recently, more research has been focusing on the use of monolingual data for NMT. Previous work combines NMT models with separately trained language models (Gülçehre et al.2015). In (Sennrich et al.2015), authors showed that target-side monolingual data can greatly enhance the decoder model. They do not propose any changes in the network architecture, but rather pair monolingual data with automatic Back-translations and treat it as additional training data. Contrary to this, (Zhang and Zong2016) exploit source-side monolingual data by employing the neural network to generate the synthetic large-scale parallel corpus and multi-task learning to predict the translation and the reordered source-side monolingual sentences simultaneously.

Similarly, recent studies have shown different approaches to exploiting monolingual data to improve NMT. In (Caglar Gulcehre and Bengio2015), authors presented two approaches to integrating a language model trained on monolingual data into the decoder of an NMT system. Similarly, (Domhan and Hieber2017) focus on improving the decoder with monolingual data. While these studies show improved overall translation quality, they require changing the underlying neural network architecture. In contrast, Back-translation allows one to generate a parallel corpus that, consecutively, can be used for training in a standard NMT implementation as presented by (Rico Sennrich and Birch016a), authors used 4.4M sentence pairs of authentic human-translated parallel data to train a baseline English to German NMT system that is later used to translate 3.6M German and 4.2M English target-side sentences. These are then mixed with the initial data to create human + synthetic parallel corpus which is then used to train new models.

In (Alina Karakanta and van Genabith2018), authors use back-translation data to improve MT for a resource-poor language, namely Belarusian (BE). They transliterate a resource-rich language (Russian, RU) into their resource-poor language (BE) and train a BE to EN system, which is then used to translate monolingual BE data into EN. Finally, an EN to BE system is trained with that back-translation data.

Our method has some differences from the above methods. As described in the above, synthetic parallel data have been widely used to boost the performance of NMT. In this work, we further extend their application by training NMT with synthetic parallel data by using Google Translate system. Moreover, our method investigating Back-translation in Neural Machine Translation for the English-Vietnamese language pair in the legal domain.

3 Our method

In Machine Translation, translation quality depends on training data. Generally, machine translation systems are usually trained on a very large amount of parallel corpus. Currently, a high-quality parallel corpus is only available for a few popular language pairs. Furthermore, for each language pair, the size of specific domains corpora and the number of domains available are limited. The English-Vietnamese is resource-poor language pair thus parallel corpus of many domains in this pair is not available or only a small amount of this data. However, monolingual data for these domains are always available, so we want to leverage a very large amount of this helpful monolingual data for our domain adaptation task in neural machine translation for English-Vietnamese pair.

The main idea in this paper, that is leveraging domain monolingual data in the target language for domain adaptation task by using Back-translation technique and Google Translate system. In this section, we present an overview of the NMT system which is used in our experiments and the next we describe our main idea in detail.

3.1 Neural Machine Translation

Given a source sentence $x = (x_1, ..., x_m)$ and its corresponding target sentence $y = (y_1, ..., y_n)$, the NMT aims to model the conditional probability $p(y|x)$ with a single large neural network. To parameterize the conditional distribution, recent studies on NMT employ the encoder-decoder architecture (Kalchbrenner and Blunsom2013; Kyunghyun Cho and Bengio014b; Ilya Sutskever and Le2014). Thereafter, the attention mechanism (Dzmitry Bahdanau and Bengio2014; Minh-Thang Luong and Manning2015b) has been introduced and successfully addressed the quality degradation of NMT when dealing with long input sentences (Kyunghyun Cho and Bengio014a).

In this study, we use the attentional NMT architecture proposed by (Dzmitry Bahdanau and Bengio2014). In their work, the encoder, which is a bidirectional recurrent neural network, reads the source sentence and generates a sequence of source representations $h = (h_1, ..., h_m)$. The decoder, which is another recurrent neural network, produces the target sentence one symbol at a time. The log conditional probability thus can be decomposed as follows:

$$\log p(y|x) = \sum_{i=1}^{n} \log p(y_t|y_{<t}, x) \quad (1)$$

where $y_{<t} = (y_1, ..., y_{t-1})$. As described in Equation 2, the conditional distribution of $p(y_t|y_{<t}, x)$ is modeled as a function of the previously predicted output y_{t-1}, the hidden state of the decoder s_t, and the context vector c_t.

$$p(y_t|y_{<t}, x) \propto \exp\{g(y_{t-1}, s_t, c_t)\} \quad (2)$$

The context vector c_t is used to determine the relevant part of the source sentence to predict y_t. It is computed as the weighted sum of source representations $h_1, ..., h_m$. Each weight α_{ti} for h_i implies the probability of the target symbol y_t being aligned to the source symbol x_i:

$$c_t = \sum_{i=1}^{m} \alpha_{ti} h_i \quad (3)$$

Given a sentence-aligned parallel corpus of size N, the entire parameter θ of the NMT model is jointly trained to maximize the conditional probabilities of all sentence pairs $\{(x^n, y^n)\}_{n=1}^{N}$:

$$\theta^* = arg\max_{\theta} \sum_{n=1}^{N} \log p(y^n|x^n) \quad (4)$$

where θ^* is the optimal parameter.

3.2 Back-translation using Google's Neural Machine Translation

In recent years, machine translation has grown in sophistication and accessibility beyond what we imaged. Currently, there are a number of online translation services ranging in ability, such as Google Translate[1], Bing Microsoft Translator[2], Babylon Translator[3], Facebook Machine Translation, etc. The Google Translate service is one of the most used machine services because of its convenience.

The Google Translate is launched in 2006 as a statistical machine translation, Google Translate has improved dramatically since its creation. Most significantly in 2017, Google moved away from Phrase-Based Machine Translation and was replaced by Neural Machine Translation (GNMT) (Johnson et al.2017). According to Googles own tests, the accuracy of the translation depends on the languages translated. Many languages have even low accurate because of their complexity and differences.

The Back-translation techniques, the first trains an intermediate system on the parallel data which is used to translate the target monolingual data into the source language. The result is a parallel corpus where the source side is synthetic machine translation output while the target is text written by humans. The synthetic parallel corpus is then simply added to the parallel corpus available to train a final system that will translate from the source to the target language. Although simple, this method has been shown to be helpful for phrase-based translation (Bojar and Tamchyna2011b), NMT (Rico Sennrich and Birch2016) as well as unsupervised MT (Guillaume Lample and Ranzato2018). Although here we focus on adapting English to Vietnamese and investigate, experiment on legal domain data. However, this method can be also applied to many other different domains for this language pair.

To take advantages of the Google Translate and helpfulness of domain monolingual data, we use the back-translation technique combine with the Google Translate to synthesize parallel corpus for training our translation system. Our method is described in detail in Figure 1.

[1] https://translate.google.com

[2] https://www.bing.com/translator

[3] https://translation.babylon-software.com/

In Figure 1, our method includes 3 stages, with details as follows:

- **Stage 1:** In this stage, we use Google Translate to translate domain monolingual data in Vietnamese *(target language side)*. The output of this stage is a translation in English *(source language side)*. This technique is called Back-translation. In this case, using the high-quality model to back-translate domain-specific monolingual target data, and then building a new model with this synthetic training data, might be useful for domain adaptation.

- **Stage 2:** In this stage, at first we synthesize parallel corpus by combine input domain monolingual data with output translation in stage 1, because input monolingual data in the legal domain, therefore we consider this synthetic parallel corpus is also in the legal domain. Next, we mix synthetic parallel corpus with an original parallel corpus which is provided by the IWSLT2015[4] workshop *(this corpus in general domain)*, this is the most interesting scenario which allows us to trace the changes in quality with increases in synthetic-to-original parallel data ratio.

- **Stage 3:** With the parallel corpus mixed in stage 2, we conduct training NMT systems from English to Vietnamese and evaluate translation quality in the legal domain and general domain.

4 Experiments setup

In this section, we describe the data sets used in our experiments, data preprocessing, the training and evaluation in detail.

4.1 Datasets and Preprocessing

Datasets We experiment on the data sets of the English-Vietnamese language pair. All experiments, we consider two different domains that are legal domain and general domain. The summary of the parallel and monolingual data is presented in Table 1.

[4] http://workshop2015.iwslt.org/

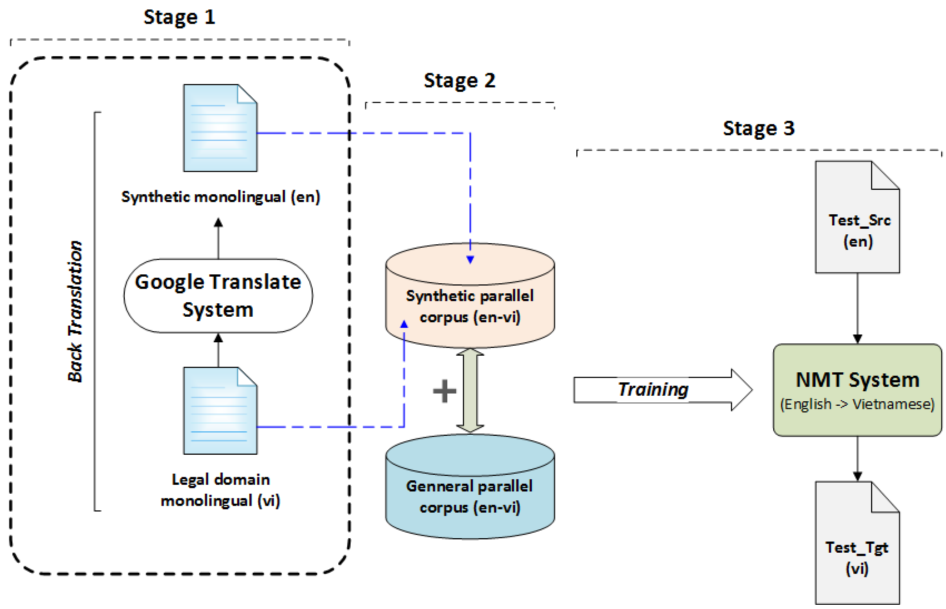

Figure 1: An illustration for our method, includes 3 stages: 1) Back-translation legal domain monolingual text by using Google Translate system; 2) synthesize parallel data from synthetic monolingual and legal domain monolingual in stage 1, and 3) combine synthetic parallel corpus with general parallel corpus for training NMT system

- For training baseline systems, we use the English-Vietnamese parallel corpus which is provided by IWSLT2015 *(133k sentence pairs)*, this corpus was used as general domain training data and tst2012/tst2013 data sets were selected as validation *(val)* and test data respectively.
- For creating the source side data *(English)*, we use 100k sentences in legal domain in target side *(Vietnamese)*.
- To evaluation, we use 500 sentence pairs in legal domain and 1,246 sentence pairs in general domain *(tst2013 data set)*.

Preprocessing Each training corpus is tokenized using the tokenization script in Moses (Koehn et al.2007) for English. For cleaning, we only applied the script *clean-n-corpus.perl* in Moses to remove lines in the parallel data containing more than 80 tokens.

In Vietnamese, a word boundary is not white space. White spaces are used to separate syllables in Vietnamese, not words. A Vietnamese word consist of one or more syllables. We use vnTokenizer (Phuong et al.2013) for word segmentation. However, we only used for separation marks such as dots, commas and other special symbols.

4.2 Settings

We have trained a Neural Machine Translation system by using the OpenNMT[5] toolkit (Klein et al.2018) with the seq2seq architecture of (Sutskever et al.2014), this is a state-of-the-art open-source neural machine translation system, started in December 2016 by the Harvard NLP group and SYSTRAN. This architecture is formed by an encoder, which converts the source sentence into a sequence of numerical vectors, and a decoder, which predicts the target sentence based on the encoded source sentence. In our NMT models is trained with the default model, which consists of a 2-layer Long Short-Term Memory (LSTM) network (Luong et al.2015) with 500 hidden units on both the encoder/decoder and the general attention type of (Minh-Thang Luong and Manning2015a).

For translation evaluation, we use standard BLEU score metric (Bi-Lingual Evaluation Understudy) (Kishore Papineni and Zhu2002) that is currently one of the most popular methods of automatic ma-

[5] http://opennmt.net/

Data Sets		Language	
		English	Vietnamese
Training	Sentences	**133316**	
	Average Length	16.62	16.68
	Words	1952307	1918524
	Vocabulary	40568	28414
Val	Sentences	**1553**	
	Average Length	16.21	16.97
	Words	13263	12963
	Vocabulary	2230	1986
General_test	Sentences	**1246**	
	Average Length	16.15	15.96
	Words	18013	16989
	Vocabulary	2708	2769
Legal_test	Sentences	**500**	
	Average Length	15.21	15.48
	Words	7605	7740
	Vocabulary	1530	1429

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

chine translation evaluation. The translated output of the test set is compared with different manually translated references of the same set.

4.3 Experiments and Results

In our experiments, we train NMT models with parallel corpus composed of: (1) synthetic data only; (2) IWSLT 2015 parallel corpus only; and (3) a mixture of parallel corpus and synthetic data. We trained 5 NMT systems and evaluated the quality of translation on the general domain data and the legal domain data. We also compare the translation quality of our systems with Google Translate, Our systems are described as follows:

- ***The system are built using IWSLT2015 data only:*** This baseline system is trained on general domain data which is provided by IWSLT2015 workshop. Training data *(133k sentences pairs)* and tst2012 data set were selected as validation *(val)*, we call this system is ***Baseline***.
- ***The system are built using synthetic data only:*** Such systems represent the case where no parallel data is available but monolingual data can be translated via an existing MT system and provided as a training corpus to a new NMT system. This case we use 100k sentences in Vietnamese in the legal domain and use Google Translate system for Back-translation. The synthetic parallel data is used for training NMT system and tst2012 data set were selected as validation *(val)*, this system is called ***Synthetic***.
- ***The system are built using mixture of parallel corpus and synthetic data:*** This is the most interesting scenario which allows us to trace the changes in quality with increases in synthetic-to-orginal data ratio. we train 2 NMT systems, the first system is trained on IWSLT2015 data *(133k sentences pairs)* + Synthetic *(50k sentences pairs)* and second system is trained on IWSLT2015 *(133k sentences pairs)* + Synthetic *(100k sentences pairs)*, and tst2012 data set were selected as validation *(val)*, these systems is called ***Baseline_Syn50*** and ***Baseline_Syn100*** respectively.

Our NMT systems are evaluated in the general domain and legal domain. We also compare translation quality with Google Translate on the same test domain data set. Experiment results are shown by the bleu score as table 2 and table 3.

As the results in table 2 and table 3, the Baseline NMT system achieved 25.43 BLEU score in general domain but reduced to 19.23 in the legal domain. After applying Back-translation, the results are im-

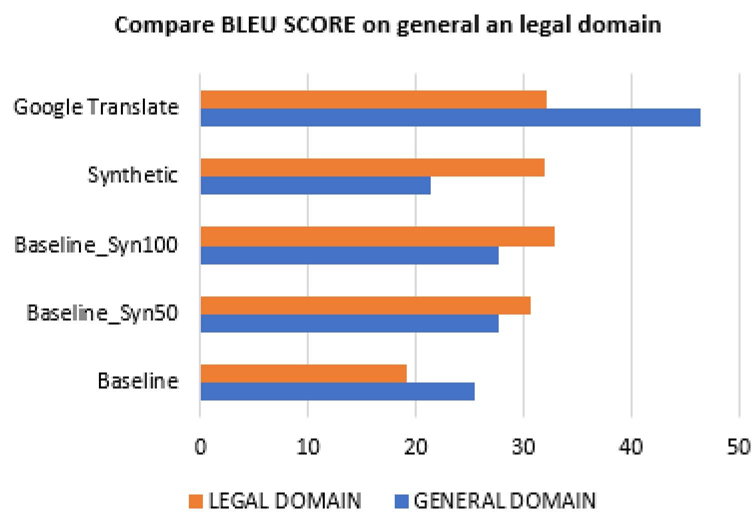

Figure 2: Comparison of translation quality when translating in the legal domain and general domain.

SYSTEM	BLEU SCORE
Baseline	**25.43**
Baseline_Syn50	**27.74**
Baseline_Syn100	**27.68**
Synthetic	**21.42**
Google Translate	**46.47**

Table 2: The experiment results of our systems in the general domain

SYSTEM	BLEU SCORE
Baseline	**19.23**
Baseline_Syn50	**30.61**
Baseline_Syn100	**32.88**
Synthetic	**31.98**
Google Translate	**32.05**

Table 3: The experiment results of our systems in the legal domain

proved, significantly outperforming strong baseline systems, our method improves translation quality in legal domain up to 13.65 BLEU points over baseline system and 2.25 BLEU points over baseline system in general domain.

In Figure 2 is shown the comparison of translation quality when translating in the legal domain and general domain. In general domain, Google Translate's bleu score is 46.47 points, the baseline system is 25.43 points and bleu score of our systems are higher than the baseline system, reaching 27.68; 27.74 points respectively. In the legal domain, Google Translate's bleu score is 32.05 points, the baseline system is 19.23 points and bleu score of our systems are higher than the baseline system, reaching 31.98, 32.61 and 32.88 points respectively. Thus, Back-translation uses Google Translate for English - Vietnamese language pair in the legal domain can improve the translation quality of the English - Vietnamese translation system.

5 Analysis and discussions

The Back-translation technique enables the use of synthetic parallel data, obtained by automatically translating cheap and in many cases available information in the target language into the source language. The synthetic parallel data generated in this way is combined with parallel texts and used to improve the quality of NMT systems. This method is simple and it has been also shown to be helpful for machine translation.

We have experimented with different synthetic data rates and observed effects on translation results. However, we have not investigated to answer issues for adapting the legal domain in NMT of English-Vietnamese language pair such as:

- Does back-translation direction matter?
- How much monolingual back-translation data is necessary to see a significant impact in MT quality?

- Which sentences are worth back translating and which can be skipped?

Overall, we are becoming smarter in selecting incremental synthetic data in NMT that helps improve both: performance of the systems and translation accuracy.

6 Conlustion

In this work, we presented a simple but effective method to adapt general neural machine translation systems into the legal domain for English-Vietnamese language pairs. We empirically showed that the quality of the NMT system is selected for Back-translation for synthetic parallel corpus generation very significant *(here we selected Google Translate for leverage advantages of this translation system)*, and neural machine translation performance can be improved by iterative back-translation in a parallel resource-poor language like Vietnamese. Our method improved translation quality by BLEU score up to 13.65 points, results significantly outperforming strong baseline systems on the general domain and legal domain.

In future work, we also want to explore the effect of adding synthetic parallel data to other resource-poor domains of English - Vietnamese language pair. We will investigate the true merits and limits of Back-translation.

Acknowledgments

This work is funded by the project: Building a machine translation system to support translation of documents between Vietnamese and Japanese to help managers and businesses in Hanoi approach Japanese market, under grant number TC.02-2016-03.

References

Alina Karakanta, J. D. and van Genabith, J. (2018). Neural machine translation for low resource languages without parallel corpora. Machine Translation, 32, 23pp.

Bertoldi, N. and Federico, M. (2009). Domain adaptation for statistical machine translation with monolingual resources. In Proceedings of the fourth workshop on statistical machine translation. Association for Computational Linguistics, pages 182189.

Bojar, O. and Tamchyna, A. (2011a). Improving translation model by monolingual data. In Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT@EMNLP 2011, pages 330336.

Bojar, O. and Tamchyna, A. (2011b). Improving translation model by monolingual data. In Workshop on Statistical Machine Translation.

Caglar Gulcehre, Orhan Firat, K. X. K. C. L. B. H.-C. L. F. B. H. S. and Bengio, Y. (2015). On using monolingual corpora in neural machine translation. CoRR, abs/1503.03535.

Domhan, T. and Hieber, F. (2017). Using target side monolingual data for neural machine translation through multi-task learning. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 15001505.

Dzmitry Bahdanau, K. C. and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Guillaume Lample, Alexis Conneau, L. D. and Ranzato, M. (2018). Unsupervised machine translation using monolingual corpus only. In International Conference on Learning Representations (ICLR).

Gülçehre, Ç., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H., Bougares, F., Schwenk, H., and Bengio, Y. (2015). On using monolingual corpora in neural machine translation. *CoRR*, abs/1503.03535.

Hua Wu, H. W. and Zong, C. (2008). Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, pages 9931000.

Ilya Sutskever, O. V. and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems. pages 31043112.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In EMNLP. volume 3, page 413.

Kishore Papineni, Salim Roukos, T. W. and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) pp. 311-318.

Klein, G., Kim, Y., Deng, Y., Nguyen, V., Senellart, J., and Rush, A. (2018). OpenNMT: Neural machine translation toolkit. In *Proceedings of the 13th Conference of the Association for Machine Translation in the*

Americas (Volume 1: Research Papers), pages 177–184, Boston, MA. Association for Machine Translation in the Americas.

Koehn, P. (2010). Statistical machine translation. Cambridge University Press.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Kyunghyun Cho, Bart Van Merrienboer, C. G. D. B. F. B.-H. S. and Bengio, Y. (2014b). Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Kyunghyun Cho, Bart van Merrienboer, D. B. and Bengio, Y. (2014a). On the properties of neural machine translation: Encoder-decoder approaches. In Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST8).

Luong, M., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *CoRR*, abs/1508.04025.

Minh-Thang Luong, H. P. and Manning, C. D. (2015a). Effective approaches to attention-based neural machine translation. In Proc of EMNLP.

Minh-Thang Luong, H. P. and Manning, C. D. (2015b). Effective approaches to attentionbased neural machine translation. arXiv preprint arXiv:1508.04025.

Nicola Ueffing, Gholamreza Haffari, A. S. (2007). Transductive learning for statistical machine translation. In Annual Meeting-Association for Computational Linguistics. volume 45, page 25.

Patrik Lambert, Holger Schwenk, C. S. a. S. A.-R. (2011). Investigations on translation model adaptation using monolingual data. In Proceedings of the Sixth Workshop on Statistical Machine Translation. Association for Computational Linguistics, pages 284293.

Phuong, L.-H., Nguyen, H., Roussanaly, A., and Ho, T. (2013). A hybrid approach to word segmentation of vietnamese texts.

Rico Sennrich, B. H. and Birch, A. (2016). Improving neural machine translation models with monolingual data. Conference of the Association for Computational Linguistics (ACL).

Rico Sennrich, B. H. and Birch, A. (2016a). Improving neural machine translation models with monolingual data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8696.

Sennrich, R., Haddow, B., and Birch, A. (2015). Improving neural machine translation models with monolingual data. *CoRR*, abs/1511.06709.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proc. NIPS*, Montreal, CA.

Zhang, J. and Zong, C. (2016). Exploiting source-side monolingual data in neural machine translation. pages 1535–1545.