Enhancing the quality of Phrase-table in Statistical Machine Translation for Less-Common and Low-Resource Languages

Minh-Thuan Nguyen*, Van-Tan Bui†, Huy-Hien Vu‡, Phuong-Thai Nguyen*, Chi-Mai Luong§
*Department of Computer Science, University of Engineering and Technology, VNU Hanoi
{thuannm, thainp}@vnu.edu.vn
†Department of Information Technology, University of Economic and Technical Industries
{bvtan}@uneti.edu.vn
‡University of Engineering and Technology, VNU Hanoi
{hienvuhuy}@gmail.com
§Department of Language and Speech Processing, Institute of Information Technology
{lcmai}@ioit.ac.vn

***Abstract*—The phrase-table plays an important role in traditional phrase-based statistical machine translation (SMT) system. During translation, a phrase-based SMT system relies heavily on phrase-table to generate outputs. In this paper, we propose two methods for enhancing the quality of phrase-table. The first method is to recompute phrase-table weights by using vector representations similarity. The remaining method is to enrich the phrase-table by integrating new phrase-pairs from an extended dictionary and projections of word vector presentations on the target-language space. Our methods produce an attainment of up to 0.21 and 0.44 BLEU scores on in-domain and cross-domain (Asian Language Treebank - ALT) English - Vietnamese datasets respectively.**

Keywords-statistical machine translation; phrase-table; vector representation similarity; extend dictionary

I. INTRODUCTION

Phrase-based statistical machine translation (PBSMT) systems have been developed for years and attain success in practice. The core of PBSMT is the phrase-table, which contains words and phrases for SMT system to translate. In the translation process, sentences are split into distinguished parts [1][2]. At each step, for a given source phrase, the system will try to find the best candidate amongst many target phrases as its translation based mainly on phrase-table. Hence, having a good phrase-table possibly makes translation systems improve the quality of translation. However, attaining a rich phrase-table is a challenge since the phrase-table is extracted and trained from large amounts of bilingual corpora which require much effort and financial support, especially for less-common languages such as Vietnamese, Laos, etc.

Latterly, there are several approaches to address this impediment. [3] proposed a method of using new scores generated by a Convolution Neural Network which indicate the semantic relatedness of phrase pairs. They attained an improvement of approximately 0.55 BLEU score. However, their method is suitable for medium-size corpora and create more scores for the phrase-table which can increase the computation complexity of all translation systems.

[4] utilized techniques of pivot languages to enrich their phrase-table. Their phrase-table is made of *source-pivot* and *pivot-target* phrase-tables. As a result of this combination, they attained a significant improvement of translation. Similarly, [5] used a method based on pivot languages to calculate the translation probabilities of *source-target* phrase pairs. Unfortunately, the methods based on pivot languages are not able to apply for the less-common languages.

[6] improved the translation quality by using phrase pairs from an augmented dictionary. They first augmented the dictionary using simple morphological variations and then assigned probabilities to entries of this dictionary by using co-occurrence frequencies collected from bilingual data. However, their method needs a lot of bilingual corpora to estimate accurately the probabilities for dictionary entries, which is not available for low-resource languages.

Recently, techniques using word embedding receive much interest from natural language processing communities. Word embedding is a vector representation of words which conserves semantic information and their contexts words in [7][8]. Additionally, we can exploit the advantage of embedding to represent words in diverse distinction spaces [8]. Inspired by the work of [6] and [8], we propose two methods to enhance the quality of a phrase-table by recomputing weights in phrase-table based on a vector representation similarity of source-language-space and target-language-space, and by generating new phrase-table entries from dictionaries and projections of word vector representations.

In this work, our contribution focuses on enhancing the phrase-table quality by recomputing phrase weights and integrating new translations into the phrase-table. In order to recompute phrase weights, we convert all phrases in the phrase-table and all words in the lexical table to vectors by using vector representation models, which are trained from source-language and target-language monolingual data. Afterward, we use a transformation matrix trained from a small bilingual data to project source-language vectors to the target-language vector space. Then we calculate the cosine similarity between the projected source-language

vector and the target-language vector. In order to integrate new translations into the phrase-table, we use entries in an extended dictionary and generate new phrase pairs by using projections of word vector presentations on the target-language space.

The rest of this paper is organized as follows: Section II describes a method to recompute phrase-table weights via the similarity between vectors. In Section III, we show how to extend phrase-table by using a dictionary with simple morphological variations, generating new phrase pairs, and combining phrase pairs into a traditional phrase-table. Settings and results of our experiments on different bilingual corpus including UET and Asian Language Treebank - ALT dataset are shown in Section IV. We show our conclusion and future works in Section V.

II. Using vector representation similarity to recompute phrase-table weights

Phrase scoring is one of the most important parts in a statistical machine translation system. It estimates weights for phrase pairs based on a large bilingual corpus. Therefore, in less-common and low-resource languages, the estimation is often inaccurate. In order to resolve this problem, we recompute phrase weights by using monolingual data. The traditional phrase-table in an SMT system normally contains four weights: *inverse phrase translation probability, inverse lexical weighting, direct phrase translation probability, and direct lexical weighting*. In order to recompute those weights, we borrow the idea of using a transformation matrix between word vectors to explore similarities among languages [8]. The detail of our method is shown as follows:

Word Vector Representation: We first build two vector representation models for source and target languages by using large amounts of monolingual data, then convert all entries including source and target phrases to vectors.

Train Transformation Matrix: In order to train the linear transformation matrix between the languages, we use a small bilingual corpus containing pairs in a dictionary and short sentences (with a length of sentence is 1 to 4) in a bilingual corpus. With a transformation matrix W, we calculate the projection of the source vector in the target space as follows:

$$We_i = z_i \tag{1}$$

where e_i is the vector of the source word or phrase i and z_i is the projection of the source vector in the target space.

In practice, learning the transformation matrix W can be considered as an optimization problem and it can be solved by minimizing the following error function using a gradient descent method:

$$min_w \sum_{i=1}^{n} \|We_i - f_i\|^2 \tag{2}$$

where e_i is the vector of source word, f_i is the vector of target word in the translation pair i.

Recompute phrase weights: In order to recompute phrase weights, we use an assumption that if two words or phrases in two different languages are translations of each other, the similarity score of them will reflect high value. Therefore, we estimate the probability $p(f_i|e_i)$ of the target word f_i given the source word e_i by the similarity between f_i and the projection of e_i in the target space as shown in Equation **(3)**, where f_i is the target vector, e_i is the source vector, W is the transformation matrix from source to target space, $similarity(f_i, We_i)$ is the formula to measure the similarity between two vectors. In our work, we use cosine similarity as Equation **(4)**:

$$p(f_i|e_i) \approx similarity(f_i, We_i) \tag{3}$$

$$similarity(f_i, We_i) = \frac{f_i.We_i}{\|f_i\|\|We_i\|} \tag{4}$$

In order to recompute phrase translation probabilities, we first train a transformation matrix to project vectors from source to target space and a transformation matrix to project vectors from target to source space. Then use Equation **(3)** to estimate the *direct phrase translation probability* and *inverse phrase translation probability*.

In order to recompute lexical weightings, we have the lexical weight of a phrase f given the phrase e are computed by the following Equation in [9].

$$P_w(f|e,a) = \prod_{i=1}^{n} \frac{1}{|\{j|(i,j) \in a\}|} \sum_{\forall(i,j)\in a} w(f_i|e_j) \tag{5}$$

where $w(f_i|e_j)$ which is word translation score, is estimated by the Equation **(3)**. In order to recompute *direct lexical weighting* and *inverse lexical weighting*, we also train two transformation matrices for projection from source to target space and vice versa.

III. Extending phrase-table and phrase-tables combination

In less-common and low-resource languages, the phrase-table is often obtained from only bilingual corpora which are sparse and do not cover all words and phrase-pairs of these languages. Hence, the phrase-tables do not contain enough words and phrases for translation. To enhance the quality of the translation, we add new translations, which are generated from dictionaries and projections of word vector representations, into the phrase-table.

A. Extending dictionary

In order to increase the number of phrases in the phrase-table, we are able to use traditional dictionaries to add new phrases and words into the phrase-table. Traditional dictionaries usually use base-form of a word to represent an entry. Other word forms such as plural, morphological, etc. only appear in explanations or examples. Thus, we are able to use those word forms to fulfill our phrase-table. This is necessary due to the sparseness of phrases in bilingual corpora for uncommon languages such as Vietnamese, Laos, etc. In this work, from a traditional Vietnamese - English dictionary, we create an extended dictionary by generating all possible morphological variations of nouns and verbs, conjugations for verbs, plural forms for nouns,

Table I
SOME NEW TRANSLATION PAIRS

Noun pairs	New pairs	Verb pairs	New pairs
chăn,noun,blanket	chăn,a blanket chăn,blankets	chăn,verb,tend	chăn,tends chăn,tending chăn,tended
ngói,noun,tile	ngói,a tile ngói,tiles	ăn,verb,eat	ăn,eats ăn,eating ăn,ate ăn,eaten

and singular forms with the article. The sample of this step is shown in Table **I**.

B. Generating new phrase pairs using projections of word vector representations

Adding new phrase and word pairs from a dictionary into the phrase-table is a simple and effective way to enrich the phrase-table. However, having a good dictionary is also costly and time-consuming. Therefore, we propose a method for creating new phrase pairs by using projections of word vector representations. The method includes three main steps shown below:

Step 1: We first create two vector representation models for source and target languages by using large monolingual data. Then, by using a small bilingual data and techniques shown in Section **II**, a transformation matrix among languages is learned to transform word vectors from source to target space.

Step 2: We extract all phrases with a fixed length in the source monolingual data. In this step, the fixed length is 3 or 4.

Step 3: For each extracted phrase, we find its translation in the target language as follows:

Step 3.1: For each word in the phrase, find its projection in the target space by using the transformation matrix, then retrieve top-k most similar words of the projections in the target language. For example, consider a Vietnamese phrase *các cầu_thủ trẻ* (some young players), we found top-5 most similar words to each word in English:

- các (some): *various, diversity, some, many, indigenous*
- cầu_thủ (player): *player, golfer, midfielder, batsman, footballer*
- trẻ (young): *young, middle-aged, dark-haired, child, woman*

Step 3.2: From the results of *Step 3.1*, we find the most reasonable sequence in the target language by using Viterbi algorithm [16] which chooses the best sequence by globally minimizing the sum of the state costs and transitions cost. In our work, we consider each top-k words of one word in the phrase to be one layer, and each word in the above top-k words to be one state. We set state costs to **0**. We also use the probabilities of bigrams as transition costs between two states. For instance, the transition cost of moving from state w_1 to state w_2 is calculated as follows:

$$trans\_cost(w_2|w_1) = 1 - \frac{count(w_1 w_2)}{count(w_1)} \qquad (6)$$

Step 3.3: We filter the phrases obtained from the *Step 3.2* in the target language. A phrase is accepted if it satisfies two conditions. Firstly, all words in a phrase are in one sentence. Secondly, the maximum character-length distance between the two words of the phrase in this sentence is less than a threshold. In our work, the threshold is the character-length of the considered phrase including spaces. For example, after running Viterbi in the Step **3.2**, we had the phrase $some\ player\ young$ (shown in Table **II**). When filtering this phrase in the target language, we found that all words in this phrase appear in the sentence *Loughguile had **some young players** on their side* and in this sentence, the maximum character-length distance between the two words is the character-length distance between the last word $players$ (27) and the first word $some$ (16). This distance is 11 (27-16) which is less than the threshold (17). Therefore, $some\ young\ players$ is opted for a translation of the phrase *các cầu_thủ trẻ*.

Table **II** shows some new phrase pairs which were generated by using our method. The first column indicates source phrases while the second column describes the results after using the Viterbi algorithm to find the best sequence as mentioned in step **3.2**. The third column shows the results after filtering the phrase obtained in step **3.2** in the target language.

Table II
THE SAMPLE OF NEW PHRASE PAIRS GENERATED BY OUR METHOD

Vietnamese phrase	After Viterbi	After filtering
không rõ nguồn_gốc	not clear origin	origins are not clear
các cầu_thủ trẻ	some player young	some young players
hàng giả hàng	goods fake goods	fake goods

C. Phrase-tables combination

In order to combine phrase pairs into a traditional phrase-table, we first generate a phrase-table for the phrase pairs by considering them as a small bilingual corpus, then using the Moses toolkit [1] to align words, extract phrases and score for the phrases. In phrases scoring, we can use the statistic method in Moses or our method shown in Section **II**. Afterward, phrase-tables were combined by the linear interpolation of individual model weights. There are four weights in the traditional phrase-table: the phrase translation probabilities $p(e|f)$ and $p(f|e)$, and the lexical weights $lex(f|e)$ and $lex(e|f)$. The linear interpolation is defined as follows:

$$p(x|y;\lambda) = \sum_{i=1}^{n} \lambda_i p_i(x|y) \qquad (7)$$

where λ_i being the interpolation weight of each model i, and with $(\sum_i \lambda_i) = 1$.

In order to compute the interpolation weights, we follow the approach in [10]. First, we create a tune set by selecting phrase pairs with high probabilities from two phrase-tables that we want to combine. Then, we select weights that minimize a cross-entropy value of this tune set. For example, in our work, to combine the extended dictionary phrase-table with the base phrase-table generated from the bilingual corpus, we first created a tune set containing phrase pairs from the extended dictionary and the base phrase-table. After tuning, the

chosen weights to combine the two phrase-tables were: $[0.68, 0.32]; [0.69, 0.32]; [0.69, 0.32]; [0.7, 0.3]$ which corresponding to the four weights in the phrase-table: Direct phrase translation probability $p(f|e)$, Direct lexical weighting $lex(f|e)$, Inverse phrase translation probability $p(e|f)$, Inverse lexical weighting lex $(e|f)$, where f is source language and e is the target language.

IV. Experiments and results

A. Experiment settings

Extend dictionary: In order to expand the dictionary (shown in Section **III**), we used a filtered dictionary including 19.376 translation pairs, which are extracted from the Vietnamese English dictionary[1]. Then, to add word forms for nouns (7.865 words) and verbs (6.759 words), we used the NodeBox library[2]. As a result, we obtained an extended dictionary containing 44.799 translation pairs.

Word Vector Representation: In order to train source-language and target-language vector models, we used Fasttext's CBOW model in [11]. We utilized Vietnamese text in Leipzig Corpora [12] for the monolingual text of the source language (Vietnamese), and British National Corpus (BNC) in [13] for the monolingual text of the target language (English). The size of these corpora could be seen in Table **III**. Each of model is trained with 50 epochs, the vector space for word embedding was 200-dimensional.

Table III
Monolingual corpora

	Number of Words	Number of Unique Words
Vietnamese	106.275.253	615.877
English	116.863.827	664.168

Train Transformation Matrices: In order to recompute lexical weightings (in Section **II**), we trained transformation matrices by using a small bilingual corpus containing 7.976 translation pairs, which were **only** selected from the filtered dictionary. To recompute phrase translation probabilities (in the Section **II**), we trained transformation matrices using a small bilingual corpus including 8.188 short sentences (length of those sentence are between 1 to 4 in the bilingual data) and 7.976 translation pairs in the Vietnamese-English dictionary.

Generate new phrase pairs: In order to generate new phrase pairs (shown in Section **III**), we used two vector models for Vietnamese and English language as mentioned above. To select a proper phrase pair from many possible phrase pairs, we use *Viterbi* algorithm in [16] to calculate the best path with highest probabilities. We also used the British National Corpus to filter phrases in the target language. As a result, we obtained 100.436 new Vietnamese-English phrase pairs.

Translation System: In all our experiments, we trained our phrase-based statistical machine translation models by using Moses system [1] with Vietnamese as the source language and English as the target language. Note that we use Moses as translation system in all below experiments. In order to clarify the effect of our methods to enhance the quality of the phrase-table, we conducted our experiments on different sizes of the dataset [13] shown in Table **IV**. We also evaluated our method on the external dataset by using the model trained and turned from Big UET dataset and testing the quality of translation on the Vietnamese-English parallel dataset obtained from Asian Language Treebank (ALT) Parallel Corpus in [17] in Table **IV**. The detail of the translation system setting can be described as follow: the maximum sentence and maximum phrase length are 80 and 7 respectively. We followed the default settings of Moses in [1]. All Vietnamese sentences were segmented by Vietnamese Word Tokenizer tool in [14].

Table IV
Bilingual corpora

	Training set	Tuning set	Testing set
Small UET dataset	50.000	1.000	1.000
Big UET dataset	271.822	29.892	15.884
ALT dataset	-	-	20.106

For the experiments using *Small UET dataset*, we conducted 9 following experiments:

- *base*: the baseline of Moses system.
- *base + fd*: We combine the original phrase table[3] with the phrase-table generated from the filtered dictionary by using Moses system.
- *base + ed*: We combine the original phrase-table with the phrase-table generated from the extended dictionary by using Moses system.
- *base + fd_r*: We recompute weights of the phrase-table generated from the filtered dictionary and then combine this phrase-table with the original phrase-table.
- *base + ed_r*: We recomputed weights of the phrase-table generated from the extended dictionary and then combine this phrase-table with the original phrase-table.
- *r*: We recomputed weights of the original phrase-table. Then we use these new weights to replace the original weights in the phrase-table.
- *base +r*: We combine the original phrase-table with the phrase-table obtained from the Experiment *r*.
- *base + r + ed*: We combine the phrase-table obtained in the Experiment *base + r* with the phrase-table generated from the extended dictionary.
- *base + r + n*: We add new phrase pairs generated by our proposed method (shown in the Section **III**) into the phrase-table obtained in the Experiment *base + r*

For the experiments using *Big UET dataset*, we conducted 4 following experiments:

- *base:* the phrase-based SMT baseline by only using Moses system.

[1]https://www.informatik.uni-leipzig.de/~duc/Dict/install.html
[2]https://www.nodebox.net/code/index.php/Linguistics

[3]the original phrase-table: the phrase-table generated from the *Small UET dataset*

- *base + r:* We recomputed weights of the original phrase-table and then combine the new weights with the original weights.
- *base + r + ed:* We combine the phrase-table obtained in the Experiment *base + r* with the phrase-table generated from the extended dictionary.
- *base + r + n:* We add new phrase pairs generated by our proposed method (shown in the Section **III**) into the phrase-table obtained in the Experiment *base + r*

Phrase-tables combination: We used the following tune sets shown in Table **V** to combine two phrase-tables in the above experiments.

B. Experiment results

The result of the experiments is shown in Table **VI** in term of the BLEU score [15]. In all experiments, using an extended dictionary shows better results than using a filtered dictionary. This can be explained by adding a number of entries into the filtered dictionary in Section **III**. Similarly, using dictionary information in the experiments *base + fd* and *base + ed* offers better results than the baseline.

The experiments *base + fd_r* and *base + ed_r* indicate that scoring phrase-table weights by using word vector representation similarity is more effective than that scoring of Moses system. The explanation for this effect is a characteristic of corpora. Moses's method relies heavily on spare bilingual data while our method uses dense monolingual data.

The experiment *r* shows that weights recomputed by word vector representation similarity in phrase-table are able to attain 82% of the BLEU score of Moses system. This means by using major monolingual data and small bilingual data, we create a relatively accurate system comparing to the original Moses which only use bilingual data. In the experiment *base + r*, results of our translation are higher than the baseline in both Big and Small UET dataset, indicating that combining the original Moses's phrase-table and the phrase-table in the experiment *r* enhances an accuracy of phrase-table weights.

In the two remaining experiments, our approach of using the two mentioned methods for enhancing the quality of the phrase-table retrieve better results than the others and the baseline. Notably, the experiment *base + r + n* acquires the highest BLEU score which is 1.36 and 0.21 higher than the baseline in the Small and Big UET Data respectively. The reason is that the number of entries in the phrase-table created by projections of word vector representation on the target-language-space is much higher than those entries of the phrase-table created by the extended dictionary.

In our experiments, we did not take experiments *base+fd, base+ed, base+fd_r* and *base+ed_r* for Big UET Data since we only aim to observe an impact of external phrase-pairs on the small dataset.

C. Examples of translation

We show some translation examples of our translation systems using Big UET dataset in Table **VII**. In the Example 1, it can be seen that the result of the *base+r+n* is similar to the *reference sentence* while the remaining results are incorrect. The explanation is that in our approach, the new phrase pair (*sẽ hồi_phục trong*; *will recover in*), which does not appear in both the extended dictionary and the original phrase-table, was created in the step of generating new phrase-pair (section **III**). However, in the Example 2, all results are incorrect and the result of *base+r+n* is the worst. In our analysis, *bác_sĩ phan_văn_nghiệm trưởng_phòng* was translated to *the chief of the department*. The reason for this incorrect translation is that the new pair *(bác_sĩ phan_văn_nghiệm trưởng_phòng; the chief of the department)* is added directly to the phrase-table. It can be explained that our method cannot produce good enough phrase pairs in this case.

V. Conclusion

In this paper, we proposed two methods to enhance the quality of the phrase-table in PMSMT system. The first method is to recompute weights of phrase-tables by exploiting the dense of monolingual data and the vector representation similarity. The second method is to generate new entries for phrase-table by using a dictionary and by creating new phrase pairs from projections of word vector representation on the target-language space. Our methods help translation systems overcome the spare data of less-common and low-resource language. Using both of two methods, the phrase-table quality has been significantly improved. As a result, the BLEU score increased by absolute 0.21, 1.36 and 0.44 on Vietnamese - English dataset (Big UET, Small UET, and ALT dataset respectively). However, there are some drawbacks in our approach since our method created incorrect entries for the phrase-table. In the future, we will work on specific cases of generating bad phrase pairs, try to apply our methods in a bigger dictionary and integrate more knowledge in ALT corpus such as part-of-speech (POS) tags, syntactic analysis annotations, etc. in order to improve the quality of the translation system.

VI. Acknowledgment

This publication is the output of the ASEAN IVO (http://www.nict.go.jp/en/asean_ivo/index.html) project Open Collaboration for Developing and Using Asian Language Treebank and financially supported by NICT (http://www.nict.go.jp/en/index.html).

References

[1] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst, "Moses: Open source toolkit for statistical machine translation," in Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions, 2007, pp. 177-180.

[2] Philipp Koehn. 2009. "Statistical machine translation," Cambridge University Press.

[3] Peyman Passban, Chris Hokamp, Andy Way, and Qun Liu. 2016. "Improving phrase-based SMT using cross- granularity embedding similarity," In The 19th Annual Conference of the European Association for Machine Translation (EAMT), pages 129–140, Riga, Latvia.

Table V
THE SIZE OF TUNE SETS FOR PHRASE-TABLES COMBINATION

	base + fd	**base + ed**	**base + fd_r**	**base + ed_r**	**base + r**	**base + r + ed**	**base + r + n**
Small UET dataset	23.309				21.819		22.166
Big UET dataset	-	-	-	-	100.000	23.015	26.255

Table VI
RESULTS ON SMALL, BIG UET AND ALT DATASET.

	base	**base+fd**	**base+ed**	**base+fd_r**	**base+ed_r**	**r**	**base+r**	**base+r+ed**	**base+r+n**
Small UET data	26.44	26.66	26.82	26.69	27.08	21.69	26.73	27.1	27.8
Big UET data	28.02	-	-	-	-	-	28.21	28.16	**28.23**
ALT data	19.81	-	-	-	-	-	20.18	20.11	**20.25**

Table VII
TRANSLATION EXAMPLES ON BIG UET DATASET

	Example 1	**Example 2**
source	cô ấy **sẽ hồi_phục trong** 1 tháng nữa	**bác_sĩ phan_văn_nghiệm trưởng_phòng** cấp_cứu là một người rất tận_tâm với bệnh_nhân
reference	she **will recover in** a month	**dr. phan văn nghiệm** , the chief of the emergency department , is very dedicated to patients
base	she will be recovered in a month	the doctor phan_văn_nghiệm emergency bureau 's a very dedicated to the patient
base+r	she will be recovered in a month	the doctor phan_văn_nghiệm emergency bureau is a very dedicated to the patient
base+r+ed	she will be recovered in a month	the doctor phan_văn_nghiệm emergency bureau is a very dedicated to the patient
base+r+n	she **will recover in** a month	emergency **the chief of the department** 's a very dedicated to the patient

[4] Yiming Cui, Conghui Zhu, Xiaoning Zhu, Tiejun Zhao and Dequan Zheng. 2013. "Phrase Table Combination Deficiency Analyses in Pivotbased SMT," In Proceedings of 18th International Conference on Application of Natural Language to Information Systems, pages 355-358.

[5] Zhu, Xiaoning, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao. "Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs," In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, pages 1665–1675, 2014.

[6] Stephan Vogel. 2004. "Augmenting Manual Dictionaries for Statistical Machine Translation Systems," In 2003 Proceedings of LREC, Lisbon, Portugal. pp. 1593-1596.

[7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. "Efficient estimation of word representations in vector space," CoRR, abs/1301.3781.

[8] Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013b. "Exploiting similarities among languages for machine translation," CoRR, abs/1309.4168.

[9] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, 2003, pp. 48-54.

[10] Sennrich, Rico (2012), "Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation," In: Proceedings of EACL 2012.

[11] Bojanowski, Piotr; Grave, Edouard; Joulin, Armand; Mikolov, Tomas: "Enriching Word Vectors with Subword Information," In Transactions of the Association for Computational Linguistics, 5 (2017), S. 135–146

[12] D. Goldhahn, T. Eckart & U. Quasthoff: "Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages," In: Proceedings of the 8th International Language Ressources and Evaluation (LREC'12), 2012.

[13] Hien Vu Huy, Tung-Lam Nguyen Phuong-Thai Nguyen, and M.L. Nguyen, "Boostrapping phrase-based statistical machine translation via wsd integration", In Proceeding of the Sixth International Joint Conference on Natural Language Processing (IJCNLP 2013), pp. 1042-1046, 2013.

[14] Hông Phuong L.., Thi Minh Huyên N., Roussanaly A., and Vinh H.T. (2008), "A Hybrid Approach to Word Segmentation of Vietnamese Texts," In: Martín-Vide C., Otto F., Fernau H. (eds) Language and Automata Theory and Applications. LATA 2008. Lecture Notes in Computer Science, vol 5196. Springer, Berlin, Heidelberg

[15] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu. "BLEU: a method for automatic evaluation of machine translation," In ACL, 2002.

[16] J. Forney, G.D., "The Viterbi algorithm," Proc. IEEE, vol. 61, no. 3, pp. 268-278, 1973.

[17] Hammam Riza, Michael Purwoadi, Gunarso, Teduh Uliniansyah, Aw Ai Ti, Sharifah Mahani Aljunied, Luong Chi Mai, Vu Tat Thang, Nguyen Phuong Thai, Vichet Chea, Rapid Sun, Sethserey Sam, Sopheap Seng, Khin Mar Soe, Khin Thandar Nwet, Masao Utiyama, Chenchen Ding. (2016) "Introduction of the Asian Language Treebank" Oriental COCOSDA.