Baltic J. Modern Computing, Vol. 4 (2016), No. 2, pp. 190–202

Pivoting Methods and Data for Czech-Vietnamese Translation via English

Duc Tam HOANG, Ondřej BOJAR

Charles University in Prague, Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics

hoangdt@comp.nus.edu.sg, bojar@ufal.mff.cuni.cz

Abstract. The statistical approach to machine translation (MT) relies heavily on large parallel corpora. For many language pairs, this can be a significant obstacle. A promising alternative is pivoting, i.e. making use of a third language to support the translation. There are a number of pivoting methods, but unfortunately, they were not evaluated in comparable settings. We focus on one particular language pair, Czech↔Vietnamese translation, with English as the pivoting language, and provide a comparison of several pivoting methods and the baseline (direct translation). Besides the experiments and analysis, another contribution is the datasets that we have collected and prepared for the three languages.

Keywords: Statistical Machine Translation, Czech-Vietnamese, parallel corpus, pivoting methods, phrase table triangulation, system cascades

1 Introduction

Large parallel corpora are of utmost importance for statistical machine translation (SMT) for producing reliable translations. Unfortunately, for most pairs of living languages, the amount of available parallel data is not sufficient. "Pivoting" methods make use of a third language ("pivot language") to support the translation.

Over past years, a number of pivoting methods have been proposed. Most of the works were conducted using *multi-parallel corpora* such as Europarl (Koehn, 2005), where the same text is available in more than two languages. In a realistic condition, the two corpora, source-pivot corpus and pivot-target corpus, are *independent*, i.e. coming from different sources. We expect that some of the approaches are more beneficial in only one of the two conditions and for sure, some approaches utilizing multilingual corpora are not applicable for independent corpora at all (Kumar et al., 2007; Chen et al., 2008).

In this work, we carry out experiments to directly compare several methods of pivoting. We select Czech and Vietnamese, a relatively unexplored language pair, for the experiments. English is chosen for the role of pivot language because it offers the largest parallel corpora with both Czech and Vietnamese.

This paper has two main contributions. (1) The paper evaluates a wide range of pivoting methods in a directly comparable setting and under the more realistic condition where the parallel corpora are independent (as opposed to multi-parallel). (2) It is the first study which focuses on machine translation between Czech and Vietnamese. It describes and publishes the corpora that we have collected and processed.

The remainder of this paper is organized as follows. Section 2 discusses related work of pivoting methods. Section 3 describes the dataset that we collected, prepared and released. Section 4 presents experimental set up, results and discussions. Finally, Section 5 concludes the paper.

2 Pivoting Methods

Pivoting is formulated as the translating task from a source language to a target language through one or more pivot languages. An important, yet mostly overlooked aspect in pivoting is the relation between the source-pivot and pivot-target corpora. For example, Chen et al. (2008) reduce the size of the phrase table by filtering out phrase pairs if they are not linked by at least one common pivot phrase. Kumar et al. (2007) combine word alignments using multiple pivot languages to correct the alignment errors trained on the source-target parallel data. Both methods (implicitly) rely on the fact that the corpora contain the same sentences available in multiple languages. While this is a reasonable assumption for a *multi-parallel corpus*, the methods are not applicable for *independent parallel corpora*.

In our study, we compare pivoting methods which can be applied under the perhaps more realistic condition that the source-pivot and pivot-target corpora are independent (Tiedemann, 2012a; Tiedemann and Nakov, 2013). This section discusses such methods and highlights their difference and potential. Each method has a number of configuration options which significantly affect the translation quality, we explore them empirically in Section 4 below.

2.1 Synthetic Corpus/Phrase Table

The synthetic corpus method (Gispert and Mariño, 2006; Galuščáková and Bojar, 2012) and the phrase table (PT) translation method (called synthetic phrase table) (Wu and Wang, 2007) aim to generate training data from MT output. Specifically, an MT system, which translates pivot language into the source or target language, is employed to translate a corpus or a phrase table of the other language pair. The result is a source→target corpus or phrase table with one side "synthetic", i.e. containing MT translated data. The synthetic corpus or phrase table is then used to build the source→target MT system.

Using MT translated data is generally seen as a bad thing. The model can easily reproduce errors introduced by the underlying MT system. In practice, however, machine-generated translations need not be always harmful, especially when they compensate for the lack of direct bilingual training data. For example, Gispert and Mariño (2006) report impressive English↔Catalan translation results by translating the English-Spanish corpus using a Spanish→Catalan MT system. The results are on par with the translation quality of English↔Spanish translation. Similarly, Galuščáková and Bojar (2012) observe that pivoting through Czech was better than direct translation from English to Slovak, due to a large difference in training data size.

Between the two methods, the task of translating a phrase table poses different challenges compared to the task of translating a corpus. Phrasal input is generally much shorter than a sentence and a lot of contextual information is lost (even considering the limited scope of existing language models).

2.2 Phrase Table Triangulation

The phrase table triangulation method (Cohn and Lapata, 2007; Zhu et al., 2014), sometimes called simply triangulation, generates an artificial source-target phrase table by directly joining two phrase tables (source-pivot and pivot-target) on common pivot phrases.

Once the tables are combined, approaches to triangulating the two phrase tables diverge in how they set the scores for the phrases. There are two options for estimating the necessary feature scores of the new phrase table: multiplying the original posterior probabilities or manipulating the original co-occurrence counts of phrases.

The first option views the triangulation as a generative probabilistic process on two sets of phrase pairs, s-t and p-t. Assuming the independent relations between three languages, the conditional distribution $p(s|t)$ is estimated over source-target phrase pair s-t by marginalising out the pivot phrase p:

$$\begin{aligned} p(s|t) &= \sum_p p(s|p,t) \times p(p|t) \\ &\approx \sum_p p(s|p) \times p(p|t) \end{aligned} \tag{1}$$

Afterwards, the feature values of identical phrases pairs are combined in the final phrase table. Either the scores are summed up or maximized (i.e. taking the higher of the score values).

The second option estimates the co-occurrence count of the source and target phrases $c(s,t)$ from the co-occurrence counts $c(s,p)$ and $c(p,t)$ of the component phrase pairs. Afterwards, the feature scores are estimated by the standard phrase extraction (Koehn, 2010).

$$c(s,t) = \sum_p f(c(s,p), c(p,t)) \tag{2}$$

In Equation 2, function f is the desired approximation function. Zhu et al. (2014) proposed four functions f: minimum, maximum, arithmetic mean and geometric mean.

Phrase table triangulation methods have received much attention, yet they have not been tested with two disjoint and independent corpora.

2.3 System Cascades

A widely popular method, system cascades (Utiyama and Isahara, 2007), simply uses two black-box machine translation systems in a sequence. The first system translates the input from the source language into the pivot language. The second system picks up the pivot hypothesis and translates it into the target language.

Formally, the problem of finding the best sentence $\hat{e}$ for a foreign input sentence f is defined as maximizing the translation score from source sentence f to a pivot sentence p, then from p to target sentence e:

$$\hat{e} \approx \arg\max_{e,p_i} p_{smt}(p_i|f) \times p_{smt}(e|p_i) \quad (3)$$

where p_i is a pivot hypothesis of the first MT system and serves as the input of the second system.

Because investigating all possible pivot sentences p is too expensive, p is chosen from the list of n-best translations of the source sentence. Sometimes, the first system is not capable of providing a list of possible translations and pivot hypotheses are limited to $n = 1$, taking the top candidate only.

2.4 Phrase Table Interpolation for System Combination

Each of the pivoting methods described above leads to a separate MT system. This opens a possibility of combining these systems, hoping that the strengths of one method would offset the weaknesses of other methods. We choose to combine multiple systems by linearly interpolating translation models. This method, called "phrase table interpolation", is defined as follows:

$$p(e|f;\lambda) = \sum_{i=1}^{n} \lambda_i p_i(e|f) \quad (4)$$

where λ_i is the interpolation weight of translation model i and satisfies the condition $\sum_i \lambda_i = 1$.

We note that the system cascades method does not have a single phrase table. It directly uses the two SMT systems, rather than building a new SMT system. It thus does not lend itself to this combination method. We circumvent the problem by creating a synthetic phrase table from the development and test sets, each translated with the cascades method. We pair the translated text with the original text to create a small synthetic corpus. A phrase table is then extracted from the synthetic corpus and used in the combination.

3 Dataset Created and Released

Czech and Vietnamese are the national languages of the Czech Republic and Vietnam, respectively. Furthermore, the two languages are not under-resourced on their own, but the amount of bilingual corpora between them is very limited despite the large Vietnamese community living in the Czech Republic. So far, no effort has been put into developing an MT tool specifically for this language pair.

We wish to investigate the potential of pivoting methods for translating between Czech and Vietnamese. After carefully examining the potential of all possible pivot languages, we decide to select English as the sole pivot language. It is the only language that provides sufficient resources to act as a bridge between Vietnamese and Czech.

We created and released two sets of multilingual datasets: a set of test data and a set of parallel corpora.

3.1 WMT Test Data

Our test set was derived from the WMT 2013 shared task,[1] which consists of 3000 aligned sentences from newspapers. We opted for the 2013 set, because more recent WMT test sets were no longer multi-parallel across all the languages. The WMT 2013 test set spanned across six languages (Czech, English, German, French, Spanish and Russian) and we extended it to include Vietnamese.

Table 1. Statistics of test data

	# sentences	# words
Czech	3,000	48,472
English	3,000	56,089
Vietnamese	3,000	75,804

Our contribution was created by human translators working in two stages. The first stage delivered a Vietnamese translation from the English side of the WMT 2013 test set, sometimes by post-editing machine-translated text. The second stage was a careful check to arrive at fluent Vietnamese text. Finally, we prepared a multi-lingual test set for Czech, English and Vietnamese. Table 1 gives the statistics of the test set.

3.2 Training Data

The training data is composed of parallel corpora among the source, target and pivot languages. For Czech-English language pair, we used CzEng 1.0, a Czech-English parallel corpus (Bojar et al., 2012) to train the translation model. For Czech-Vietnamese and English-Vietnamese, we collected available bitexts from the Internet as there were no ready-made corpora sufficient to train the translation models.

[1] http://www.statmt.org/wmt13

Table 2. Statistics of Czech-Vietnamese training data

	Original		**Cleaned**	
	Czech	Vietnamese	Czech	Vietnamese
# sentences	1,337,199	1,337,199	1,091,058	1,091,058
# words	9,128,897	12,073,975	6,718,184	7,646,701
# unique words	224,416	68,237	195,446	59,737

Table 3. Statistics of English-Vietnamese training data

	Original		**Cleaned**	
	English	Vietnamese	English	Vietnamese
# sentences	2,035,624	2,035,624	1,113,177	1,113,177
# words	16,638,364	17,565,580	8,518,711	8,140,876
# unique words	91,905	78,333	69,513	58,286

We collected data from two main sources: OPUS[2] and TED talks.[3] OPUS is a growing multilingual corpus of translated open source documents. It covers over 90 languages and includes data from several domains (Tiedemann, 2012b). The majority of Vietnamese-English and Vietnamese-Czech bitexts in OPUS were subtitles from motion pictures. As such, these bitexts were not always close translations; due to various constraints of the domain, the texts were often just paraphrases. The later source contained selected TED talks which were provided in English and equipped with transcripts in Czech and/or Vietnamese. There were 1198 talks for which English and Vietnamese transcripts are available. There were 784 TED talks for which Czech and Vietnamese transcripts are available.

Our preliminary analysis indicated that the collected datasets were noisy to the extent that the noise would harm the performance of SMT approaches. Hence, we opted for a semi-automatic cleanup of the corpora (both Czech-Vietnamese and English-Vietnamese). We improved the corpus quality by two steps: normalizing and filtering. The normalizing step cleaned up the corpora based on some typical formatting patterns in subtitles and transcripts (e.g. we tried to rejoin sentences spanning over multiple subtitles). The filtering step relied on the filtering tool used in the development of the CzEng corpus (Bojar et al., 2012). We trained the tool on a set of 1,000 sentence pairs which had been selected randomly from the corpus and manually annotated. Overall, the normalization and filtering reduced the size of the Czech-Vietnamese corpus by about 32.25% and the size of the English-Vietnamese corpus by about 51.29% (the number of words). The statistics of the training data is shown in Table 2 and 3. Our analysis showed that the cleaning phrase helped in improving the performance of the translation model trained on the collected datasets.

[2] http://opus.lingfil.uu.se

[3] https://www.ted.com/talks

4 Experiments

We empirically evaluate the pivoting methods in the context of Czech↔Vietnamese translation. We also carry out a brief evaluation on the quality of Czech↔English and English↔Vietnamese translations. This provides an insight into the corpus quality, which affects the final performance of pivoting methods.

4.1 Setup

The experiments are carried out using using Moses framework (Koehn et al., 2007). Instead of Moses standard EMS, we use Eman (Bojar and Tamchyna, 2013) to manage the large number of experiments.

We use the standard phrase-based SMT approach which follows the log-linear model. The model features include the translation model, language model, distance-based reordering, word penalty and phrase penalty (no lexicalized reordering model). The translation models are trained on the parallel data that we have prepared (see Section 3). Word alignments are created automatically on the bitexts using Giza++ (Och and Ney, 2003), followed by the standard phrase extraction (Koehn et al., 2003). Three language models are trained using the KenLM language modeling toolkit (Heafield, 2011) with the order of 5.

For the tuning and final evaluation, we split the prepared Czech-English-Vietnamese WMT 2013 set into two parts: the first 1500 sentences as the development set and the remaining 1500 sentences as the test set. The log-linear model is optimized by tuning on the development data with minimum error rate training (MERT, Och (2003)) as the tuning method and BLEU as the tuning metric (Papineni et al., 2002).

The pivoting methods are implemented and processed using the available data that we have. The experimental results are evaluated using BLEU (as implemented in Moses scorer; single-reference, lowercased, and in the tokenization used by the MT system). We also carry out manual evaluation for the final results.

4.2 Baseline Systems

We first build the SMT system by training on the direct parallel data for all 6 translation directions among Czech, English and Vietnamese. Of the 6 component systems, we use the SMT systems trained on the direct Czech↔Vietnamese parallel data as the baseline system.

Table 4 shows the experimental results of six component systems on the test set. We can see that the Czech→Vietnamese and Vietnamese→Czech baseline systems attain very low results (10.59 and 7.62 BLEU points). This is not surprising. Despite the preparation step, the Czech↔Vietnamese training data is still noisy. The essence of transcribed bitexts is paraphrasing, which may be correct in a particular context but incorrect in general. Furthermore, the properties of the examined languages (Czech inflective with very rich morphology, Vietnamese analytic with rather fixed word order) render the Czech-Vietnamese translation as a difficult problem.

Our analysis shows that the component systems for English↔Vietnamese translation perform relatively well. This is attributed by the similarity between English and

Table 4. Performance of baseline systems by direct translation

Direction	Label	BLEU
Czech→English	cs→en	23.23
English→Czech	en→cs	15.26
Vietnamese→English	vi→en	33.88
English→Vietnamese	en→vi	34.45
Czech→Vietnamese	cs→vi	**10.59**
Vietnamese→Czech	vi→cs	**7.62**

Vietnamese, notably the small number of inflectional morphemes. With the collected dataset, we attain competitive results compared to current English↔Vietnamese MT translation.

4.3 Results of Pivoting Methods

4.3.1 Phrase Table Translation We choose to conduct the phrase table translation method, which is similar to the *synthetic corpus* method. To create synthetic Czech↔Vietnamese PTs, there are two options:

1. Translating the English side of English↔Vietnamese phrase tables into Czech using the English→Czech component MT system.
2. Translating the English side of Czech↔English phrase tables into Vietnamese using the English→Vietnamese component MT system.

After translation, the probabilities and lexical weights are kept from the original phrase tables.

Table 5. Performance of synthetic phrase table method

Option	vi→cs	cs→vi
Translating English↔Vietnamese phrase table	7.34	9.67
Translating Czech↔English phrase table	8.40	12.09
Direct Translation (Baseline)	7.62	10.59

Table 5 shows the performance of the two options. We see that translating the large CzEng 1.0 phrase table by the small systems achieves better results than the other way around, regardless of the translation direction. We note that not only the CzEng 1.0 PT has a better coverage, but also the English→Vietnamese system delivers translations of a relatively good quality. The English→Czech system faces the problem of incorrect word forms even though the morphemes are correct. We also note that the PT translation method which involves translating Czech↔English phrase table attains better results than the baseline systems. This shows the potential of pivoting methods over the direct translation.

4.3.2 Phrase Table Triangulation We followed two specific options to conduct phrase table triangulation. Each option in turn offers a number of ways to merge the feature values of identical pivoted phrase pairs.

1. Pivoting posterior probabilities, merging by the summation or maximization function
2. Pivoting the co-occurrence counts, approximating by the minimum, maximum, arithmetic mean or geometric mean function

For each of the translation directions, these two options result in six phrase tables which have the same phrase pairs but different feature values. Table 6 shows the performance of all the setups.

Table 6. Comparison between the six options of PT triangulation method

Option	**Function**	**vi→cs**	**cs→vi**
1	summation	**7.44**	**10.28**
1	maximization	7.21	9.64
2	minimum	7.24	9.86
2	maximum	6.38	7.64
2	arithmetic mean	6.25	6.95
2	geometric mean	7.05	9.24
Direct Translation (Baseline)		7.62	10.59

First, we can see that both options of the triangulation method receive lower BLEU scores, compared to the phrase table translation method. The result is rather interesting because the triangulation method has an appealing description. It is generally considered a good system, sometimes outperforming direct translation. The primary reason for the failure here is the high level of noise created by triangulation. The method doubles the amount of noise in both phrase tables, thus decreasing the overall performance. Moreover, as our corpora are independent, the overlapping part is small. This results in a low coverage of phrases.

Second, re-estimating co-occurrence counts appears to be less effective than combining the probabilities directly. The primary reason is the difference between two phrase tables. The Czech-English phrase table is much larger than the English-Vietnamese phrase table. As the co-occurrence counts are biased either the large PT or the small PT, thus minimizing the difference between valid and noise phrase pairs. Hence, the noisy pairs acquire probabilities as high as the valid pairs. When the co-occurrence counts are

Table 7. Performance of system cascades method

n	1	2	5	10	20	30	50	75	100
cs→vi	9.05	9.19	9.33	9.50	9.70	9.70	9.80	**9.82**	9.82
vi→cs	13.35	13.51	13.65	13.71	13.77	**13.83**	13.73	13.75	13.79

biased towards the large PT (i.e. the maximum and arithmetic mean functions), the high number of common phrases worsens the probabilities.

Another observation shows that computation of the new probability favours summation over maximization. It is reasonable that the final probability of a source- target pairs should be computed over all middle-phrases rather than just one phrase. One unit (word or phrase) may have more than one translation in other language.

4.3.3 System Cascades For system cascades, we use the component systems to translate each step of the process. There are two directions of translation, which lead to two different settings for the system cascades method.

For Vietnamese→Czech system cascades method, we first use the Vietnamese→English component MT system to translate the input from Vietnamese into English. We then use the English→Czech component MT system to translate the English sentence into Czech.

For Czech→Vietnamese system cascades method, we first use the Czech→English component MT system to translate the input from Czech into English. We then use the English→Vietnamese component MT system to translate the English sentence into Vietnamese.

In our experiments, we select n from $\{1, 2, 5, 10, 20, 30, 50, 75, 100\}$ to verify the effectiveness of using n-best translations instead of just selecting the top hypothesis. The list of n-best translations allows the second system to compensate for errors of the first system's single-best output, thus producing a better translation.

Table 7 confirms our claim that the n-best list of hypotheses helps system cascades. Furthermore, the system cascades method achieves higher results than the baseline system and other pivoting methods. The promising performance of system cascades comes from the fact that the method uses complete translations. During the translation process, pivoting sentences are broken into phrases separately for each of the two phrase tables. Only a small portion of phrases remains intact during the process. In most of the cases, the segmentation into phrases is different for the pivot-target translation and for the source-pivot translation.

4.3.4 Combination through Phrase Table Interpolation We adopt the uniform weights to perform phrase table interpolation, which has shown to be robust (Cohn and Lapata, 2007). We adapt all four features of the standard Moses SMT translation model: the phrase translation probabilities and the lexical weights.

Table 8. Automatic evaluation of Czech↔Vietnamese translation

Method	**PT Size**	**vi→cs**	**cs→vi**
Direct Translation	8.70M	7.62	10.59
PT Translation	53.21M	8.40	12.09
PT Triangulation	61.50M	7.44	9.86
System Cascades	0.08M	9.82	**13.83**
Combination (PT Interpolation)	95.00M	**10.12**	**13.80**

Table 8 summarizes our experimental results using automatic scoring. It includes the results of the individual systems and the combined system, which is built based on the interpolated phrase table.

We further conduct manual evaluation over the final results of Czech→Vietnamese translation. We perform relative ranking among 5 systems, the established practice of WMT. To interpret this 5-way ranking, we adopt the technique used by WMT until 2013 (before TrueSkill): we extract the 10 pairwise comparisons from each ranking. For a given system, we report the proportion of pairs in which the system was ranked equally or higher than its competitor (out of all pairs where the system was evaluated), see the column "≥ Others" in Table 9. Additionally, we report a simpler interpretation of the 5-way ranking following Bojar et al. (2011). Each 5-way ranking is called a "block" and we report how often each system was among the winners in this block. Since we are comparing 5 systems, all our blocks include all systems, so "≥ All in block" simply means the rate of wins.

Table 9. Manual evaluation of Czech→Vietnamese translation

Method	**≥ Others**	**≥ All in Block**
Direct Translation	0.76	0.56
PT Translation	0.71	0.48
PT Triangulation	0.77	0.56
System Cascades	**0.86**	**0.56**
Combination (PT Interpolation)	**0.85**	**0.60**

Tables 8 and 9 provide the same picture: the system combination improves a little over the system cascades method.

We note that the performance of a specific method heavily depends on languages, domains and corpora in question. For example, system cascades achieved the best results with our datasets and the performance of phrase table translation is better when translating the larger (Czech-English) phrase table with the smaller (English-Vietnamese) MT system than the other way around, regardless of the final translation direction (Czech↔Vietnamese) using the translated phrase table.

5 Conclusion

We carried our a set of experiments with baseline direct translation and three types of pivoting methods, optionally concluded by a last step that combines the different approaches to a single system, improving over each of the individual components. Our comparative study suggests that in absence of a multi-parallel corpus, simple cascading of systems outperforms methods manipulating the phrase table.

To support further experiments in Czech↔Vietnamese machine translation, we assembled and described two training corpora and created one test set. The corpora are available in the Lindat repository:

– http://hdl.handle.net/11234/1-1594 (WMT13 Vietnamese Test Set)
– http://hdl.handle.net/11234/1-1595 (CsEnVi Pairwise Parallel Corpus)

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 645452 (QT21).

This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

References

Bojar, Ondřej and Aleš Tamchyna. 2013. The design of Eman, an experiment manager. *The Prague Bulletin of Mathematical Linguistics*, 99:39–56.

Bojar, Ondřej, Miloš Ercegovčević, Martin Popel, and Omar Zaidan. 2011. A Grain of Salt for the WMT Manual Evaluation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 1–11, Edinburgh, Scotland, July. Association for Computational Linguistics.

Bojar, Ondřej, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel, and Aleš Tamchyna. 2012. The joy of parallelism with CzEng 1.0. In *Proceedings of the 2012 International Conference on Language Resources and Evaluation*.

Chen, Yu, Andreas Eisele, and Martin Kay. 2008. Improving statistical machine translation efficiency by triangulation. In *Proceedings of the International Conference on Language Resources and Evaluation*.

Cohn, Trevor and Mirella Lapata. 2007. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Galuščáková, Petra and Ondřej Bojar. 2012. Improving SMT by Using Parallel Data of a Closely Related Language. In *Proceedings of the Fifth International Conference Baltic Human Language Technologies*, volume 247 of *Frontiers in AI and Applications*, pages 58–65, Amsterdam, Netherlands. IOS Press.

Gispert, Adrià De and José B. Mariño. 2006. Catalan-english statistical machine translation without parallel corpus: Bridging through spanish. In *Proceedings of 5th International Conference on Language Resources and Evaluation*, pages 65–68.

Heafield, Kenneth. 2011. KenLM: faster and smaller language model queries. In *Proceedings of the 2011 Sixth Workshop on Statistical Machine Translation*.

Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*.

Koehn, Philipp, 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 *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit*, volume 5, pages 79–86.

Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge University Press.

Kumar, Shankar, Franz Josef Och, and Wolfgang Macherey. 2007. Improving word alignment with bridge languages. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Och, Franz Josef. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*.

Tiedemann, Jörg and Preslav Nakov. 2013. Analyzing the use of character-level translation with sparse and noisy datasets. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 676–684, Hissar, Bulgaria, September. INCOMA Ltd. Shoumen, BULGARIA.

Tiedemann, Jörg. 2012a. Character-based pivot translation for under-resourced languages and domains. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 141–151, Avignon, France, April. Association for Computational Linguistics.

Tiedemann, Jorg. 2012b. Parallel data, tools and interfaces in OPUS. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

Utiyama, Masao and Hitoshi Isahara. 2007. A comparison of pivot methods for phrase-based statistical machine translation. In *Proceedings of the 2007 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies*.

Wu, Hua and Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Zhu, Xiaoning, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao. 2014. 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*.

Received May 3, 2016 , accepted May 10, 2016