Syntactic Matching Methods in Pivot Translation

Akiva Miura[†], Graham Neubig[†,††], Katsuhito Sudoh[†] and Satoshi Nakamura[†]

The pivot translation is useful method for translating between languages that contain little or no parallel data by utilizing equivalents in an intermediate language such as English. Commonly, phrase-based or tree-based pivot translation methods merge source–pivot and pivot–target translation models into a source–target model. This tactic is known as triangulation. However, the combination is based on the surface forms of constituent words, and it often produces incorrect source–target phrase pairs because of interlingual differences and semantic ambiguities in the pivot language. The translation accuracy is thus degraded. This paper proposes a triangulation approach that utilizes syntactic subtrees in the pivot language to avoid incorrect phrase combinations by distinguishing pivot language words by their syntactic roles. The results of the experiments conducted on the United Nations Parallel Corpus demonstrate that the proposed method is superior to other pivot translation approaches in all tested combinations of languages.

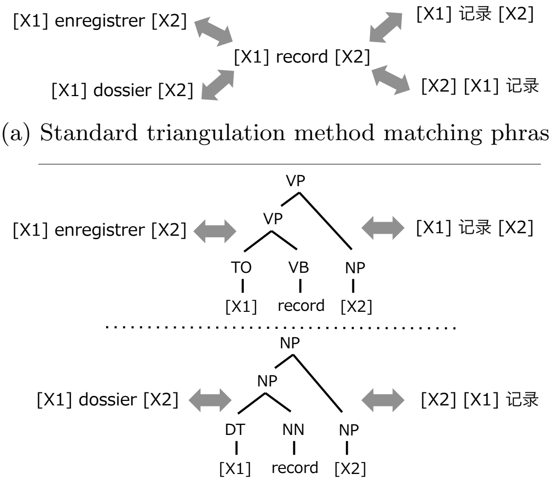
Key Words: *Pivot Translation, Machine Translation, Parallel Corpus, Low-Resourced Language Pairs, Syntactic Analysis*

1 Introduction

Translation effected using models trained on larger parallel corpora can achieve greater accuracy (Dyer, Cordova, Mont, and Lin 2008) in statistical machine translation (SMT) (Brown, Pietra, Pietra, and Mercer 1993). Unfortunately, most language pairs are restricted in terms of readily available parallel corpora: some have fewer than 100k sentence pairs; others do not contain any. This paucity is especially true of language pairs that do not include English, and the problem is difficult to overcome because it would cost millions of dollars to manually produce a high-quality parallel corpus.

One effective solution to surmount the scarceness of bilingual data is to introduce a pivot language that contains existing parallel data with respect to both the source and target languages (de Gispert and Mariño 2006). The triangulation is both popular and effective among the various methods that employ pivot languages (Utiyama and Isahara 2007; Cohn and Lapata 2007). This process first combines source–pivot and pivot–target translation models (TMs) into a source–

[†] Graduate School of Information Science, Nara Institute of Science and Technology
[††] Language Technologies Institute, Carnegie Mellon University



(a) Standard triangulation method matching phrases

(b) Proposed triangulation method matching subtrees

Figure 1 Example of disambiguation by parse subtree matching (Fr-En-Zh), [X1] and [X2] are non-terminals for sub-phrases.

target model and then translates data using this merged model. The procedure of triangulating two TMs into one has been examined for different frameworks of SMT and its effectiveness has been confirmed both for Phrase-Based SMT (PBMT) (Koehn, Och, and Marcu 2003; Utiyama and Isahara 2007) and for Hierarchical Phrase-Based SMT (Hiero) (Chiang 2007; Miura, Neubig, Sakti, Toda, and Nakamura 2015). However, word interlingual differences in word usage and word sense ambiguity cause difficulties in accurately learning the correspondences between the source and target phrases. Thus, the accuracy of triangulated models is lesser than the precision attained by models trained on direct parallel corpora.

In the triangulation method, source–pivot and pivot–target phrase pairs are connected as source–target pairs if a common pivot-side phrase is available. Figure 1-(a) illustrates a sample standard triangulation on the Hiero TM, which combines the hierarchical rules of phrase pairs by matching pivot phrases with equivalent surface forms. This example also demonstrates the problems of ambiguity: the English word "record" can correspond to several different parts-of-speech according to the context. More broadly, phrases that include this word can also potentially take different grammatical structures, but it is impossible to uniquely identify these constructions unless information is provided with regard to the surrounding context.

This varying syntactic structure will influences translation. For example, the French verb "enregistrer" corresponds to the English verb "record". At the same time, the French noun "dossier" matches the noun form of the English word "record". In a more extreme instance, Chinese does not incorporate inflections depending on the part-of-speech of the word. Thus, although the word order changes, the Chinese term "记录" is used even in contexts where record is employed as a different grammatical category. These specifics might result in the incorrect connection of "[X1] enregistrer [X2]" and "[X2] [X1] 记录" even though the proper correspondence of "[X1] enregistrer [X2]" and "[X1] dossier [X2]" would be "[X1] 记录 [X2]" and "[X2] [X1] 记录" respectively.. Hence, a superficial phrase matching method based solely on the surface form of the pivot often combines incorrect phrase pairs, causing translation errors if the translation scores of the mached pairs are estimated to be higher than the actual correspondences.

Given this background, it is hypothesized that the disambiguation of these cases would be simpler if necessary syntactic information such as phrase structures is considered during the pivoting process. To incorporate this intuition into the models introduced in this paper, the authors propose a method that considers the syntactic information of the pivot phrase as shown in Figure 1-(b). In this manner, the model distinguishes the translation rules extracted in contexts within which the English symbol string "[X1] record [X2]" behaves as a verbal phrase from situations in which the same string acts as a nominal phrase.

Specifically, the method posited in this paper is based on Synchronous Context-Free Grammars (SCFGs) (Aho and Ullman 1969; Chiang 2007), which are widely used in tree-based machine translation frameworks. First, Section 2 of the paper provides a quick review of SCFGs. The baseline triangulation method that only uses the surface forms for performing the triangulation is detailed in Section 3, and two methods for triangulation based on syntactic matching are proposed in Section 4. The first method places a hard restriction on the exact matching of parse trees (Section 4.1) included in translation rules, whereas the second places a softer limitation and allows partial matches (Section 4.2). Experiments of pivot translation on the United Nations Parallel Corpus (Ziemski, Junczys-Dowmunt, and Pouliquen 2016) were performed by the authors to investigate the proposed method's impact on pivot translation quality. The results of these investigations are presented in Section 5. These findings demonstrate that the posited process indeed provides significant gains in accuracy (of up to 2.3 BLEU points), in almost all tested combinations of five languages with English used as the pivot language. In addition, as an auxiliary result, the authors compared pivot translations effected through the use of the proposed method to those made through zero-shot neural machine translation. These outcomes confirm that the triangulation of symbolic TMs still significantly outperforms neural MT in the zero-resource scenario.[1]

2 Machine Translation Framework

2.1 Synchronous Context-Free Grammars (SCFGs)

In this section, the authors initially deal with SCFGs, particularly hierarchical phrase-based translation (Hiero) (Chiang 2007), which are widely used in machine translation. The elementary structures used in translation in SCFGs are synchronous rewrite rules with aligned pairs of source and target symbols on the right side as in

$$X \rightarrow \langle \bar{s},\ \bar{t} \rangle \tag{1}$$

where X is the head symbol of the rewrite rule, and $\bar{s}$ and $\bar{t}$ are both strings of terminals and non-terminals on the source and target sides respectively. Each string in the right-side pair has the same number of indexed non-terminals, and identically indexed non-terminals correspond to each-other. A synchronous rule could also, for example, take the form of

$$X \rightarrow \left\langle X_0 \ \text{ of } \ X_1,\ X_1 \text{ 的 } X_0 \right\rangle . \tag{2}$$

Synchronous rules can be extracted based on parallel sentences and automatically obtained word alignments. Each extracted rule is scored with phrase translation probabilities in both directions $\phi(\bar{s}|\bar{t})$ and $\phi(\bar{t}|\bar{s})$, lexical translation probabilities in both directions $\phi_{lex}(\bar{s}|\bar{t})$ and $\phi_{lex}(\bar{t}|\bar{s})$, a word penalty counting the terminals in $\bar{t}$, and a constant phrase penalty of 1.

At the time of translation, the decoder searches for the target sentence that maximizes the derivation probability, which is defined as the sum of the scores of the rules used in the derivation, and the log of the language model (LM) probability over the target strings. When not considering an LM, it is possible to efficiently find the best translation for an input sentence using the CKY+ algorithm (Chappelier, Rajman, et al. 1998). When using an LM, the expanded search space is further reduced based on the limit on expanded edges, or total states per span, through a procedure such as cube pruning (Chiang 2007).

[1] A preliminary version of this paper has presented in (三浦，Neubig，中村 2016b) and (Miura, Neubig, Sudoh, and Nakamura 2017).

2.2 Hierarchical Rules

The rules used in Hiero are specifically discussed in this section. Hierarchical rules are composed of the initial head symbol S and synchronous rules containing terminals and singular type of non-terminal X.[2] Hierarchical rules are extracted using the same phrase extraction procedure employed in phrase-based translation (Koehn et al. 2003) based on word alignments, followed by a step that performs a recursive extraction of hierarchical phrases (Chiang 2007).

For example, hierarchical rules could take the form of

$$X \rightarrow \left\langle \text{Officers},\ \text{主席団 成員} \right\rangle, \tag{3}$$

$$X \rightarrow \left\langle \text{the Committee},\ \text{委員会} \right\rangle, \tag{4}$$

$$X \rightarrow \left\langle X_0 \text{ of } X_1,\ X_1 \text{ 的 } X_0 \right\rangle. \tag{5}$$

From these rules, the input sentence can be translated by the derivation:

$$\begin{aligned} S \quad &\rightarrow \quad \langle X_0,\ X_0 \rangle \\ &\Rightarrow \quad \left\langle X_1 \text{ of } X_2,\ X_2 \text{ 的 } X_1 \right\rangle \\ &\Rightarrow \quad \left\langle \text{Officers of } X_2,\ X_2 \text{ 的 主席団 成員} \right\rangle \\ &\Rightarrow \quad \langle \text{Officers of the Committee}, \\ &\qquad\qquad \text{委員会 的 主席団 成員} \rangle. \end{aligned}$$

The advantage of Hiero is that it is able to achieve relatively superior word reordering accuracy (compared to other symbolic SMT alternatives such as standard phrase-based MT) without language-dependent processing. On the other hand, since it does not use syntactic information and tries to extract all possible combinations of rules, Hiero tends to extract very large translation rule tables, and it is also likely to be less syntactically faithful in its derivations.

2.3 Explicitly Syntactic Rules

The use of synchronous context-free grammar or synchronous tree-substitution grammar (Graehl and Knight 2004) rules forms an alternative to the Hiero rules. These options explicitly take into account the syntax of the source-side (tree-to-string rules), the target-side (string-to-tree rules), or both (tree-to-tree rules). The tree-to-string (T2S) rules, for example, utilize parse

[2] It is also standard to include a glue rule $S \rightarrow \langle X_0,\ X_0 \rangle$, $S \rightarrow \langle S_0\ X_1,\ S_0\ X_1 \rangle$, $S \rightarrow \langle S_0\ X_1,\ X_1\ S_0 \rangle$ to fall back on when standard rules cannot result in a proper derivation.

trees on the source language side, and the head symbols of the synchronous rules are not limited to S or X, but instead use non-terminal symbols corresponding to the phrase structure tags of a given parse tree. Thus, T2S rules could take the form of

$$X_{\mathrm{NP}} \rightarrow \langle (\text{NP (NNS Officers)}),\ \text{主席団 成員} \rangle , \tag{6}$$

$$X_{\mathrm{NP}} \rightarrow \langle (\text{NP (DT the) (NNP Committee)}),\ \text{委員会} \rangle , \tag{7}$$

$$X_{\mathrm{PP}} \rightarrow \left\langle (\text{PP (IN of) } X_{\mathrm{NP},0}),\ X_0 \text{ 的} \right\rangle , \tag{8}$$

$$X_{\mathrm{NP}} \rightarrow \left\langle (\text{NP } X_{\mathrm{NP},0}\ X_{\mathrm{PP},1}),\ X_1\ X_0 \right\rangle . \tag{9}$$

Here, parse subtrees of the source language rules are set in the form of S-expressions. From these rules, the translation can be effected from the parse tree of the input sentence by the derivation:

$$\begin{array}{lcl} x_{\mathrm{ROOT}} & \rightarrow & \left\langle x_{\mathrm{NP},0},\ x_0 \right\rangle \\ & \Rightarrow & \left\langle (\text{NP } x_{\mathrm{NP},1}\ x_{\mathrm{PP},2}),\ x_2\ x_1 \right\rangle \\ & \Rightarrow & \left\langle (\text{NP (NP (NNS Officers) } x_{\mathrm{PP},2})),\ x_2 \text{ 主席団 成員} \right\rangle \\ & \overset{*}{\Rightarrow} & \left\langle \begin{array}{l} \text{(NP} \\ \quad \text{(NP (NNS Officers))} \\ \quad \text{(PP (IN of)} \\ \qquad \text{(NP (DT the)} \\ \qquad\quad \text{(NNP Committee))))} \end{array},\ \text{委員会 的 主席団 成員} \right\rangle \end{array}$$

It is hence possible in T2S translation to obtain a result that conforms to the grammar of the source language. Also, as an advantage of this method, the number of less-useful synchronous rules extracted by syntax-agnostic methods such as Hiero is reduced. This decrease makes it possible to learn more compact rule tables and allows for faster translation.

3 Pivot Translation Methods

Several methods for SMT using pivot languages have been proposed, including *cascade* methods that consecutively translate from source to pivot then pivot to target (de Gispert and Mariño 2006), *synthetic* methods that machine-translate the training data to generate a pseudo-parallel corpus (de Gispert and Mariño 2006), and *triangulation* methods that obtain a source–target phrase/rule table by merging source–pivot and pivot–target table entries with identical pivot language phrases (Cohn and Lapata 2007). The triangulation method is particularly notable for producing higher quality translation results in comparison to other pivot methods (Utiyama and Isahara 2007; Miura et al. 2015), and this approach has thus been employed as the grounding for the work presented in this paper.

3.1 Triangulation of TMs

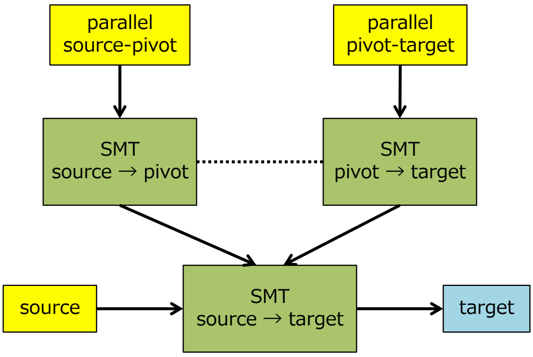


Figure 2 Triangulation of translation models

SCFG training procedure stores the extracted and scored phrase pairs from a bilingual corpus into a structured file called the *rule table*. Figure 2 presents a diagram of the *triangulation* of the source–pivot and pivot–target rule tables into the source–target table.

Triangulation for SCFGs searches T_{SP} and T_{PT} for source–pivot and pivot–target rules that have common pivot symbols and synthesizes these into source–target rules to create rule table T_{ST}:

$$\begin{aligned} &X \rightarrow \left\langle \overline{s},\ \overline{t} \right\rangle \in T_{ST} \\ &s.t.\ X \rightarrow \left\langle \overline{s}, \overline{p} \right\rangle \in T_{SP}\ \wedge\ X \rightarrow \left\langle \overline{p}, \overline{t} \right\rangle \in T_{PT}. \end{aligned} \tag{10}$$

Phrase translation probability $\phi(\cdot)$ and lexical translation probability $\phi_{lex}(\cdot)$ are estimated for all combined source–target phrases according to:

$$\phi\left(\overline{t}|\overline{s}\right) = \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi\left(\overline{t}|\overline{p}, \overline{s}\right) \phi\left(\overline{p}|\overline{s}\right) \approx \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi\left(\overline{t}|\overline{p}\right) \phi\left(\overline{p}|\overline{s}\right), \tag{11}$$

$$\phi\left(\overline{s}|\overline{t}\right) = \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi\left(\overline{s}|\overline{p}, \overline{t}\right) \phi\left(\overline{p}|\overline{t}\right) \approx \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi\left(\overline{s}|\overline{p}\right) \phi\left(\overline{p}|\overline{t}\right), \tag{12}$$

$$\phi_{lex}\left(\overline{t}|\overline{s}\right) = \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi_{lex}\left(\overline{t}|\overline{p}, \overline{s}\right) \phi_{lex}\left(\overline{p}|\overline{s}\right) \approx \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi_{lex}\left(\overline{t}|\overline{p}\right) \phi_{lex}\left(\overline{p}|\overline{s}\right), \tag{13}$$

$$\phi_{lex}\left(\overline{s}|\overline{t}\right) = \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi_{lex}\left(\overline{s}|\overline{p}, \overline{t}\right) \phi_{lex}\left(\overline{p}|\overline{t}\right) \approx \sum_{\overline{p} \in T_{SP} \cap T_{PT}} \phi_{lex}\left(\overline{s}|\overline{p}\right) \phi_{lex}\left(\overline{p}|\overline{t}\right), \tag{14}$$

where $\overline{s}$, $\overline{p}$, and $\overline{t}$ are the phrases in the source, pivot, and target, respectively, and the construction $\overline{p} \in T_{SP} \cap T_{PT}$ indicates that $\overline{p}$ is contained in both phrase tables T_{SP} and T_{PT}. Word penalty and phrase penalty $X \rightarrow \langle \overline{s},\ \overline{t} \rangle$ are set as the same values of $X \rightarrow \langle \overline{p},\ \overline{t} \rangle$.

3.2 Problems of Pivot-Side Ambiguity

Although triangulation is known to achieve higher translation accuracy than other simple methods and although it has become a popular and standard form of pivot translation nowadays, the problem of ambiguity still remains. This subsection describes the causes of the difficulties and provides pertinent examples.

In triangulation, Equations (11)–(14) are based on the memoryless channel model, which assumes

$$\phi\left(\overline{t}|\overline{p},\overline{s}\right) = \phi\left(\overline{t}|\overline{p}\right), \tag{15}$$

$$\phi\left(\overline{s}|\overline{p},\overline{t}\right) = \phi\left(\overline{s}|\overline{p}\right). \tag{16}$$

In Equation (15), for example, it is presumed that, given the pivot and source phrases, the translation probability of the target phrase is not affected by the source phrase. Nonetheless, it is easy to produce examples where this assumption does not hold.

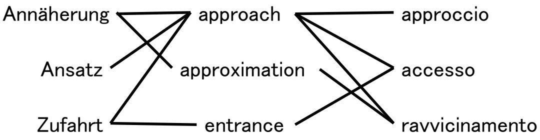


Figure 3 An example of ambiguity in De-En-It triangulation.

Figure 3 exemplifies three words in German and Italian. Each of the terms corresponds to the polysemic English word "approach." In such a case, it is extremely difficult to find associated source–target phrase pairs and to estimate the translation probabilities properly. As a result, pivot translation is significantly more ambiguous than standard translation.

The authors of this paper have previously proposed a pivot translation method that uses the triangulation of Synchronous CFG rule tables to a Multi-Synchronous CFG (MSCFG) (Neubig, Arthur, and Duh 2015) rule table that remembers the pivot as shown in Figure 4. This method performs the translation using pivot LMs (Miura et al. 2015), and experiments have established

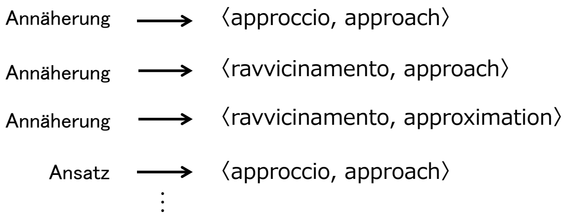

Annäherung ⟶ ⟨approccio, approach⟩
Annäherung ⟶ ⟨ravvicinamento, approach⟩
Annäherung ⟶ ⟨ravvicinamento, approximation⟩
Ansatz ⟶ ⟨approccio, approach⟩
⋮

Figure 4 An example of the triangulation method remembering the pivot.

that the process is effective in cases when a strong pivot LM exists.

This previously conceived method is effective in instances where the existing source–pivot and pivot–target parallel corpora are not large (i.e., containing less than hundreds of thousands of sentence pairs) and, conversely, where the available pivot monolingual corpus is sizable. Since the MSCFG decoder demands an immense quantity of memory and computational time and it is difficult to accomplish distributed processing, it is not realistic to scale up within the same framework. Further, although the MSCFG decoder helps in selecting the appropriate translation rules in a pivot LM, it cannot essentially solve the problem of ambiguity and of inappropriately connected rules, which remain in the triangulated rule table as noise. This paper attempts to elucidate the manner in which the noisy rules in triangulated TMs can be reduced to bring them closer to the accuracy attained by directly trained TMs.

4 Triangulation with Syntactic Matching

The previous section outlined the standard triangulation method and marked that the pivot-side ambiguity causes an incorrect estimation of translation probability and that the translation accuracy might decrease for this reason. To address this problem, it is desirable to be able to distinguish pivot-side phrases that have different syntactic roles or meanings, even if the symbol strings are equivalent. The next two subsections describe two methods of discerning pivot phrases that take on syntactically discrete roles: the first technique involves exact matching of parse trees; the second pertains to soft matching.

4.1 Exact Matching of Parse Subtrees

In the exact matching method, the pivot–source and pivot–target T2S TMs are first trained by parsing the pivot side of parallel corpora. Next, these data are stored into rule tables as T_{PS} and T_{PT}, respectively. The synchronous rules of T_{PS} and T_{PT} correspondingly take the form of $X \rightarrow \langle \hat{p}, \overline{s} \rangle$ and $X \rightarrow \langle \hat{p}, \overline{t} \rangle$, where $\hat{p}$ is a symbol string that expresses the pivot-side parse subtree (S-expression), and $\overline{s}$ and $\overline{t}$ express the source and target symbol strings in that order. The procedure of synthesizing source–target synchronous rules essentially follows Equations (11)–(14), except that it utilizes T_{PS} instead of T_{SP} (the direction of probability features is reversed) and the pivot subtree $\hat{p}$ instead of pivot phrase $\overline{p}$. In this case, $\overline{s}$ and $\overline{t}$ do not have syntactic information, and thus, the synthesized synchronous rules should be hierarchical rules as explained in Section 2.2.

The matching conditions of this method are more stringent in their constraints than the correspondence of superficial symbols in standard triangulation and thus potentially lessen incorrect connections of phrase pairs, resulting in a more reliable triangulated TM. Conversely, the number of connected rules decreases as well in this restricted triangulation, and hence, the coverage of the triangulated model might be reduced. Therefore, it is important to create TMs that are both reliable and that comprise superior coverage.

4.2 Partial Matching of Parse Subtrees

To prevent the problem of the reduction of coverage in the exact matching method, the authors of this paper propose a partial matching method that retains the coverage of standard triangulation by allowing the connection of incompletely equivalent pivot subtrees. To estimate translation probabilities in partial matching, the *weighted triangulation* generalizing Equations (11)–(14) of standard triangulation with the weight function $\psi(\cdot)$ must first be defined as in

$$\phi\left(\overline{t}|\overline{s}\right) = \sum_{\hat{p_T}} \sum_{\hat{p_S}} \phi\left(\overline{t}|\hat{p_T}\right) \psi\left(\hat{p_T}|\hat{p_S}\right) \phi\left(\hat{p_S}|\overline{s}\right), \tag{17}$$

$$\phi\left(\overline{s}|\overline{t}\right) = \sum_{\hat{p_S}} \sum_{\hat{p_T}} \phi\left(\overline{s}|\hat{p_S}\right) \psi\left(\hat{p_S}|\hat{p_T}\right) \phi\left(\hat{p_T}|\overline{t}\right), \tag{18}$$

$$\phi_{lex}\left(\overline{t}|\overline{s}\right) = \sum_{\hat{p_T}} \sum_{\hat{p_S}} \phi_{lex}\left(\overline{t}|\hat{p_T}\right) \psi\left(\hat{p_T}|\hat{p_S}\right) \phi_{lex}\left(\hat{p_S}|\overline{s}\right), \tag{19}$$

$$\phi_{lex}\left(\overline{s}|\overline{t}\right) = \sum_{\hat{p_S}} \sum_{\hat{p_T}} \phi_{lex}\left(\overline{s}|\hat{p_S}\right) \psi\left(\hat{p_S}|\hat{p_T}\right) \phi_{lex}\left(\hat{p_T}|\overline{t}\right), \tag{20}$$

where $\hat{p_S} \in T_{SP}$ and $\hat{p_T} \in T_{PT}$ are, respectively, the pivot parse subtrees of source–pivot and pivot–target synchronous rules. By adjusting $\psi(\cdot)$, the magnitude of the penalty for instances of incompletely matched connections can be controlled. If it is defined that $\psi(\hat{p_T}|\hat{p_S}) = 1$ when $\hat{p_T}$ is equal to $\hat{p_S}$ and that otherwise $\psi(\hat{p_T}|\hat{p_S}) = 0$, Equations (17)–(20) are equivalent to Equations (11)–(14).

The better estimation of $\psi(\cdot)$ is not trivial, and the co-occurrence counts of $\hat{p_S}$ and $\hat{p_T}$ are not available. Therefore, a heuristic estimation method is introduced as

$$\psi(\hat{p_T}|\hat{p_S}) = \frac{w(\hat{p_S}, \hat{p_T})}{\sum_{\hat{p} \in T_{PT}} w(\hat{p_S}, \hat{p})} \cdot \max_{\hat{p} \in T_{PT}} w(\hat{p_S}, \hat{p}), \tag{21}$$

$$\psi(\hat{p_S}|\hat{p_T}) = \frac{w(\hat{p_S}, \hat{p_T})}{\sum_{\hat{p} \in T_{SP}} w(\hat{p}, \hat{p_T})} \cdot \max_{\hat{p} \in T_{SP}} w(\hat{p}, \hat{p_T}), \tag{22}$$

$$w(\hat{p_S}, \hat{p_T}) = \begin{cases} 0 & (flat(\hat{p_S}) \neq flat(\hat{p_T})) \\ \exp\left(-d\left(\hat{p_S}, \hat{p_T}\right)\right) & (otherwise) \end{cases}, \tag{23}$$

$$d(\hat{p_S}, \hat{p_T}) = TreeEditDistance(\hat{p_S}, \hat{p_T}), \tag{24}$$

where $flat(\hat{p})$ returns only the sequence of leaf elements, or the symbol string of $\hat{p}$ keeping non-terminals,[3] and $TreeEditDistance(\hat{p_S}, \hat{p_T})$ is the minimum cost of a sequence of operations (contract, un-contract, and modify the label of an edge) that are required to transform $\hat{p_S}$ into $\hat{p_T}$ (Klein 1998).

According to Equations (21)–(24), it may be assured that the incomplete match of pivot subtrees leads to $d(\cdot) \geq 1$ and penalizes in a manner that $\psi(\cdot) \leq 1/e^d \leq 1/e$, and an exact match of subtrees causes a value of $\psi(\cdot)$ that is at least $e \approx 2.718$ times larger than those obtained by the utilization of partially matched subtrees.

5 Experiments

5.1 Experimental Set-Up

5.1.1 Evaluation of Pivot SMT methods

To investigate the effect of the proposed approach, the authors evaluated translation accuracy through pivot translation experiments conducted on the United Nations Parallel Corpus (UN6Way) (Ziemski et al. 2016). UN6Way is a line-aligned multilingual parallel corpus that includes data in English (En), Arabic (Ar), Spanish (Es), French (Fr), Russian (Ru) and Chinese (Zh), and accounts for different families of languages. The corpus contains over 11M sentences for each language pair and is therefore deemed suitable for multilingual translation tasks such as pivot translation. English was fixed as the pivot language for the experiments reported in this paper because it is the language most frequently employed for this function. Its utilization as the

[3] For example, given $\hat{p} = (NP(NP(NNS\text{Officers}))(PP(IN\text{of})(NP(DT\text{the})(NNP\text{Committee}))))$, then, $flat(\hat{p})$ = "Officers of the Committee".

pivot language thus yields the positive side-effect of the readily available accurate phrase structure parsers, benefits the proposed method. Pivot translation was performed on all combinations of the other five languages, and the accuracy of each method was compared. For tokenization, SentencePiece,[4] an unsupervised text tokenizer and detokenizer, was adopted. Although it is designed primarily for use in neural MT, it was confirmed that SentencePiece also helps in reducing training time and that it also improves translation accuracy in the authors' previously posited Hiero model. A single shared tokenization model was first trained by feeding a total of 10M sentences from the data of all six languages, setting the maximum shared vocabulary size to be 16k, and all available text was tokenized with the trained model. English raw text was used without SentencePiece tokenization for phrase structure analysis and for training Hiero and T2S TMs on the pivot side. To generate parse trees, the Ckylark PCFG-LA parser (Oda, Neubig, Sakti, Toda, and Nakamura 2015) was utilized, and lines over 60 tokens in length were filtered out from all parallel data to ensure the accuracy of parsing and alignment. Once the sorting was accomplished, about 7.6M lines remained. Since Hiero requires a large quantity of computational resources for training and decoding, the decision was taken to use only the first 1M lines to train each TM, instead of the entirety of the available training data.[5] Travatar (Neubig 2013) was employed as a decoder. Hiero and T2S TMs were utilized and were trained with Travatar's rule extraction code. 5-gram LMs were trained over the target side of the same parallel data utilized for training TMs using KenLM (Heafield 2011). The first 1,000 lines of the 4,000 lines of test and dev sets were used for testing and parameter tuning, respectively. For the evaluation of translation results, the text was de-tokenized with SentencePiece and re-tokenized with the tokenizer from the Moses toolkit (Koehn, Hoang, Birch, Callison-Burch, Federico, Bertoldi, Cowan, Shen, Moran, Zens, Dyer, Bojar, Constantin, and Herbst 2007) for Arabic, Spanish, French and Russian. The Chinese text was re-tokenized with KyTea tokenizer (Neubig, Nakata, and Mori 2011) and was then evaluated with the utilization of case-sensitive BLEU-4 (Papineni, Roukos, Ward, and Zhu 2002), RIBES (Isozaki, Hirao, Duh, Sudoh, and Tsukada 2010) and NIST (Doddington 2002).

Five translation methods were assessed:

Cascade Hiero:

[4] https://github.com/google/sentencepiece

[5] The authors use the same 1M pivot-side sentences in source-pivot and pivot-target parallel data, even though it is not a realistic situation in pivot translation. This setting is useful to investigate how pivot translation tasks degrade accuracy from the ideal condition by comparing with scores of Direct methods. The previous work of the authors (Miura et al. 2015; 三浦，Neubig，Sakti，戸田，中村 2016a) has not found large differences in translation accuracy between situations of using the same pivot sentences and no common pivot sentences.

vocabulary size:	16k (shared)
source embedding size:	512
target embedding size:	512
output embedding size:	512
encoder hidden size:	512
decoder hidden size:	512
LSTM layers:	1
attention type:	MLP
attention hidden size:	512
optimizer type:	Adam
loss integration type:	mean
batch size:	2048
max iteration:	200k
dropout rate:	0.3
decoder type:	Luong+ 2015

Table 1 Main parameters of NMT training

Sequential pivot translation with source–pivot and pivot–target Hiero TMs (weak baseline).

Tri. Hiero:

Triangulating source–pivot and pivot–target Hiero TMs into a source–target Hiero TM using the traditional method (baseline, Section 3.1).

Tri. TreeExact

Triangulating pivot–source and pivot–target T2S TMs into a source–target Hiero TM using the proposed exact matching of pivot subtrees (proposed 1, Section 4.1).

Tri. TreePartial

Triangulating pivot–source and pivot–target T2S TMs into a source–target Hiero TM using the proposed partial matching of pivot subtrees (proposed 2, Section 4.2).

Direct Hiero:

Translating with a Hiero TM directly trained on the source-target parallel corpus without using a pivot language (as an oracle).

5.1.2 Comparison with Neural MT

Recent investigations (Firat, Sankaran, Al-Onaizan, Yarman Vural, and Cho 2016; Johnson, Schuster, Le, Krikun, Wu, Chen, Thorat, Viégas, Wattenberg, Corrado, Hughes, and Dean 2017) have found that neural machine translation systems can gain the ability to perform translations with zero parallel resources by training on multiple sets of bilingual data. However, previous work has not examined the competitiveness of these methods in comparison to pivot-based symbolic SMT frameworks such as PBMT or Hiero. In this section, a zero-shot NMT model and other pivot NMT methods are compared to the pivot-based Hiero models. The NMTKit [6] was adopted to train and evaluate NMT models. The detailed parameters of training the NMT models are shown in Table 1.

For additional translation methods were assessed:

Cascade NMT:

Sequential pivot translation with source-pivot and pivot-target NMTs.

Synthetic NMT:

Generating pseudo-parallel corpus synthesized by translating pivot-side of source-pivot parallel corpus with pivot-target NMT (Sennrich, Haddow, and Birch 2016; Firat et al. 2016).

Zero-Shot NMT:

Training single shared model with pvt $\leftrightarrow$ $\{$src,target$\}$ parallel data according to (Johnson et al. 2017).

Direct NMT:

Translating with NMT directly trained on the source-target parallel corpus without using pivot language (for comparison).

The training data for Cascade NMT, Synthetic NMT, and Zero-Shot NMT were the same 1M sentences used for the Pivot Hiero methods (source–pivot and pivot–target corpora). The training data for Direct NMT was identical to that utilized for Direct Hiero.

5.2 Results

Comparison of Direct models among different language pairs and frameworks: Before pivot translation tasks were compared, the performances of SMT and NMT in Direct translation tasks were ascertained. Table 2 demonstrates the BLEU-4, RIBES, and NIST scores for each Direct translation task and language pair. The results confirm the tendency that directly trained NMT models achieve high translation accuracy even in the case of translation between languages of different families On the other hand, these scores are drastically reduced in situations that do not offer source–target parallel corpora for training.

Since the parameters were optimized for the BLEU score and not for RIBES and NIST,

[6] https://github.com/odashi/nmtkit

Source	Target	BLEU [%] / RIBES [%] / NIST Score	
		Direct Hiero	Direct NMT
Ar	*En*	40.00 / **83.98** / **7.838**	**40.16** / 83.33 / 7.710
Es	*En*	49.50 / 87.41 / **9.208**	**50.51** / **87.91** / 9.130
Fr	*En*	42.13 / 83.91 / **8.321**	**43.12** / **84.70** / 8.220
Ru	*En*	**41.15** / **83.75** / **8.026**	40.88 / 82.99 / 7.831
Zh	*En*	34.19 / 78.31 / **7.368**	**35.50** / **80.64** / 7.249
En	Ar	28.18 / 80.11 / 6.717	**30.58** / **80.84** / **6.793**
	Es	49.70 / 86.91 / **9.201**	**50.61** / **88.57** / 9.187
	Fr	40.57 / 82.40 / 8.033	**41.56** / **84.24** / **8.112**
	Ru	31.63 / 79.61 / 6.777	**34.76** / **80.68** / **6.979**
	Zh	33.07 / 80.89 / 8.170	**38.05** / **83.78** / **8.176**
Ar	Es	**38.49** / 82.85 / **7.442**	38.25 / **83.09** / 7.288
	Fr	**33.34** / **79.97** / **6.828**	33.16 / 78.80 / 6.641
	Ru	24.63 / 75.55 / **5.813**	**27.00** / **75.87** / 5.745
	Zh	27.27 / 76.31 / **6.827**	**30.04** / **79.48** / 6.771
Es	Ar	**27.18** / **79.19** / **6.350**	26.02 / 78.73 / 6.184
	Fr	**43.24** / **85.60** / **8.240**	41.83 / 83.82 / 7.924
	Ru	28.83 / 77.84 / **6.434**	**30.65** / **78.70** / 6.429
	Zh	27.08 / 75.29 / 7.037	**32.36** / **80.85** / **7.320**
Fr	Ar	**25.10** / **77.51** / **5.854**	23.28 / 76.29 / 5.732
	Es	**45.20** / **86.48** / **8.317**	44.49 / 85.35 / 8.294
	Ru	27.42 / **77.11** / **6.016**	**28.29** / 75.80 / 5.963
	Zh	25.84 / 74.55 / 6.619	**29.10** / **78.78** / **6.833**
Ru	Ar	22.53 / 76.03 / **5.722**	**23.19** / **76.32** / 5.636
	Es	37.60 / 82.04 / **7.496**	**38.67** / **82.09** / 7.409
	Fr	**34.05** / **79.88** / **6.945**	33.26 / 78.57 / 6.764
	Zh	28.03 / 76.31 / **7.083**	**31.39** / **79.13** / 6.993
Zh	Ar	**20.09** / 70.59 / **5.382**	17.73 / **73.59** / 5.210
	Es	**30.66** / 74.43 / **6.580**	28.05 / **78.33** / 6.502
	Fr	**25.97** / 71.87 / **6.012**	24.35 / **74.76** / 5.954
	Ru	**21.16** / 69.27 / **5.280**	19.59 / **72.44** / 5.218

Table 2 Comparison of SMT and NMT in multilingual Direct translation tasks. Bold face indicates higher evaluation score for each language-pair and measurement.

higher RIBES / NIST scores do not necessarily imply the superior performance of the MT framework. They merely demonstrate the side effects obtained by optimizing for BLEU. However, the comparison of RIBES and NIST scores shows the obvious propensities of SMT and NMT frameworks. In almost language pairs, Direct Hiero outperformed Direct NMT in NIST scores, even though its BLEU and RIBES scores were lower than NMT. It is known that NMT faces difficulties with rare words and that it tends to fail in the translation of vocabulary that is not frequently used. This discovery reflects the fact that NIST is recognized as a measurement tool that values the translation accuracy of content words that generally occur in lower frequencies in comparison to function words. Conversely, NMT is capable of acquiring higher naturalness and fluency in word sequences that include function words, since BLEU provides weightage to the

Source	Target	1-Gram / 2-Gram / 3-Gram / 4-Gram Precision / Brevity Penalty [%]	
		Direct Hiero	Direct NMT
Ar	*En*	67.03 / 44.98 / 33.28 / 25.52 / 100.0	65.88 / 45.34 / 33.82 / 25.76 / 100.0
Es	*En*	74.83 / 54.48 / 42.78 / 34.42 / 100.0	74.76 / 55.40 / 44.03 / 35.70 / 100.0
Fr	*En*	68.97 / 47.02 / 35.39 / 27.46 / 100.0	69.33 / 48.30 / 36.78 / 28.61 / 99.53
Ru	*En*	68.19 / 46.00 / 34.29 / 26.66 / 100.0	66.72 / 45.43 / 34.17 / 26.96 / 100.0
Zh	*En*	65.17 / 39.89 / 27.16 / 19.34 / 100.0	63.38 / 40.94 / 29.07 / 21.06 / 100.0
En	Ar	57.37 / 34.13 / 22.19 / 14.73 / 99.65	58.75 / 37.66 / 26.15 / 18.43 / 95.16
	Es	74.17 / 54.89 / 43.29 / 34.88 / 100.0	74.28 / 56.49 / 45.22 / 36.39 / 98.72
	Fr	65.89 / 45.46 / 34.20 / 26.44 / 100.0	67.27 / 47.71 / 36.80 / 28.95 / 96.45
	Ru	58.51 / 37.38 / 26.63 / 19.67 / 96.67	60.10 / 40.67 / 30.36 / 23.22 / 95.96
	Zh	71.36 / 42.22 / 27.23 / 18.31 / 94.45	70.33 / 45.45 / 31.57 / 22.78 / 97.73

Table 3 Components of BLEU score for English-related translation tasks.

Source	Target	BLEU Score [%]				
		Direct	Cascade (baseline)	Tri. Hiero (baseline)	Tri. TreeExact (proposed 1)	Tri. TreePartial (proposed 2)
Ar	Es	38.49	30.95	34.20	‡ 34.97	‡ **35.94**
	Fr	33.34	25.08	29.93	‡ 30.68	‡ **30.83**
	Ru	24.63	18.70	22.94	‡ 23.94	‡ **24.15**
	Zh	27.27	21.77	22.78	‡ **25.17**	‡ 25.07
Es	Ar	27.18	22.72	22.97	‡ 24.09	‡ **24.45**
	Fr	43.24	35.40	38.74	‡ 39.62	‡ **40.12**
	Ru	28.83	22.43	26.35	‡ 27.25	‡ **27.41**
	Zh	27.08	23.36	24.54	25.00	† **25.16**
Fr	Ar	25.10	19.88	21.65	21.40	† **22.13**
	Es	45.20	37.75	40.16	‡ 41.03	‡ **41.99**
	Ru	27.42	20.64	24.71	† 25.24	‡ **25.64**
	Zh	25.84	21.79	23.16	**23.56**	23.53
Ru	Ar	22.53	18.71	19.82	19.86	**20.35**
	Es	37.60	31.33	34.56	34.96	‡ **35.62**
	Fr	34.05	27.11	30.75	† 31.43	‡ **31.67**
	Zh	28.03	21.81	24.88	25.07	**25.12**
Zh	Ar	20.09	14.82	16.66	17.01	‡ **17.73**
	Es	30.66	23.15	27.84	27.99	**28.05**
	Fr	25.97	19.55	23.82	24.34	† **24.35**
	Ru	21.16	14.79	18.63	‡ 19.58	‡ **19.59**

Table 4 Comparison of each method. Bold face indicates the highest BLEU score in pivot translation, and daggers indicate statistically significant gains over Tri. Hiero ($\dagger : p < 0.05, \ddagger : p < 0.01$).

accuracy of word sequences (n-grams) and RIBES is known to give importance word order.

The table also demonstrates that the evaluation scores of language-pairs that do not contain English are much lower than those that include English. For example, the BLEU scores of Arabic–English, English–French and Arabic–French in Direct Hiero are 40.00, 40.57, and 33.34, respectively. It is thought that sentence pairs that do not include English are more noisy than those that contain English since the multilingual corpus was primarily constructed by sourcing

Source	Target	Number of source-side unique phrases/words			
		Tri. Hiero	Tri. TreePartial	Tri. TreeExact	*loss in coverage* [%]
Ar	Es	2.646M / 5,077	2.646M / 5,077	2.580M / 5,072	2.494 / 0.985
	Fr	2.658M / 5,071	2.658M / 5,071	2.589M / 5,067	2.596 / 0.079
	Ru	2.406M / 5,088	2.406M / 5,088	2.347M / 5,085	2.452 / 0.059
	Zh	2.386M / 5,040	2.386M / 5,040	2.324M / 5,034	1.844 / 0.119
Es	Ar	2.013M / 5,188	2.013M / 5,188	1.942M / 5,182	3.527 / 0.116
	Fr	2.129M / 5,210	2.129M / 5,210	2.062M / 5,205	3.147 / 0.096
	Ru	2.037M / 5,197	2.037M / 5,197	1.978M / 5,191	2.896 / 0.115
	Zh	1.986M / 5,180	1.986M / 5,180	1.920M / 5,175	3.323 / 0.097
Fr	Ar	2.233M / 5,316	2.233M / 5,316	2.176M / 5,310	2.553 / 0.113
	Es	2.366M / 5,342	2.366M / 5,342	2.302M / 5,337	2.705 / 0.094
	Ru	2.266M / 5,318	2.266M / 5,318	2.203M / 5,311	2.780 / 0.132
	Zh	2.215M / 5,321	2.215M / 5,321	2.162M / 5,313	2.393 / 0.150
Ru	Ar	2.505M / 5,644	2.505M / 5,644	2.437M / 5,637	2.715 / 0.124
	Es	2.536M / 5,682	2.536M / 5,682	2.478M / 5,677	2.287 / 0.088
	Fr	2.531M / 5,665	2.531M / 5,665	2.479M / 5,661	2.055 / 0.071
	Zh	2.515M / 5,688	2.515M / 5,688	2.466M / 5,682	1.948 / 0.105
Zh	Ar	1.556M / 9,474	1.556M / 9,474	1.480M / 9,428	4.884 / 0.486
	Es	1.570M / 9,555	1.570M / 9,555	1.504M / 9,523	4.200 / 0.335
	Fr	1.568M / 9,520	1.568M / 9,520	1.499M / 9,490	4.401 / 0.315
	Ru	1.593M / 9,487	1.593M / 9,487	1.518M / 9,457	4.708 / 0.316

Table 5 Comparison of rule table coverage in proposed triangulation methods.

from English documents as the pivot.

Performance of English-related translation tasks: Pivot translation tasks should depend strongly on the performance of the source–pivot translation and should rely even more compellingly on the pivot–target translation since the pivot–target translation essentially comprises the upper bound performance of generating target sentences for the given pivot-side input. It is natural that the translation for pairs of languages belonging to different families exhibits a different trend with regard to accuracy. Table 2 illustrates that the TMs of English–Spanish and English–French achieve higher evaluation scores, perhaps because they exhibit relatively closer language structures than the other evaluated English-relative language pairs. Although English–Arabic, English–Russian, and English–Chinese translation achieve poorer accuracy, each of these likely result from different language features, such as morphology, word order, and diversity of expression.

Table 3 illustrates the components of the BLEU score evaluation, including the precision of 1-grams through 4-grams and the brevity penalty (Papineni et al. 2002). This table demonstrates that English–Chinese translation achieved higher accuracy in translating words, or 1-gram precision than language pairs that comprises Arabic, French, and Russian targets. However, the

Source	Target	Noise Ratio in Triangulated Table [%]		
		Tri. Hiero	Tri. TreeExact	Tri. TreePartial
Ar	Es	78.40	**63.61** (-14.79)	68.51 (-9.88)
	Fr	81.39	**67.31** (-14.08)	72.22 (-9.17)
	Ru	81.87	**69.23** (-12.64)	73.84 (-8.03)
	Zh	75.70	**63.06** (-12.64)	67.72 (-7.98)
Es	Ar	80.03	**64.97** (-15.06)	69.80 (-10.23)
	Fr	81.55	**65.30** (-16.26)	70.61 (-10.94)
	Ru	81.45	**68.02** (-13.43)	72.67 (-8.78)
	Zh	74.05	**61.60** (-12.45)	65.90 (-8.15)
Fr	Ar	81.77	**67.69** (-14.08)	72.39 (-9.38)
	Es	80.94	**64.94** (-16.00)	70.20 (-10.74)
	Ru	82.77	**69.84** (-12.93)	74.52 (-8.25)
	Zh	76.14	**64.29** (-11.85)	68.53 (-7.61)
Ru	Ar	82.15	**70.15** (-12.00)	74.60 (-7.55)
	Es	79.80	**67.16** (-12.64)	71.68 (-8.12)
	Fr	82.41	**70.10** (-12.31)	74.76 (-7.65)
	Zh	76.07	**64.31** (-11.76)	68.67 (-7.40)
Zh	Ar	80.05	**66.90** (-13.15)	71.23 (-8.82)
	Es	77.94	**65.53** (-12.41)	69.70 (-8.24)
	Fr	78.24	**68.54** (-9.70)	72.51 (-5.73)
	Ru	79.80	**67.07** (-12.73)	71.27 (-8.53)

Table 6 Comparison of noise ratio in triangulated rule table

precision of 2-grams through 4-grams, or the accuracy of translating word sequences is relatively lower in English–Chinese and this low BLEU score is primarily caused by the low 4-gram precision. This result reflects the fact that word inflections do not exist in Chinese, and instead, the word order takes on significant syntactic roles.

Conversely, the table also clarifies that English–Arabic and English–Russian translation achieved lower precision even relating to 1-grams. This lack of accuracy could be caused by the fact that Arabic and Russian are known for their morphological richness, and it is thus more difficult for MT to translate the source words into correct forms of target words than in the case of other, more morphologically simple languages.

Translation accuracy of pivot translation methods: The results of the experiments that used all combinations of pivot translation tasks via English for five languages are shown in Table 4. These outcomes exhibit that the proposed partial matching method of pivot subtrees in triangulation outperformed the standard triangulation method for all language pairs and that it achieved higher or almost equal scores than the proposed exact matching method. The exact matching method also outperformed the standard triangulation method in the majority of the language pairs, but has a lesser improvement than the partial matching method. As demonstrated

Source	Target	Distribution Error Rate (MAE / RMSE) [%]		
		Tri. Hiero	Tri. TreeExact	Tri. TreePartial
Ar	Es	14.16 / 10.62	14.49 / 11.06	13.96 / 10.56
	Fr	13.01 / 9.72	13.52 / 10.19	12.90 / 9.65
	Ru	12.64 / 9.51	12.33 / 9.24	12.03 / 8.97
	Zh	15.88 / 11.96	13.69 / 10.42	13.81 / 10.42
Es	Ar	13.90 / 10.29	13.84 / 10.30	13.44 / 9.92
	Fr	13.39 / 10.61	14.51 / 11.30	13.95 / 10.89
	Ru	12.81 / 9.71	12.92 / 9.70	12.52 / 9.38
	Zh	16.02 / 12.09	13.94 / 10.69	14.01 / 10.64
Fr	Ar	13.40 / 9.98	13.10 / 9.76	12.70 / 9.38
	Es	14.25 / 11.38	14.39 / 11.29	14.05 / 11.03
	Ru	12.58 / 9.58	12.46 / 9.37	11.98 / 8.99
	Zh	15.45 / 11.74	13.34 / 10.28	13.40 / 10.23
Ru	Ar	12.68 / 9.35	12.36 / 9.16	11.98 / 8.79
	Es	13.27 / 10.05	13.68 / 10.54	13.12 / 10.00
	Fr	12.29 / 9.28	12.84 / 9.78	12.13 / 9.17
	Zh	15.34 / 11.72	13.13 / 10.10	13.25 / 10.11
Zh	Ar	12.57 / 9.11	12.86 / 9.39	12.57 / 9.09
	Es	13.25 / 9.78	13.58 / 10.16	13.22 / 9.79
	Fr	12.86 / 9.49	12.67 / 9.44	12.25 / 9.07
	Ru	12.22 / 9.14	12.40 / 9.31	12.12 / 9.03

Table 7 Comparison of distribution error rate in triangulated rule table

by the authors' previous research undertaking, the sequential pivot translation was uniformly weaker than all triangulation methods.

Effect on coverage: Table 5 presents the outcomes of the comparison of the coverage achieved by each proposed triangulation method. This table confirms that Tri.TreeExact reduced the number of unique phrases by several percentage points and Tri.TreePartial kept the same coverage as Tri.Hiero. Especially, triangulated TMs from Chinese with exact matching contain substantially fewer source phrases and significantly source words, and harmed coverage up to 4.884% as Chinese contains many characters and values the reordering of short tokens instead of inflections. This anomaly could constitute one of the reasons for the difference in improvement stability with regard to the partial and exact matching methods.

Noise reduction: The main motivation of using parse trees in the proposed methods is to prevent the inappropriate connection of phrase correspondences and reduce the noise in rule tables. To investigate the manner in which the syntactic matching methods succeed in removing noisy rules, an analysis of noise ratio was conducted. Noisy rules must contain source and target phrases having no correspondence in meaning, though this decision cannot be made for all phrase pair candidates in rule tables. It was therefore assumed that directly trained TMs that could avail of a source–target parallel corpus would demonstrate a fine approximation close to the ideal distribution of translation probability. To compute the noise ratio $noise(T_{tri}|T_{dir})$ of triangulated rule table T_{tri} with directly trained table T_{dir}

$$noise\left(T_{tri}|T_{dir}\right) = \frac{\sum_{\left(\bar{s},\bar{t}\right)\in T_{tri}\setminus T_{dir}} \phi\left(\bar{t}|\bar{s}\right)}{\sum_{\left(\bar{s},\bar{t}\right)\in T_{tri}} \phi\left(\bar{t}|\bar{s}\right)} \tag{25}$$

was defined, where $\phi(\bar{t}|\bar{s})$ represents the forward translation probability that can be considered the most important feature of the rule table. Table 6 displays the calculated noise ratio of the rule table for each triangulation method and language pair. This result shows that, although triangulated rule tables contain many noisy rules, the syntactic matching methods are indeed successful in reducing them. Tri.TreeExact decreased noisy rules, up to -16.26%, and Tri.TreePartial lessened noisy rules up to -10.94%. The reason why the noise reduction rate of Tri.TreePartial is lower than that of Tri.TreeExact is that the former weakens the influence of noisy rules instead of removing them to retain the coverage.

Improvement of probability estimation: Although syntactic matching methods help in reducing noisy rules, there is no guarantee that they can improve the estimation of translation probabilities. Table 7 exhibits the mean absolute error (MAE) and the root-mean-square error (RMSE) for the distribution of forward translation probability scores of triangulated rule tables in comparison to directly trained rule tables. To calculate MAE and RMSE, the noisy rules that were not contained in directly trained rule tables were ignored to separate the different factors. The results evince that Tri.TreePartial reduced MAE and RMSE, making the distribution closer to the ideal in almost all language pairs. On the other hand, Tri.TreeExact did not diminish the errors in a stable manner. This consequence may be induced by the fact that the restricted matching conditions of Tri.TreeExact exclude many unmatched phrase pair candidates and may remove even those translation rules that are not noisy. It may therefore be posited that the softening of restrictions pertaining to matching conditions aids in the improvement of the estimation of translation probabilities.

Comparison with NMT:

Table 8 presents the BLEU score of each translation task and language pair. Perhaps, Synthetic NMT outperformed Cascade NMT for the majority of language pairs because multi-layer NNs are robust for noisy training data and can optimize the trained model with fine-tuning tech-

Source	Target	BLEU Score [%]						
		Direct Hiero	*Direct NMT*	Tri. TreePartial	Cascade Hiero	Cascade NMT	Synthetic NMT	Zero-Shot NMT
Ar	Es	38.49	38.25	35.94	30.95	31.62	32.35	8.18
	Fr	33.34	33.16	30.83	25.08	26.91	29.51	8.57
	Ru	24.63	27.00	24.15	18.70	21.67	21.81	5.79
	Zh	27.27	30.04	25.07	21.77	23.70	25.63	5.04
Es	Ar	27.18	26.02	24.45	22.72	21.21	23.01	5.22
	Fr	43.24	41.83	40.12	35.40	31.84	36.57	15.04
	Ru	28.83	30.65	27.41	22.43	23.60	25.97	7.57
	Zh	27.08	32.36	25.16	23.36	26.03	27.31	8.62
Fr	Ar	25.10	23.28	22.13	19.88	18.66	18.83	8.08
	Es	45.20	44.49	41.99	37.75	32.93	36.78	14.37
	Ru	27.42	28.29	25.64	20.64	20.87	23.60	8.77
	Zh	25.84	29.10	23.53	21.79	23.14	24.96	11.95
Ru	Ar	22.53	23.19	20.35	18.71	19.71	19.21	3.18
	Es	37.60	38.67	35.62	31.33	31.25	31.22	10.42
	Fr	34.05	33.26	31.67	27.11	27.34	29.10	9.76
	Zh	28.03	31.39	25.12	21.81	24.25	25.46	9.46
Zh	Ar	20.09	20.17	17.73	14.82	16.89	18.01	10.38
	Es	30.66	32.69	28.05	23.15	26.01	27.80	6.13
	Fr	25.97	27.68	24.35	19.55	23.35	25.46	7.12
	Ru	21.16	23.17	19.59	14.79	18.40	20.53	3.21

Table 8 Comparison of SMT and NMT in pivot translation tasks.

niques. On the other hand, the fine-tuning is available for Cascade NMT only separately for the source–pivot and pivot–target TMs and not for the whole pipelined system.

In the setting of the current experiments, although bilingually trained NMT systems were competitive to or outperformed Hiero-based models, the zero-shot translation was uniformly weaker. This outcome could be the result of using only a single LSTM layer for each encoder and decoder or because there was an insufficient quantity of parallel corpora or language pairs. It may therefore be posited that, although zero-shot translation demonstrated reasonable results in some settings, successful zero-shot translation systems are difficult to build, and pivot-based symbolic MT systems such as PBMT or Hiero might still be competitive alternatives.

Qualitative analysis: A translated sentence for which pivot-side ambiguity is resolved in the syntactic matching methods is presented as an example:

Source Sentence in French:

La Suisse encourage **tous les États** **parties** à soutenir le travail conceptuel que fait actuellement le Secrétariat .

Corresponding Sentence in English:

Switzerland encourages all parties to support the current conceptual work of the secretariat.

Reference in Spanish:

Suiza alienta a **todos los Estados partes** a que apoyen la actual labor *conceptual* de la Secretaría .

Direct Hiero:

Suiza alienta a todos los Estados partes a que apoyen el trabajo conceptual que se examinan en la Secretaría . (BLEU+1: 55.99, RIBES: 91.47, NIST: 9.687)

Tri. Hiero:

Suiza *conceptuales* para apoyar la labor que en estos momentos la Secretaría alienta a todos los Estados Partes . (BLEU+1: 29.74, RIBES: 53.92, NIST: 6.204)

Tri. TreeExact:

Suiza alienta a **todos los Estados Partes** a apoyar la labor conceptual que actualmente la Secretaría . (BLEU+1: 43.08, RIBES: 88.36, NIST: 8.625)

Tri. TreePartial:

Suiza alienta a **todos los Estados Partes** a apoyar la labor conceptual que actualmente la Secretaría . (same with Tri. TreeExact)

Direct NMT:

Suiza alienta a todos los Estados partes a que apoyen la labor conceptual que está realizando la Secretaría . (BLEU+1: 66.94, RIBES: 95.79, NIST: 11.22)

Synthetic NMT:

Suiza alentó a todos los Estados Partes a que apoyen el trabajo conceptual que actualmente está a la Secretaría . (BLEU+1: 28.94, RIBES: 88.01, NIST: 7.267)

The results of Tri.TreeExact and Tri.TreePartial are identical in this example. The digest of the derivation process in Tri. Hiero is

$$\begin{aligned} S &\Rightarrow \langle X_0,\ X_0 \rangle \\ &\Rightarrow \langle \textit{La Suisse } X_1,\ \textit{Suiza } X_1 \rangle \\ &\Rightarrow \langle \textit{La Suisse } X_2 \ .,\ \textit{Suiza } X_2 \ . \rangle \\ &\Rightarrow \left\langle \textit{La Suisse } \underline{X_3} \textit{ partie } \utilde{X_4} \ .,\ \textit{Suiza } \utilde{X_4}\ \underline{X_3} \textit{ Partes } . \right\rangle \\ &\Rightarrow \ldots \end{aligned}$$

On the other hand, the digest of the derivation process in Tri.TreeExact/Tri.TreePartial is

$S \Rightarrow \langle X_0,\ X_0 \rangle$
$\Rightarrow \langle \textit{La Suisse } X_1,\ \textit{Suiza } X_1 \rangle$
$\Rightarrow \langle \textit{La Suisse encourage } X_2\ X_3,\ \textit{Suiza alienta a } X_2\ X_3 \rangle$
$\Rightarrow \langle \textit{La Suisse encourage tous les } X_4 \textit{ partie } X_3,\ \textit{Suiza alienta a todos } X_4 \textit{ Partes } X_3 \rangle$
$\Rightarrow \langle \textit{La Suisse encourage tous les Etats partie } X_3,\ \textit{Suiza alienta a todos los Estados Partes } X_3 \rangle$
$\Rightarrow \ldots$

Here, it is observed that the derivation in Tri.Hiero uses rule $X \rightarrow \langle X_0 \text{ parties } X_1, X_1\ X_0 \text{ Partes} \rangle$[7] which causes the incorrect reordering of phrases, followed by steps of incorrect word selection.[8] On the other hand, the derivation in Tri.TreeExact and Tri.TreePartial uses rule $X \rightarrow \langle \text{tous les } X_0 \text{ parties}, \text{ todos } X_0 \text{ Partes} \rangle$[9] as synthesized from the T2S rules with the common pivot subtree

(NP (DT all) (NP' X_{NNP}(NNS parties)). It can thus be proved that the derivation improves the word-selection and word-reordering using this rule.

The following example presents a translated sentence for which the exact matching method loses the necessary translation rule and instead degrades accuracy.

Source Sentence in Chinese

秘书长 关于 “ 人力 资源 管理 改革 概览 : 流 动性 ” 的 报告

Corresponding Sentence in English:

Report of the Secretary-General on the overview of human resources management reform : mobility

Reference in Russian

Доклад Генерального секретаря об обзоре хода реформы системы управления людскими ресурсами : мобильность

Direct Hiero:

Доклад Генерального секретаря об реформе управления людскими ресурсами : обзор мобильность » (BLEU+1: 45.47, RIBES: 92.29, NIST: 9.143)

[7] The words emphasized with underline and wavy-underline in the example correspond to X_0 and X_1 respectively.

[8] For example, the word “conceptuales” with italic face in Tri.Hiero takes the wrong form and position.

[9] The words emphasized in bold face in the example correspond to the rule.

Tri. Hiero:

Доклад Генерального секретаря о реформе управления людскими ресурсами обзор : Мобильность » (BLEU+1: 26.25, RIBES: 86.67, NIST: 5.962)

Tri. TreeExact:

Доклад Генерального секретаря мобильности обзор реформе управления людскими ресурсами » : ". (BLEU+1: 26.19, RIBES: 85.66, NIST: 5.652)

Tri. TreePartial:

Доклад Генерального секретаря о Обзор реформе системы управления людскими ресурсами : мобильность ". (BLEU+1: 48.76, RIBES: 89.54, NIST: 8.107)

Direct NMT

Доклад Генерального секретаря о “ Обзоре реформы управления людскими ресурсами : мобильность ” (BLEU: 45.37, RIBES: 91.22, NIST: 8.652)

Synthetic NMT

Доклад Генерального секретаря о реформе управления людскими ресурсами : доклад Генерального секретаря о реформе управления людскими ресурсами : доклад Генерального секретаря (BLEU+1: 21.16, RIBES: 61.48, NIST: 3.608)

In this example, the corresponding Russian word form of the Chinese word “流动性” (mobility) is “мобильность.” However, Tri.TreeExact places this word in the incorrect case form “мобильности” and also positions it far from the correct placing since the translation rule connecting “流动性” with the correct form “мобильность” is lost in the process of exact matching, and this rule is maintained in Tri.Hiero and Tri.TreePartial. The selection of the incorrect case form often causes misplacing because of LMs, and it is affirmed that the results obtained by the use of Tri.TreeExact contain more incorrect word forms and positions.

5.3 Related Work

Up to this point, representative pivot translation methods in SMT have been explained. Other related research studies in pivot translation are primarily based on the triangulation for PBMT and focuse on discussions to further improve accuracy (Zhu, He, Wu, Zhu, Wang, and Zhao 2014; Levinboim and Chiang 2015; Dabre, Cromieres, Kurohashi, and Bhattacharyya 2015). The process of correctly estimating the translation probability is a problem in triangulation.

Zhu et al. (2014) have proposed an estimation method of source–target translation probability by estimating source-target co-occurrence counts first instead of the direct estimation from source–pivot and pivot–target translation probabilities (Equations 11–14). They have reported that stable translation accuracy can be obtained even in the triangulation of two phrase tables with unbalanced table size.

Levinboim and Chiang (2015) have asserted that it is especially difficult to estimate word-level translation probability for phrase correspondence in the triangulation stage. Subsequently, they have proposed a method for improving the quality of the triangulation by estimating the translation probability even for the correspondence of words which cannot be directly observed, using a distributed expression of words (Mikolov, Yih, and Zweig 2013).

This paper focuses on pivot translation using English as the pivot language, though it is also known that translation accuracy varies on the manner in which a pivot language is selected. The influence of the choice of the pivot language on pivot translation has been discussed in detail by Paul et al. (2009). In reality, there are few situations in which the pivot language can be selected from multiple viable candidates, though in the ideal scenario where bilingual corpora of the same scale can be obtained via several languages, a pivot language having a similar language structure as the source or target language should be chosen.

Additionally, it is not necessary to limit the number of pivot languages to one, and methods that consider the simultaneous use of multiple pivot languages have also been proposed. Representatives of such purposing include methods such as aggregating multiple source–target phrase/rule tables obtained by triangulation with respective pivot languages into one table with linear interpolation and those that accomplish searching via the simultaneouse use of multiple TMs (Dabre et al. 2015).

Alternatively, training methods of multilingual NMT, which improve translation accuracy by causing translation tasks of multiple language pairs to be trained as a common encoder, have been also proposed (Dong, Wu, He, Yu, and Wang 2015; Zoph and Knight 2016; Johnson et al. 2017).

6 Conclusion

In this paper, the authors have proposed a new method of pivot translation to resolve the pivot-side syntactic ambiguity. This proposed method introduces an explicitly syntax-aware matching condition to find the correct correspondences between source-pivot and pivot-target translation rules. It can thus produce more reliable models. The results on the multilingual translation experiments revealed significant improvements in MT evaluation scores for the proposed method for all tested language pairs that possessed larger indirectly parallel corpora than (Miura et al. 2015), of 1M sentence pairs and that had access to the additional resource of English syntactic parsing. A syntactic matching method that allowed partial matching successfully reduced the number of noisy translation rules. This noise decrease enhanced the estimation of translation probabilities and better translation accuracy. This method is effective instances where that accurate syntactic parsers for the pivot language are available, and t is practical to use for pivot translations that allow access to larger quantities of parallel corpora.

To estimate translation probabilities, a heuristic that had no guarantee of being optimal was introduced. The smoothing technique of TMs could present an effective solution as one of the directions to improve the estimation of probability scores. It is common to apply smoothing methods such as back-off and interpolation when the coverage of a single model is poor. In their experimental setting, the authors have incorporated reliable and high-coverage models, and it should be easy to combine them by linear interpolation with fixed coefficients. In in the future, therefore, the authors are planning to explore more refined estimation methods that utilize machine learning.

Incidentally, the previously proposed method of pivot translation (Miura et al. 2015) uses MSCFG models that possess the potential to access information from the source, target, and pivot languages. For example, it should be possible to combine the proposed methods in Section 4 and (Miura et al. 2015) and to let the MSCFG model remember the pivot tree structures.

As a more advanced method for future research, it should be possible to devise compounded MSCFG models that can store both pivot-side and source-side syntactic information, thereby realizing translations with higher reproducibility of source information. The authors have mentioned that pivot translation presents the problem of the loss of source language information, which is affected by the expressiveness of the pivot language. In fact, this problem is frequent not just in MT but also for human translators. For example, since modern English is known for its simple morphology that does not include complicated grammatical conjugations such as personal suffixes, linguistic modalities such as number, case, and gender are lost when translating into English. This deficiency yields a translation that is different from the original meaning from English into another language. With the method posited in this paper, the authors aim to achieve machine translation outcomes that preserve the linguistic information of the original content by combining it with pivot-side syntactic structures.

Acknowledgement

Part of this work was supported by JSPS KAKENHI Grant Numbers JP16H05873 and JP17H06101. The authors also thank the anonymous reviewers for their insightful comments.

References

Aho, A. V. and Ullman, J. D. (1969). "Syntax Directed Translations and the Pushdown Assembler." *Journal of Computer and System Sciences*, **3** (1), pp. 37–56.

Brown, P. F., Pietra, V. J., Pietra, S. A. D., and Mercer, R. L. (1993). "The Mathematics of Statistical Machine Translation: Parameter Estimation." *Computational Linguistics*, **19**, pp. 263–312.

Chappelier, J.-C., Rajman, M., et al. (1998). "A Generalized CYK Algorithm for Parsing Stochastic CFG." In *Proc. TAPD*, Vol. 98, pp. 133–137. Citeseer.

Chiang, D. (2007). "Hierarchical Phrase-Based Translation." *Computational Linguistics*, **33** (2), pp. 201–228.

Cohn, T. and Lapata, M. (2007). "Machine Translation by Triangulation: Making Effective Use of Multi-Parallel Corpora." In *Proc. ACL*, pp. 728–735.

Dabre, R., Cromieres, F., Kurohashi, S., and Bhattacharyya, P. (2015). "Leveraging Small Multilingual Corpora for SMT Using Many Pivot Languages." In *Proc. NAACL*, pp. 1192–1202.

de Gispert, A. and Mariño, J. B. (2006). "Catalan-English Statistical Machine Translation without Parallel Corpus: Bridging through Spanish." In *Proc. of LREC 5th Workshop on Strategies for developing machine translation for minority languages*, pp. 65–68.

Doddington, G. (2002). "Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics." In *Proc. ARPA Workshop on Human Language Technology*, pp. 138–145.

Dong, D., Wu, H., He, W., Yu, D., and Wang, H. (2015). "Multi-Task Learning for Multiple Language Translation." In *Proc. ACL*, pp. 1723–1732.

Dyer, C., Cordova, A., Mont, A., and Lin, J. (2008). "Fast, Easy, and Cheap: Construction of

Statistical Machine Translation Models with MapReduce." In *Proc. WMT*, pp. 199–207.

Firat, O., Sankaran, B., Al-Onaizan, Y., Yarman Vural, F. T., and Cho, K. (2016). "Zero-Resource Translation with Multi-Lingual Neural Machine Translation." In *Proc. EMNLP*, pp. 268–277.

Graehl, J. and Knight, K. (2004). "Training Tree Transducers." In *Proc. NAACL*, pp. 105–112.

Heafield, K. (2011). "KenLM: Faster and Smaller Language Model Queries." In *Proc, WMT*, pp. 187–197.

Isozaki, H., Hirao, T., Duh, K., Sudoh, K., and Tsukada, H. (2010). "Automatic Evaluation of Translation Quality for Distant Language Pairs." In *Proc. EMNLP*, pp. 944–952.

Johnson, M., Schuster, M., Le, Q., 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." *TACL*, **5**, pp. 339–351.

Klein, P. N. (1998). "Computing the Edit-Distance Between Unrooted Ordered Trees." In *Proc. of European Symposium on Algorithms*, pp. 91–102.

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 *Proc. ACL*, pp. 177–180.

Koehn, P., Och, F. J., and Marcu, D. (2003). "Statistical Phrase-Based Translation." In *Proc. NAACL*, pp. 48–54.

Levinboim, T. and Chiang, D. (2015). "Supervised Phrase Table Triangulation with Neural Word Embeddings for Low-Resource Languages." In *Proc. EMNLP*, pp. 1079–1083.

Mikolov, T., Yih, W.-t., and Zweig, G. (2013). "Linguistic Regularities in Continuous Space Word Representations." In *Proc. NAACL*, pp. 746–751.

Miura, A., Neubig, G., Sakti, S., Toda, T., and Nakamura, S. (2015). "Improving Pivot Translation by Remembering the Pivot." In *Proc. ACL*, pp. 573–577.

Miura, A., Neubig, G., Sudoh, K., and Nakamura, S. (2017). "Tree as a Pivot: Syntactic Matching Methods in Pivot Translation." In *Proc. WMT*, pp. 90–98.

Neubig, G. (2013). "Travatar: A Forest-to-String Machine Translation Engine based on Tree Transducers." In *Proc. ACL Demo Track*, pp. 91–96.

Neubig, G., Arthur, P., and Duh, K. (2015). "Multi-Target Machine Translation with Multi-Synchronous Context-free Grammars." In *Proc. NAACL*, pp. 484–491.

Neubig, G., Nakata, Y., and Mori, S. (2011). "Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis." In *Proc. ACL*, pp. 529–533.

Oda, Y., Neubig, G., Sakti, S., Toda, T., and Nakamura, S. (2015). "Ckylark: A More Robust

PCFG-LA Parser.” In *Proc. NAACL*, pp. 41–45.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). “BLEU: a Method for Automatic Evaluation of Machine Translation.” In *Proc. ACL*, pp. 311–318.

Paul, M., Yamamoto, H., Sumita, E., and Nakamura, S. (2009). “On the Importance of Pivot Language Selection for Statistical Machine Translation.” In *Proc. NAACL*, pp. 221–224.

Sennrich, R., Haddow, B., and Birch, A. (2016). “Improving Neural Machine Translation Models with Monolingual Data.” In *Proc. ACL*, pp. 86–96.

Utiyama, M. and Isahara, H. (2007). “A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation.” In *Proc. NAACL*, pp. 484–491.

Zhu, X., He, Z., Wu, H., Zhu, C., Wang, H., and Zhao, T. (2014). “Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs.” In *Proc. EMNLP*, pp. 1665–1675.

Ziemski, M., Junczys-Dowmunt, M., and Pouliquen, B. (2016). “The United Nations Parallel Corpus v1.0.” In *Proc. LREC*, pp. 3530–3534.

Zoph, B. and Knight, K. (2016). “Multi-Source Neural Translation.” In *Proc. NAACL*, pp. 30–34.

三浦明波，Neubig Graham，Sakti Sakriani，戸田智基，中村哲 (2016a). 中間言語情報を記憶するピボット翻訳手法. 自然言語処理, **23** (5), pp. 499–528.

三浦明波，Neubig Graham，中村哲 (2016b). 木構造を中間表現とするピボット翻訳手法. 情報処理学会 第 227 回自然言語処理研究会 (SIG-NL), 9 号, pp. 1–6.

Akiva Miura: received his B.Sc. from Technion - Israel Institute of Technology in 2013, and his M.Eng. and Ph.D. in information science from Nara Institute of Science and Technology in 2016 and 2018 respectively. From 2018, he has been a researcher at Fujitsu Laboratories Ltd., where he is pursuing research in machine translation and knowledge technology.

Graham Neubig: received his B.E. from University of Illinois, Urbana-Champaign in 2005, and his M.S. and Ph.D. in informatics from Kyoto University in 2010 and 2012 respectively. He is currently an assistant professor at Carnegie Mellon University and an affiliate associate professor at Nara Institute of Science and Technology. His research interests include speech and natural language processing, with a focus on machine learning approaches for applications such as machine translation, speech recognition, and spoken dialog.

Katsuhito Sudoh: is an associate professor of Nara Institute of Science and

Technology. He received a bachelor's degree in engineering in 2000, and a master's and Ph.D. degree in informatics in 2002 and 2015, respectively, from Kyoto University. He was in NTT Communication Science Laboratories from 2002 to 2017. He currently works on machine translation and natural language processing. He is a member of the Association for Computational Linguistics (ACL), the Association of Natural Language Processing (ANLP), the Information Processing Society of Japan (IPSJ) and the Acoustical Society of Japan (ASJ)

Satoshi Nakamura: is Professor of Graduate School of Information Science, Nara Institute of Science and Technology, Honorarprofessor of Karlsruhe Institute of Technology, Germany, and ATR Fellow. He received his B.S. from Kyoto Institute of Technology in 1981 and Ph.D. from Kyoto University in 1992. He was Director of ATR Spoken Language Communication Research Laboratories in 2000-2008 and Vice president of ATR in 2007-2008. He was Director General of Keihanna Research Laboratories and the Executive Director of Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology in 2009-2010. He is currently Director of Augmented Human Communication laboratory and a full professor of Graduate School of Information Science at Nara Institute of Science and Technology. He has been Elected Board Member of International Speech Communication Association, ISCA, since June 2011 and IEEE Signal Processing Magazine Editorial Board member since April 2012, IEEE SPS Speech and Language Technical Committee Member since 2013, and IEEE Fellow since 2016.