Reordering Constraint Based on Document-Level Context

Takashi Onishi and **Masao Utiyama** and **Eiichiro Sumita**
Multilingual Translation Laboratory, MASTAR Project
National Institute of Information and Communications Technology
3-5 Hikaridai, Keihanna Science City, Kyoto, JAPAN
{takashi.onishi,mutiyama,eiichiro.sumita}@nict.go.jp

Abstract

One problem with phrase-based statistical machine translation is the problem of long-distance reordering when translating between languages with different word orders, such as Japanese-English. In this paper, we propose a method of imposing reordering constraints using document-level context. As the document-level context, we use noun phrases which significantly occur in context documents containing source sentences. Given a source sentence, zones which cover the noun phrases are used as reordering constraints. Then, in decoding, reorderings which violate the zones are restricted. Experiment results for patent translation tasks show a significant improvement of 1.20% BLEU points in Japanese-English translation and 1.41% BLEU points in English-Japanese translation.

1 Introduction

Phrase-based statistical machine translation is useful for translating between languages with similar word orders. However, it has problems with long-distance reordering when translating between languages with different word orders, such as Japanese-English. These problems are especially crucial when translating long sentences, such as patent sentences, because many combinations of word orders cause high computational costs and low translation quality.

In order to address these problems, various methods which use syntactic information have been proposed. These include methods where source sentences are divided into syntactic chunks or clauses and the translations are merged later (Koehn and

Knight, 2003; Sudoh et al., 2010), methods where syntactic constraints or penalties for reordering are added to a decoder (Yamamoto et al., 2008; Cherry, 2008; Marton and Resnik, 2008; Xiong et al., 2010), and methods where source sentences are reordered into a similar word order as the target language in advance (Katz-Brown and Collins, 2008; Isozaki et al., 2010). However, these methods did not use document-level context to constrain reorderings. Document-level context is often available in real-life situations. We think it is a promising clue to improving translation quality.

In this paper, we propose a method where reordering constraints are added to a decoder using document-level context. As the document-level context, we use noun phrases which significantly occur in context documents containing source sentences. Given a source sentence, zones which cover the noun phrases are used as reordering constraints. Then, in decoding, reorderings which violate the zones are restricted. By using document-level context, contextually-appropriate reordering constraints are preferentially considered. As a result, the translation quality and speed can be improved. Experiment results for the NTCIR-8 patent translation tasks show a significant improvement of 1.20% BLEU points in Japanese-English translation and 1.41% BLEU points in English-Japanese translation.

2 Patent Translation

Patent translation is difficult because of the amount of new phrases and long sentences. Since a patent document explains a newly-invented apparatus or method, it contains many new phrases. Learning phrase translations for these new phrases from the

Source	パッド 電極 １ １ は 、 第 １ の 絶縁 膜 で ある 層間 絶縁 膜 １ ２ を 介 して 半導体 基板 １ ０ の 表面 に 形成 されている 。
Reference	the pad electrode 11 is formed on the top surface of the semiconductor substrate 10 through **an interlayer insulation film 12 that is a first insulation film** .
Baseline output	**an interlayer insulating film 12** is formed on the surface of a semiconductor substrate 10 , a pad electrode 11 via **a first insulating film** .
Source + Zone	パッド 電極 １ １ は 、 \<**zone**\> 第 １ の \<**zone**\> 絶縁 膜 \<**/zone**\> で ある 層間 \<**zone**\> 絶縁 膜 \<**/zone**\> １ ２ \<**/zone**\> を 介 して 半導体 基板 １ ０ の 表面 に 形成 されている 。
Proposed output	pad electrode 11 is formed on the surface of the semiconductor substrate 10 through **the interlayer insulating film 12 of the first insulating film** .

Table 1: An example of patent translation.

training corpora is difficult because these phrases occur only in that patent specification. Therefore, when translating such phrases, a decoder has to combine multiple smaller phrase translations. Moreover, sentences in patent documents tend to be long. This results in a large number of combinations of phrasal reorderings and a degradation of the translation quality and speed.

Table 1 shows how a failure in phrasal reordering can spoil the whole translation. In the baseline output, the translation of "第 １ の 絶縁 膜 で ある 層間 絶縁 膜 １ ２" (*an interlayer insulation film 12 that is a first insulation film*) is divided into two blocks, "*an interlayer insulating film 12*" and "*a first insulating film*". In this case, a reordering constraint to translate "第 １ の 絶縁 膜 で ある 層間 絶縁 膜 １ ２" as a single block can reduce incorrect reorderings and improve the translation quality. However, it is difficult to predict what should be translated as a single block.

Therefore, how to specify ranges for reordering constraints is a very important problem. We propose a solution for this problem that uses the very nature of patent documents themselves.

3 Proposed Method

In order to address the aforementioned problem, we propose a method for specifying phrases in a source sentence which are assumed to be translated as single blocks using document-level context. We call these phrases "coherent phrases". When translating a document, for example a patent specification, we first extract coherent phrase candidates from the document. Then, when translating each sentence in the document, we set zones which cover the coher-

ent phrase candidates and restrict reorderings which violate the zones.

3.1 Coherent phrases in patent documents

As mentioned in the previous section, specifying coherent phrases is difficult when using only one source sentence. However, we have observed that document-level context can be a clue for specifying coherent phrases. In a patent specification, for example, noun phrases which indicate parts of the invention are very important noun phrases. In previous example, "第 １ の 絶縁 膜 で ある 層間 絶縁 膜 １ ２" is a part of the invention. Since this is not language dependent, in other words, this noun phrase is always a part of the invention in any other language, this noun phrase should be translated as a single block in every language. In this way, important phrases in patent documents are assumed to be coherent phrases.

We therefore treat the problem of specifying coherent phrases as a problem of specifying important phrases, and we use these phrases as constraints on reorderings. The details of the proposed method are described below.

3.2 Finding coherent phrases

We propose the following method for finding coherent phrases in patent sentences. First, we extract coherent phrase candidates from a patent document. Next, the candidates are ranked by a criterion which reflects the document-level context. Then, we specify coherent phrases using the rankings. In this method, using document-level context is critically important because we cannot rank the candidates without it.

3.2.1 Extracting coherent phrase candidates

Coherent phrase candidates are extracted from a context document, a document that contains a source sentence. We extract all noun phrases as coherent phrase candidates since most noun phrases can be translated as single blocks in other languages (Koehn and Knight, 2003). These noun phrases include nested noun phrases.

3.2.2 Ranking with C-value

The candidates which have been extracted are nested and have different lengths. A naive method cannot rank these candidates properly. For example, ranking by frequency cannot pick up an important phrase which has a long length, yet, ranking by length may give a long but unimportant phrase a high rank. In order to select the appropriate coherent phrases, measurements which give high rank to phrases with high termhood are needed. As one such measurement, we use C-value (Frantzi and Ananiadou, 1996).

C-value is a measurement of automatic term recognition and is suitable for extracting important phrases from nested candidates. The C-value of a phrase p is expressed in the following equation:

$$C\text{-}value(p) = \begin{cases} (l(p)-1)\,n(p) & (c(p)=0) \\ (l(p)-1)\left(n(p)-\frac{t(p)}{c(p)}\right) & (c(p)>0) \end{cases}$$

where
$l(p)$ is the length of a phrase p,
$n(p)$ is the frequency of p in a document,
$t(p)$ is the total frequency of phrases which contain p as a subphrase,
$c(p)$ is the number of those phrases.

Since phrases which have a large C-value frequently occur in a context document, these phrases are considered to be a significant unit, i.e., a part of the invention, and to be coherent phrases.

3.2.3 Specifying coherent phrases

Given a source sentence, we find coherent phrase candidates in the sentence in order to set zones for reordering constraints. If a coherent phrase candidate is found in the source sentence, the phrase is regarded a coherent phrase and annotated with a zone tag, which will be mentioned in the next section.

We check the coherent phrase candidates in the sentence in descending C-value order, and stop when the C-value goes below a certain threshold. Nested zones are allowed, unless their zones conflict with pre-existing zones. We then give the zone-tagged sentence, an example is shown in Table 1, as a decoder input.

3.3 Decoding with reordering constraints

In decoding, reorderings which violate zones, such as the baseline output in Table 1, are restricted and we get a more appropriate translation, such as the proposed output in Table 1.

We use the Moses decoder (Koehn et al., 2007; Koehn and Haddow, 2009), which can specify reordering constraints using <zone> and </zone> tags. Moses restricts reorderings which violate zones and translates zones as single blocks.

4 Experiments

In order to evaluate the performance of the proposed method, we conducted Japanese-English (J-E) and English-Japanese (E-J) translation experiments using the NTCIR-8 patent translation task dataset (Fujii et al., 2010). This dataset contains a training set of 3 million sentence pairs, a development set of 2,000 sentence pairs, and a test set of 1,251 (J-E) and 1,119 (E-J) sentence pairs. Moreover, this dataset contains the patent specifications from which sentence pairs are extracted. We used these patent specifications as context documents.

4.1 Baseline

We used Moses as a baseline system, with all the settings except distortion limit (dl) at the default. The distortion limit is a maximum distance of reordering. It is known that an appropriate distortion-limit can improve translation quality and decoding speed. Therefore, we examined the effect of a distortion-limit. In experiments, we compared dl = 6, 10, 20, 30, 40, and −1 (unlimited). The feature weights were optimized to maximize BLEU score by MERT (Och, 2003) using the development set.

4.2 Compared methods

We compared two methods, the method of specifying reordering constraints with a context document

	in (this case) , (the leading end) 15f of (the segment operating body) ((15 swings) in
w/o Context	in (this case) , (the leading end) 15f of (the segment operating body) ((15 swings) in (a direction opposite)) to (the a arrow direction) .
w/ Context	in (this case) , ((the leading end) 15f) of (((the segment) operating body) 15) swings in a direction opposite to (the a arrow direction) .

Table 3: An example of the zone-tagged source sentence. <zone> and </zone> are replaced by "(" and ")".

System	dl	J→E		E→J	
		BLEU	Time	BLEU	Time
	6	27.83	4.8	35.39	3.5
	10	30.15	6.9	38.14	4.9
	20	30.65	11.9	38.39	8.5
Baseline	30	30.72	16.0	38.32	11.5
	40	29.96	19.6	38.42	13.9
	−1	30.35	28.7	37.80	18.4
w/o Context	−1	30.01	8.7	38.96	5.9
w/ Context	−1	**31.55**	12.0	**39.21**	8.0

Table 2: BLEU score (%) and average decoding time (sec/sentence) in J-E/E-J translation.

(w/ Context) and the method of specifying reordering constraints without a context document (w/o Context). In both methods, the feature weights used in decoding are the same value as those for the baseline (dl = −1).

4.2.1 Proposed method (w/ Context)

In the proposed method, reordering constraints were defined with a context document. For J-E translation, we used the CaboCha parser (Kudo and Matsumoto, 2002) to analyze the context document. As coherent phrase candidates, we extracted all subtrees whose heads are noun. For E-J translation, we used the Charniak parser (Charniak, 2000) and extracted all noun phrases, labeled "NP", as coherent phrase candidates. The parsers are used only when extracting coherent phrase candidates. When specifying zones for each source sentence, strings which match the coherent phrase candidates are defined to be zones. Therefore, the proposed method is robust against parsing errors. We tried various thresholds of the C-value and selected the value that yielded the highest BLEU score for the development set.

4.2.2 w/o Context

In this method, reordering constraints were defined without a context document. For J-E translation, we converted the dependency trees of source sentences processed by the CaboCha parser into bracketed trees and used these as reordering constraints. For E-J translation, we used all of the noun phrases detected by the Charniak parser as reordering constraints.

4.3 Results and Discussions

The experiment results are shown in Table 2. For evaluation, we used the case-insensitive BLEU metric (Papineni et al., 2002) with a single reference.

In both directions, our proposed method yielded the highest BLEU scores. The absolute improvement over the baseline (dl = −1) was 1.20% in J-E translation and 1.41% in E-J translation. According to the bootstrap resampling test (Koehn, 2004), the improvement over the baseline was statistically significant ($p < 0.01$) in both directions. When compared to the method without context, the absolute improvement was 1.54% in J-E and 0.25% in E-J. The improvement over the baseline was statistically significant ($p < 0.01$) in J-E and almost significant ($p < 0.1$) in E-J. These results show that the proposed method using document-level context is effective in specifying reordering constraints.

Moreover, as shown in Table 3, although zone setting without context is failed if source sentences have parsing errors, the proposed method can set zones appropriately using document-level context. The Charniak parser tends to make errors on noun phrases with ID numbers. This shows that document-level context can possibly improve parsing quality.

As for the distortion limit, while an appropriate distortion-limit, 30 for J-E and 40 for E-J, improved the translation quality, the gains from the proposed method were significantly better than the gains from the distortion limit. In general, imposing strong constraints causes fast decoding but low translation quality. However, the proposed method improves the translation quality and speed by imposing appropriate constraints.

5   Conclusion

In this paper, we proposed a method for imposing reordering constraints using document-level context. In the proposed method, coherent phrase candidates are extracted from a context document in advance. Given a source sentence, zones which cover the coherent phrase candidates are defined. Then, in decoding, reorderings which violate the zones are restricted. Since reordering constraints reduce incorrect reorderings, the translation quality and speed can be improved. The experiment results for the NTCIR-8 patent translation tasks show a significant improvement of 1.20% BLEU points for J-E translation and 1.41% BLEU points for E-J translation.

We think that the proposed method is independent of language pair and domains. In the future, we want to apply our proposed method to other language pairs and domains.

References

Eugene Charniak. 2000. A Maximum-Entropy-Inspired Parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 132–139.

Colin Cherry. 2008. Cohesive Phrase-Based Decoding for Statistical Machine Translation. In *Proceedings of ACL-08: HLT*, pages 72–80.

Katerina T. Frantzi and Sophia Ananiadou. 1996. Extracting Nested Collocations. In *Proceedings of COLING 1996*, pages 41–46.

Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro, Terumasa Ehara, Hiroshi Echizen-ya, and Sayori Shimohata. 2010. Overview of the Patent Translation Task at the NTCIR-8 Workshop. In *Proceedings of NTCIR-8 Workshop Meeting*, pages 371–376.

Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010. Head Finalization: A Simple Reordering Rule for SOV Languages. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 244–251.

Jason Katz-Brown and Michael Collins. 2008. Syntactic Reordering in Preprocessing for Japanese→English Translation: MIT System Description for NTCIR-7 Patent Translation Task. In *Proceedings of NTCIR-7 Workshop Meeting*, pages 409–414.

Philipp Koehn and Barry Haddow. 2009. Edinburgh's Submission to all Tracks of the WMT 2009 Shared Task with Reordering and Speed Improvements to Moses. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 160–164.

Philipp Koehn and Kevin Knight. 2003. Feature-Rich Statistical Translation of Noun Phrases. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej 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 Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In *Proceedings of EMNLP 2004*, pages 388–395.

Taku Kudo and Yuji Matsumoto. 2002. Japanese Dependency Analysis using Cascaded Chunking. In *Proceedings of CoNLL-2002*, pages 63–69.

Yuval Marton and Philip Resnik. 2008. Soft Syntactic Constraints for Hierarchical Phrased-Based Translation. In *Proceedings of ACL-08: HLT*, pages 1003–1011.

Franz Josef Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 160–167.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, Tsutomu Hirao, and Masaaki Nagata. 2010. Divide and Translate: Improving Long Distance Reordering in Statistical Machine Translation. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 418–427.

Deyi Xiong, Min Zhang, and Haizhou Li. 2010. Learning Translation Boundaries for Phrase-Based Decoding. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 136–144.

Hirofumi Yamamoto, Hideo Okuma, and Eiichiro Sumita. 2008. Imposing Constraints from the Source Tree on ITG Constraints for SMT. In *Proceedings of the ACL-08: HLT Second Workshop on Syntax and Structure in Statistical Translation (SSST-2)*, pages 1–9.