A Comparison Study of Parsers for Patent Machine Translation

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

Abstract

Machine translation of patent documents is very important from a practical point of view. One of the key technologies for improving machine translation quality is the utilization of syntax. It is difficult to select the appropriate parser for patent translation because the effects of each parser on patent translation are not clear. This paper provides comparative evaluation of several state-of-the-art parsers for English, focusing on the effects for patent machine translation from English to Japanese. We measured how much each parser contributed to improve translation quality when the parser was used to obtain the syntax of input sentences. In addition, we examined the effects of a method using parsed document-level context containing the input sentence to determine noun phrases (Onishi et al., 2011). We conducted experiments using the NTCIR-8 patent translation task dataset. Most of the parsers improved translation quality. When the method using document-level context was applied, all of the compared parsers improved translation quality.

1 Introduction

In recent years, demands for patent machine translation have increased. With globalization comes an increase in the need for the international circulation of patent documents. It is, therefore, important to improve the quality of machine translation of patent sentences. Word ordering is the main issue in statistical machine translation of long patent sentences between language pairs with widely different word orders, such as English-Japanese. One of the key technologies for improving translation quality is the utilization of syntax to determin proper word order. The syntax of an input sentence is considered useful to determine the word order of a translated sentence.

It is difficult to select the appropriate parser for patent translation. There are mainly two reasons:

- Parsing is a difficult task, and several methods have been proposed in recent years. There are probabilistic CFG-based parser (Collins, 1997; Charniak, 2000; Klein and Manning, 2003; Petrov and Klein, 2007), dependency parser (McDonald and Pereira, 2006), and HPSG-based parser (Miyao and Tsujii, 2008).

- The effects of each parser on patent translation are not clear in the commonly used evaluations of parsers. Most state-of-the-art parsers for English were trained with the Wall Street Journal (WSJ) from the Penn Treebank corpus. Such parsers were evaluated by measuring bracketing precision and recall of the output using the WSJ from the Penn Treebank corpus. From the evaluation, it is not clear how well these models work in the other domains such as patent domain.

There is a task-oriented evaluation (Miyao et al., 2008). Miyao et al. (2008) compared parsers based on the accuracy of identifying protein-protein interaction that used the parser's output as features for machine learning models. This evaluation showed the effect of each parser for the protein-protein interaction task using biomedical papers.

There is research that studied the relation between parse accuracy and translation quality (Quirk and Corston-Oliver, 2006). This showed the relationship between a parser's training data size and the translation quality. They did not compare parsers, nor did they use a patent corpus. Research has also been done on the relationship between four parsers and translation quality of syntax-based statistical machine translation (Zhang et al., 2006). They did not use patent corpus, and only evaluated probabilistic CFG-based parsers. They used target side syntax and did not use source side syntax. To the best of our knowledge, there has been no previous research comparing the effects of parsers on patent machine translation.

In this paper, we compared the effects of several state-of-the-art parsers on patent machine translation. This research reveals how effective each parser is in patent machine translation.

There are statistical machine translation methods that use input sentence syntax: reordering constraint methods (Cherry, 2008; Marton and Resnik, 2008; Yamamoto et al., 2008; Xiong et al., 2010; Onishi et al., 2011), tree-to-string methods (Liu et al., 2006; Huang et al., 2006), and tree-to-tree methods (Cowan et al., 2006; Zhang et al., 2008; Liu et al., 2009). In this research, we used a reordering constraint method, which directly controls word order using the syntax of an input sentence for phrase based statistical machine translation, one of the widely used statistical translation methods. The syntax structure was obtained using each parser to be compared. We evaluate the effects of each parser on patent machine translation by evaluating patent machine translation quality.

Moreover, we also applied a method that used document-level context containing the input sentence to determine the noun phrases in the input sentence (Onishi et al., 2011). These results showed how their method was effective with each parser.

The rest of this paper is organized as follows: In section 2, we show the six parsers that we compared. In section 3, we explain the method of comparison. In section 4, we discuss the experiment results from the NTCIR-8 patent translation task data. In section 5, we conclude this paper.

2 Parsers

We focused on six well-known publicly available parsers. The parsers are categorized by method into three groups: probabilistic CFG parser, dependency parser, and HPSG parser.

2.1 Probabilistic CFG parser

Owing to Penn Treebank (Marcus et al., 1993), there has been a lot of research into parsers based on probabilistic CFG that output phrase structures. Fig. 1 shows an example of a phrase structure. The ways to parameterize the probabilistic models vary. In this research, we used the following four parsers:

COLLINS Collins' (1997) parser. The parser uses a lexicalized probabilistic CFG model. The tool includes three models: model 1, 2, and 3. We used model 3. Because the tool did not include a POS tagger function, we used Tsuruoka's English POS tagger (Tsuruoka and Tsujii, 2005) to get part-of-speech.

CHARNIAK Charniak's (2000) parser. The parser uses a lexicalized probabilistic CFG model. The model is based on the principle of maximum entropy.

STANFORD Stanford's parser (Klein and Manning, 2003). The parser uses an unlexicalized probabilistic CFG model. We used version 1.6.5.

BERKELEY Berkeley's parser (Petrov and Klein, 2007). The parser uses an unlexicalized probabilistic CFG model. We used release 1.1.

2.2 Dependency parser

Owing to the CoNLL shared tasks (Buchholz and Marsi, 2006; Nivre et al., 2007), research into dependency parsing have been active. Dependency structure is a tree structure in which a node is a word and an edge is the relation between a parent node and a child node. A child node modifies its parent node. Fig. 2 shows an example of a dependency tree structure. In this research, we used the following parser:

MST MacDonald and Pereira's (2006) parser. Projective dependency parsing is based on Eisner's algorithm (Eisner, 1996). We used version 0.4.3b. The tool did not contain a model.

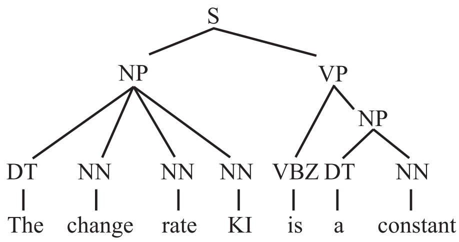

Figure 1: Penn Treebank-style phrase structure

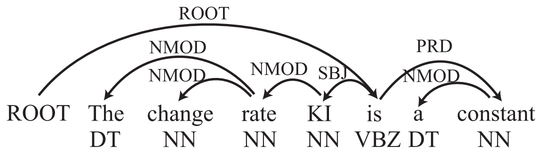

Figure 2: Dependency tree structure

We built a model using WSJ section 2 to 21 from Penn Treebank. Because the tool did not include a POS tagger function, we used Tsuruoka's English POS tagger (Tsuruoka and Tsujii, 2005) to get part-of-speech.

2.3 HPSG parser

There is a parser based on the HPSG (Pollard and Sag, 1994) theory. HPSG-based parsers analyze not only phrase structure but also deeper structures, such as the arguments of a predicate, simultaneously. We used only the phrase structures of the parse results. In this research, we used the following parser:

ENJU An HPSG parser (Miyao and Tsujii, 2008). It consists of an HPSG grammar extracted from the Penn Treebank, and a maximum entropy model trained with an HPSG Treebank derived from the Penn Treebank. We used version 2.3.1.

3 Comparison methodology

We compared parsers based on the translation quality of patent sentences translated by a phrase-based statistical machine translation with reordering constraints using syntax of input sentences. We translated from English to Japanese, whose word orders are widely different. In translation between languages with widely different word orders, it is difficult to assign the proper word order, especially with long input sentences. Input sentence syntax is useful in deciding a word order for the translated sentence. We parsed the input sentence and constrained the word order using these parsed results. The translation quality was measured using the 4-gram BLEU score (Papineni et al., 2002). There is a method that determines the noun phrases in an input sentence by using the parse results from document-level context that contains the input sentence. We applied this method and compared parsers based on the translation quality. We also examined the effects of this method on each parser and combinations of parsers.

Domain	Sentence length
Travel	7.7
News	21.0
Patent	30.3

Table 1: Average sentence length in three domains. Sentence length is the number of words per English sentence. We used the IWSLT corpus (Eck and Hori, 2005) in the travel domain, the WMT08 News Commentary corpus (Callison-Burch et al., 2008) in the news domain, and the NTCIR-8 Patent machine translation corpus (Fujii et al., 2010) in the patent domain.

First, we show the issue of patent translation. Next, we explain the methods that deal with the issue by reordering constraints using syntax of input sentences. Finally, we explain the method that estimates noun phrases using document-level context.

3.1 Patent translation

In this research, we focused on the translation of patent sentences. Patent sentence translation is difficult and the main reason for this is that patent sentences are long. As shown in Table 1, patent sentences are longer than those in other domains. In general, longer sentences cause an explosion of reordering combinations and degrade translation quality.

When we translate between languages with similar word orders, we can prevent the loss of translation quality by using distortion limits that constrain word reordering in phrase-based statistical machine translation. However, in translation between language pairs with widely different word orders, such as English-Japanese, long-distance word reordering is required when an input sentence is long. Therefore, when an input sentence for translation is long, the word order possibilities are large. This leads difficulty of determining the proper word reordering.

Below is an example of translation by our baseline system without using syntax. This example shows how a failure of word order affects the overall translation quality. Table 2 gives the meanings of the expressions in the Baseline Output.

Input sentence

> **a rotational position-detecting device 3 that is constructed of a resolver or a rotary encoder** is mounted on a shaft of a rotor , not shown , of the electric motor 1 .

Baseline Output

> 電動 機 1 の **回転 位置 検出 装置 3** 、 図示 し ない ロータ の 軸 に 装着 さ れ た **ロータリー エンコーダ や レゾルバ の で 構成 さ れる** 。

The bolded section of the input sentence was translated into two separated parts, in Gothic, in the baseline output. The bolded section of the input sentence refers to a single apparatus. Thusly, if the expression is translated as two separate expressions, the original meaning cannot be understood and is lost.

3.2 Reordering constraint for phrase based statistical machine translation

There are methods to address the reordering problem that constrain reordering using the syntax of the input sentences (Cherry, 2008; Marton and Resnik, 2008; Yamamoto et al., 2008; Xiong et al., 2010; Onishi et al., 2011). In this research, we focused on a reordering constraint method, which controls word order using the syntax of an input sentence for phrase based statistical machine translation. We investigated effects of parsers on a phrase based statistical machine translation with reordering constraints.

Using the aforementioned example, a reordering constraint that translates the bold section of the input sentence into one block reduces incorrect word ordering and improves translation quality.

For this research, we used parsers to obtain the syntax structures in input sentences, and constrained reordering to translate a noun phrase distinguished by parsing as one block.

Expressions in the output	Meanings in English
電動 機 1 の	of the electric motor 1
回転 位置 検出 装置 3	**a rotational position-detecting device 3**
図示 し ない	not shown
ロータ の 軸 に 装着 さ れ た	mounted on a shaft of a rotor
ロータリー エンコーダ や レゾルバ の で 構成 さ れる	**is constructed of a resolver or a rotary encoder**

Table 2: Meanings of expressions in the Baseline Output in order of the output.

The Moses phrase-based decoder has a function that constrains reordering using zone tags (Koehn and Haddow, 2009). Moses restricts reordering that violates zones specified by zone tags, and translates one zone to one block. We used this function of the Moses decoder to translate a noun phrase of parsed results into one block. We add zone tags that cover noun phrases to an input sentence. Zone tags can be nested if the new tag does not conflict with other existing tags. Here is an example of an input sentence with zone tags:

Input sentence with zone tags

> ⟨**zone**⟩ ⟨**zone**⟩ *a rotational position-detecting device 3* ⟨**/zone**⟩ *that is constructed of* ⟨**zone**⟩ *a resolver or a rotary encoder* ⟨**/zone**⟩ ⟨**/zone**⟩ is mounted on a shaft of a rotor , not shown , of the electric motor 1 .

Dependency structures do not explicitly express noun phrases. We regarded a subtree whose root node is a noun as a noun phrase. A “subtree” consists of a node and all of its descendent nodes.

3.3 Using document-level context

Onishi et al. (2011) proposed a method that did not use the noun phrases obtained by parsing an input sentence directly, but instead used the noun phrases determined by using the parse results of a context document, a document that contains an input sentence. This method can determine noun phrases by considering document-level consistency. The method is as follows:

1. The method parses a context document containing an input sentence.

Set	Number of sentences
Training	3.2 million
Development	2,000
Test	1,119

Table 3: Statistics for the NTCIR-8 English to Japanese patent translation task dataset.

2. The method extracts all noun phrases from the parse results.
3. The method ranks the noun phrases based on a C-value (Frantzi and Ananiadou, 1996) that gives high rank to phrases with high termhood from nested candidates.
4. The method searches the list of noun phrases (in order of rank) for expressions that appear in the input sentence.
5. The method determines the searched expression to be a noun phrase and adds zone tags if the expression does not conflict with existing zone tags.

The C-value of a phrase p is expressed in the following equation:

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

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.

Onishi et al. (2011) pointed out that since phrases with large C-values frequently occur in a context document, these phrases are considered a significant unit, i.e., a part of the invention, and are assumed to be translated as single blocks.

4 Experiment

We conducted English to Japanese patent translation experiments using the NTCIR-8 patent translation task data (Fujii et al., 2010). This data set consists of 3.2 million English-Japanese sentence pairs, development data of 2,000 sentence pairs, and test data of 1,119 sentences and their single reference data, as shown in Table 3. Furthermore, this dataset contains the patent specifications from which the test sentences were extracted. We used these patent specifications as context documents.

	BLEU	Gains from Baseline
Baseline	38.42	0
COLLINS	36.97	−1.45**
CHARNIAK	39.24	+0.82**
STANFORD	39.56	+1.06**
BERKELEY	**39.60**	+1.18**
MST	39.44	+1.02**
ENJU	39.40	+0.98**

Table 4: Comparison between parsers based on the effects of reordering constraints using the parsed results of test data.

4.1 Baseline

We used Moses for the machine translation system. The following settings were used:

- GIZA++ and grow-diag-final-and heuristics,
- 5-gram language model with interpolated modified Kneser-Ney discounting,
- msd-bidirectional-fe lexicalized reordering,
- distortion-limit = -1 (unlimited).

The feature weights were tuned by minimum error rate training using the development data.

4.2 Experiment 1

We evaluated parsers based on the effect of reordering constraints where the parse results of the test sentences were directly used. We parsed the test sentences using each parser and annotated the zone tags that cover noun phrases as described at section 3.2. In this experiment, we used the same feature weights as those for the baseline system.

Results and discussions

Table 4 gives the results of the translations using the reordering constraint of zone tags covering the noun phrases obtained directly by the parsers. "Baseline" indicates the result that did not use a parser and zone tags.

The five parsers other than COLLINS had improved the BLEU scores over the baseline BLEU score. From these results, it can be seen that the CHARNIAK, STANFORD, BERKELEY, MST, and

ENJU parsers were effective for patent machine translation.

The "**" mark in Table 4 and Table 5 denotes a statistical significant difference at the significance level of $\alpha = 0.01$ and the "*" mark denotes a statistical significant difference at the significance level of $\alpha = 0.05$ according to the bootstrap resampling test (Koehn, 2004).

The difference between BERKELY and the top parsers STANFORD, MST, and ENJU was not significant at $\alpha = 0.05$ and the difference between BERKELEY and CHARNIAK was significant at $\alpha = 0.05$. From these results, it can be seen that **BERKELEY**, **STANFORD**, **MST**, and **ENJU** were especially **effective** for patent machine translation among the six parsers when the noun phrases in an input sentence were obtained by parsing the input sentence directly.

4.3 Experiment 2

We evaluated parsers based on the effects of reordering constraints where noun phrases were determined using the parsed context documents as described in section 3.3. We also evaluated the effect of using context documents for each parser.

In addition, we used a combination of parsers in which one parser parsed a context document while another parser parsed the same context document. We used the two documents that had been parsed as parsed context documents and extracted noun phrases from them. The subsequent process is the same as described in section 3.3.

For this experiment, we used the noun phrases from a context document that had C-values greater than or equal to 1.0. Most of noun phrases had C-values greater than 1.0. We also used the same feature values as those for the baseline system.

Results and discussions

Table 5 gives the results of the translation using the reordering constraint of zone tags covering the noun phrases in the test data determined using parsed context documents. All the six parsers had improved BLEU scores over the baseline in Table 4. We also examined the effects using context documents. As shown in "Gains from without context" in Table 5, the BLEU scores using the parse results of context documents were higher than those of the results using only the parse results of the input sentences for all parsers.

	BLEU	Gains from without context
COLLINS	39.42	+2.45**
CHARNIAK	39.58	+0.34*
STANFORD	39.64	+0.08
BERKELEY	**39.76**	+0.16
MST	39.54	+0.10
ENJU	39.68	+0.28*

Table 5: Comparison between parsers based on the effects of reordering constraints using the parsed results of context documents.

	STANFORD	MST	ENJU
BERKELEY	**39.95**	39.65	39.86
STANFORD		39.79	39.92
MST			39.67

Table 6: Comparison between combinations of parsers based on the effects of reordering constraints using the parsed results of context documents. The values given are BLEU scores.

Table 6 shows the translation results with reordering constraints using context documents parsed by two parsers. We used the top four parsers in Table 4, BERKELEY, STANFORD, MST, and, ENJU for the parser combinations. The BLEU scores in Table 6, except for the parser combinations including MST, are higher than the single best BLEU scores of the two parsers in Table 5 using context.

Among the combinations and of all of the results, **the best BLEU score** was achieved by **the combination of BERKELEY and STANFORD**. The difference between the combination of BERKELEY and STANFORD with context and BERKELEY without context was significant at $\alpha = 0.05$. When comparing results, the difference between BERKELEY with context and BERKELEY without was not significant, whereas there was a significant difference between the combination of parsers with context and BERKELEY without context. From these results, it can be seen that **using context documents with combination of parsers is effective** for patent translation.

4.4 Experiment 3

We investigated the relationship between parse accuracy and translation quality. We randomly selected

Parser \Context	F-measure without	F-measure with	Cross brackets without	Cross brackets with
COLLINS	41.06	44.90	1.554	0.372
CHARNIAK	64.73	55.04	1.078	0.314
STANFORD (SF)	67.26	57.06	1.136	0.36
BERKELEY (BK)	**69.96**	56.89	0.846	0.334
MST	51.09	48.44	0.948	0.408
ENJU	62.00	56.14	0.77	0.324
BK & SF	-	60.16	-	0.35
BK & MST	-	56.94	-	0.36
BK & ENJU	-	59.25	-	**0.31**
SF & MST	-	56.70	-	0.39
SF & ENJU	-	59.14	-	0.38
MST & ENJU	-	54.64	-	0.37

Table 7: Comparison between combinations of parsers based on the effects of reordering constraints using the parsed results of context documents. “Cross brackets” shows the average number of cross brackets per sentence.

BERKELEY without context	[The conductor pattern 14a] is led out up to [*the first side*] **[**[*face 20b*] of [element 1] to be electrically connected to [the other terminal electrode 5]**]** .
BERKELEY and STANFORD with context	[[The [conductor pattern]] 14a] is led out up to [[the first side] face] 20b of [element 1] to be electrically connected to [the other [terminal electrode] 5] .

Table 8: Examples of noun phrase structures. Brackets are represented by “[” and “]”.

500 sentences from the test sentences and manually annotated them with noun phrase tags. We calculated the parse accuracy for noun phrases using the annotated corpus and a bracket-scoring program named evalb[1].

Results and discussions

Table 7 shows the parse accuracy of noun phrases. In Table 7, “without” means without using context and “with” means using context to determine noun phrases.

First, we will focus on the F-measure without context category. In this category, a high F-measure of parse accuracy almost produces high quality translation. BERKELEY’s score was the highest and it also had the highest BLEU score in Table 4.

[1] http://nlp.cs.nyu.edu/evalb/

Next, we focus on the difference between the F-measures “without” and “with” context. As shown in the results in Table 5, using context improved translation quality. However, as shown in Table 7, using context degrades the F-measure, which indicates that there must be an important factor other than the F-measure in translation quality.

Now, we focus on the difference between Cross brackets “without” and “with” context. Shown in Table 7, **using context** to determine noun phrases **reduced the average number of cross brackets** compared to the number from without using context. From this, we inferred that the reason for improvement in translation quality using context was from a reduction in cross bracket parse errors. Table 8 shows examples of noun phrases parsed by BERKELEY without context and determined by a combination of BERKELEY and STANFORD with context. The expression surrounded by cross brackets is underlined. The underlined expression crosses the noun phrase in italics. There are no cross brackets in the with-context results. Having multiple nested brackets degraded the bracketing precision of the with-context results.

Based on the analysis, we saw that parse results require not only high F-measure, but also low Cross brackets for patent machine translation.

5 Conclusion

We empirically compared the effects of six parsers on patent machine translation. We used a phrase based statistical machine translation method that used syntax structures in the source language for reordering constraints. We conducted experiments on English to Japanese patent translation using the NTCIR-8 patent translation task dataset. Most of the parsers, not only the probabilistic CFG parsers but also the dependency parser and the HPSG parser, were effective when a noun phrase reordering constraint was used. When a method that determined noun phrases using the parse results of document-level context was applied, all the parsers had effects on patent translation. The best translation quality was obtained when the Berkeley parser and the Stanford parser were used together and a method using document-level context was applied.

References

Sabine Buchholz and Erwin Marsi. 2006. Conll-x shared task on multilingual dependency parsing. In *Proceedings of CoNLL*, pages 149–164.

Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. 2008. Further meta-evaluation of machine translation. In *Proceedings of WMT*, pages 70–106.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of NAACL*, pages 132–139.

Colin Cherry. 2008. Cohesive phrase-based decoding for statistical machine translation. In *Proceedings of ACL-HLT*, pages 72–80.

Michael Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of ACL*, pages 16–23.

Brooke Cowan, Ivona Kučerová, and Michael Collins. 2006. A discriminative model for tree-to-tree translation. In *Proceedings of EMNLP*, pages 232–241.

Matthias Eck and Chiori Hori. 2005. Overview of the iwslt 2005 evaluation campaign. In *Proceedings of IWSLT*.

Jason M. Eisner. 1996. Three new probabilistic models for dependency parsing: An exploration. In *Proceedings of COLING*, pages 340–345.

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*, pages 371–376.

Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of locality. In *Proceedings of AMTA*, pages 66–73.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of ACL*, pages 423–430.

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 WMT*, pages 160–164.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP*, pages 388–395.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of COLING-ACL*, pages 609–616.

Yang Liu, Yajuan Lü, and Qun Liu. 2009. Improving tree-to-tree translation with packed forests. In *Proceedings of ACL-IJCNLP*, pages 558–566.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank. In *CL, Volume 19, Number 2*, pages 313–330.

Yuval Marton and Philip Resnik. 2008. Soft syntactic constraints for hierarchical phrased-based translation. In *Proceedings of ACL-HLT*, pages 1003–1011.

Ryan McDonald and Fernando Pereira. 2006. Online learning of approximate dependency parsing algorithms. In *Proceedings of EACL*, pages 81–88.

Yusuke Miyao and Jun'ichi Tsujii. 2008. Feature forest models for probabilistic hpsg parsing. In *CL, Volume 34, Number 1*, pages 81–88.

Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki, and Jun'ichi Tsujii. 2008. Task-oriented evaluation of syntactic parsers and their representations. In *Proceedings of ACL-HLT*, pages 46–54.

Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of CoNLL*, pages 915–932.

Takashi Onishi, Masao Utiyama, and Eiichiro Sumita. 2011. Reordering constraint based on document-level context. In *Proceedings of ACL-HLT*, pages 434–438.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of ACL*, pages 311–318.

Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *HLT-NAACL*, pages 404–411.

Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Chris Quirk and Simon Corston-Oliver. 2006. The impact of parse quality on syntactically-informed statistical machine translation. In *Proceedings of EMNLP*, pages 62–69.

Yoshimasa Tsuruoka and Jun'ichi Tsujii. 2005. Bidirectional inference with the easiest-first strategy for tagging sequence data. In *Proceedings of HLT-EMNLP*, pages 467–474.

Deyi Xiong, Min Zhang, and Haizhou Li. 2010. Learning translation boundaries for phrase-based decoding. In *Proceedings of HLT-NAACL*, 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 SSST-2*, pages 1–9.

Hao Zhang, Huizhen Wang, Tong Xiao, and Jingbo Zhu. 2006. The impact of parsing accuracy on syntax-based smt. In *Proceedings of NLP-KE*, pages 1–4.

Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lim Tan, and Sheng Li. 2008. A tree sequence alignment-based tree-to-tree translation model. In *Proceedings of ACL-HLT*, pages 559–567.