

# The RWTH Aachen System for NTCIR-9 PatentMT

Minwei Feng  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
feng@cs.rwth-aachen.de

Christoph Schmidt  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
schmidt@cs.rwth-aachen.de

Joern Wuebker  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
wuebker@cs.rwth-aachen.de

Stephan Peitz  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
peitz@cs.rwth-aachen.de

Markus Freitag  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
freitag@cs.rwth-aachen.de

Hermann Ney  
Lehrstuhl für Informatik 6  
RWTH Aachen University  
Ahornstr. 55  
D-52056 Aachen, Germany  
ney@cs.rwth-aachen.de

## ABSTRACT

This paper describes the statistical machine translation (SMT) systems developed by RWTH Aachen University for the Patent Translation task of the 9th NTCIR Workshop. Both phrase-based and hierarchical SMT systems were trained for the constrained Japanese-English and Chinese-English tasks. Experiments were conducted to compare different training data sets, training methods and optimization criteria as well as additional models for syntax and phrase reordering. Further, for the Chinese-English subtask we applied a system combination technique to create a consensus hypothesis from several different systems.

## Categories and Subject Descriptors

I.2.7 [Nature Language Processing]: machine translation

## General Terms

Experimentation

## Keywords

SMT, Patent Translation

## Team Name

RWTH Aachen

## Subtasks/Languages

Japanese-to-English PatentMT, Chinese-to-English PatentMT

## External Resources Used

Stanford Parser, MeCab, LDC Segmenter

## 1. INTRODUCTION

This is the RWTH Aachen University system paper for the Patent Translation Task of the 9th NTCIR Workshop. We submitted results for the two subtasks Japanese-English and Chinese-English. The structure of the paper is as follows: in Section 2, we describe

the baseline systems for both the Japanese-English and the Chinese-English task, including phrase-based and hierarchical SMT systems. Section 3 focuses on the system setup and additional models used for Japanese-English. Section 4 specifies the system setup and additional models used for Chinese-English. In both sections, experimental results are presented to compare different techniques. Finally, we draw some conclusions in Section 5.

## 2. TRANSLATION SYSTEMS

For the NTCIR-9 Patent Translation evaluation we utilized RWTH's state-of-the-art phrase-based and hierarchical translation systems as well as our in-house system combination framework. GIZA++ [13] was employed to train word alignments. All systems were evaluated using the automatic BLEU [14] and TER [15] metrics.

### 2.1 Phrase-Based System

We apply a phrase-based translation (PBT) system similar to the one described in [21]. Phrase pairs are extracted from a word-aligned bilingual corpus and their translation probabilities in both directions are estimated by relative frequencies. The standard feature set further includes an  $n$ -gram language model, phrase-level IBM-1 and word-, phrase- and distortion-penalties. Parameters are optimized with the downhill simplex algorithm [11] on the word graphs.

### 2.2 Hierarchical System

For the hierarchical setups described in this paper, the open source toolkit Jane [18] is employed. Jane has been developed at RWTH and implements the hierarchical approach as introduced by [1] with some state-of-the-art extensions. In hierarchical phrase-based translation, a weighted synchronous context-free grammar is induced from parallel text. In addition to contiguous *lexical* phrases, *hierarchical* phrases with up to two gaps are extracted. The search was carried out using the cube pruning algorithm [5].

The standard models integrated into the Jane baseline systems are phrase translation probabilities and lexical translation probabilities on the phrase level, each for both translation directions, length penalties on the word and phrase level, three binary features marking hierarchical phrases, glue rules as well as rules with non-terminals at the boundaries, source-to-target and target-to-source

phrase length ratios, four binary count features and an  $n$ -gram language model. The model weights are optimized with standard MERT [12] on 100-best lists.

In addition to the standard features, *parse matching* and *soft syntactic label* features, which are two models using syntactical information of the English target side, are applied as described in [16]. The motivation to add these models to the Jane system is to improve the reordering further and to obtain a more grammatically correct translation. The linguistic information necessary for these models was extracted by applying the Stanford parser<sup>1</sup> to the English target sentences.

### 2.3 System Combination

For the Chinese-English subtask, we also submitted results generated by our system combination framework. System combination is used to generate a consensus translation from multiple hypotheses produced with different translation engines, leading to a hypothesis which is better in terms of translation quality than any of the individual hypotheses. The basic concept of RWTH’s approach to machine translation system combination has been described by Matusov et al. [8, 9]. This approach includes an enhanced alignment and reordering framework. A lattice is built from the input hypotheses. The translation with the best score within the lattice according to some statistical models is then selected as the consensus translation.

### 2.4 Language Models

All language models are standard  $n$ -gram language models trained with the SRI toolkit [17] using interpolated modified Kneser-Ney smoothing. For both language pairs, we trained a language model on the target side of the bilingual data. For the Japanese-English task, parts of the monolingual United States Patent and Trademark Office have been used. For the Chinese-English task, we use the three data sets *us2003*, *us2004* and *us2005* of the above corpus. We have not used the monolingual data from the Japan Patent Office as adding these corpora did not decrease the LM perplexity on the development corpus.

### 2.5 Categorization

To reduce the sparseness of the training data in both tasks, four different categories (URLs, numbers, dates, hours) are introduced. Each word in the training data fitting into one of the categories is replaced by a unique category symbol. After the translation process, the symbol is again replaced by the original value. Chinese numerals are converted into Arabic numerals with a rule-based script.

## 3. JAPANESE-ENGLISH

### 3.1 Preprocessing

The segmentation of the Japanese text was done using the publicly available MeCab toolkit<sup>2</sup>. MeCab generates a very fine-grained tokenization, especially in the case of verbs, often splitting the verb ending into several tokens. This can sometimes lead to problems during decoding, because in reordering these tokens can be independently moved to different positions in the sentence. We therefore tried a more coarse-grained tokenization by automatically merging verb endings to a single token with a rule-based script. Moreover, all forms of the copula “である” (“to be”) and the modal verb “する” (“to do”) were merged into a single token as well.

<sup>1</sup><http://nlp.stanford.edu/software/lex-parser.shtml>

<sup>2</sup><http://mecab.sourceforge.net/>

MeCab	merged endings
設け・られ・て・いる	設け・られている
要求・さ・れ・た	要求・された
必要・で・ある	必要・である

Figure 1: The MeCab standard tokenization vs. the coarser “merged endings” tokenization

In the experiments, we refer to this tokenization as *merged endings*. See Figure 1 for some examples of the different tokenization schemes.

The katakana script is partly used to transcribe loanwords from other languages in Japanese. In the patent domain, there are many English technical terms which are transcribed in katakana. In the training data, about 8% of the tokens are katakana words. However, while the English terms may consist of several words, e.g. “clump cutter”, the Japanese transcription in the patent data was usually written as a single word, e.g. “クランプカッター” without any separation mark (“ ”). The MeCab segmenter does not automatically split these compound words.

For machine translation of German, [7] describes a frequency-based compound splitting method. We adapted this method to perform compound splitting for Japanese katakana words. We only allowed the splitting if each component has a length of at least two characters. This leads to improved word alignments, as the English technical terms and their transcriptions in Japanese have the same number of words. Further, the out-of-vocabulary (OOV) rate is reduced, because new compound terms consisting of known components can be translated. On the development data set, the number of OOVs is reduced from 178 to 122. We denote this preprocessing variant as *katakana split*.

Statistics of the training data with the different preprocessing variants are given in Table 1.

### 3.2 System setup

We use both a standard phrase-based (see Section 2.1) and a hierarchical system (see Section 2.2). GIZA++ is used to produce a word alignment for the preprocessed bilingual training data. From the word alignment we heuristically extract the standard or hierarchical phrase/rule table. We used the provided *pat-dev-2006-2007* data as development set (“*dev*”) to optimize the log-linear model parameters. As unseen test set (“*test*”) we used the NTCIR-8 intrinsic evaluation data set. The language model is a 4-gram trained only on the bilingual data. An additional language model, denoted as *usLM*, is a 4-gram trained on the bilingual data and the monolingual data sets *us2003* and *us2005*.

### 3.3 Experimental Results

Based on our observations in previous experiments [6], we chose 4BLEU-TER as the optimization criterion for the phrase-based system, as this leads to a more stable optimization. For the hierarchical system, we used the standard BLEU criterion, as 4BLEU-TER led to a degradation of performance in this case.

The experimental results are shown in Table 2. While we cannot observe significant changes in performance between the different preprocessing schemes, the combination of both *merged endings* and *katakana split* led to the best results. Using the larger language model (*usLM*) leads to another small improvement.

The clearest observation from our results is that the hierarchical paradigm is strongly superior to the standard phrase-based system with a difference of 2.6 in BLEU on *test*. One of the reasons is the substantial difference in the word order between Japanese and

bilingual corpora	Japanese				English
	MeCab	merged endings	katakana split	merged endings + katakana split	
Sentences	3,172,464				
Running Words	113,517,693	108,466,479	114,129,980	109,064,806	109,920,763
Vocabulary	150,753	150,927	122,144	122,295	112,214

Table 1: Corpus statistics for the different Japanese preprocessings of the bilingual training data.

Japanese→English	opt criterion	dev		test	
		BLEU	TER	BLEU	TER
Jane +merged endings +katakana split +syntax	BLEU	28.9	64.7	30.4	63.5
Jane +merged endings +katakana split	BLEU	28.8	64.4	30.3	63.4
Jane +katakana split	BLEU	28.4	65.3	30.2	64.1
Jane +merged endings	BLEU	27.7	66.0	29.6	64.5
Jane	BLEU	28.5	65.0	30.1	63.6
PBT +merged endings +katakana split +usLM	4BLEU -TER	25.7	65.2	27.8	63.3
PBT +merged endings +katakana split	4BLEU -TER	25.5	65.4	27.7	63.7
PBT +katakana split	4BLEU -TER	25.4	65.4	27.5	64.1
PBT +merged endings	4BLEU -TER	25.0	65.0	27.3	63.7
PBT	4BLEU -TER	25.2	65.5	27.4	64.0

Table 2: RWTH systems for the NTCIR-9 Japanese-English Patent translation task (truecase). PBT is the standard phrase-based system, Jane the hierarchical system. BLEU and TER results are in percentage.

frequency	Japanese	English
37	$X_0$ を $X_1$	$X_1 X_0$
34	$X_0$ に $X_1$	$X_1 X_0$
26	$X_0$ の $X_1$	$X_1 X_0$
23	$X_0$ は $X_1$	$X_0 X_1$
14	, $X_0$ の	of the $X_0$
12	$X_0$ に示すよう	as shown in $X_0$
12	$X_0$ する $X_1$	$X_1 X_0$
12	$X_0$ は $X_1$	$X_1 X_0$
11	$X_0$ した $X_1$	$X_1 X_0$
11	$X_0$ には $X_1$	$X_0 X_1$
11	$X_0$ の	of the $X_0$
10	$X_0$ , $X_1$ ように	$X_0$ , as $X_1$ ,
9	図 $X_0$ は,	FIG. $X_0$ is a
8	また, $X_0$ と $X_1$	the $X_0$ and the $X_1$
8	に $X_0$	$X_0$ to
8	の $X_0$	$X_0$ of
8	の $X_0$ に	to the $X_0$ of

Table 3: Excerpt of the most frequent hierarchical rules used in translation of the test set.

English. From looking at the phrase table, we can see that the hierarchical rules are very well suited to deal with this difference in word order and reorder whole phrases based on particles such as は, の, を, に, etc., which mark the end of these phrases. Table 3 shows some of the most frequent hierarchical rules used to translate *test*. The three topmost rules reorder two adjacent phrases, where the first phrase is marked by the particle を, に or の.

The fact that the hierarchical rules can capture the long range dependencies between the Japanese and English language can be seen by taking a close look at the example sentence given in Figure 2. The Japanese sentence 「本発明は、半導体ウエハなど

No.	Japanese	English
1	.	.
2	$X_0$ は, $X_1$ に関する	$X_0$ relates to a $X_1$
3	本発明	the present invention
4	$X_0$ の $X_1$	$X_1 X_0$
5	研磨方法	polishing method
6	半導体 $X_0$ など $X_1$	$X_1$ a semiconductor $X_0$ or the like
7	を研磨するため	for polishing
8	ウエハ	wafer

Table 4: Rules used for translating the example sentence from Figure 2 with the hierarchical paradigm.

を研磨するための研磨方法に関する。」 is translated into “the present invention relates to a polishing method for polishing a semiconductor wafer or the like.” It is obvious that the word order of the hierarchical translation is much better than that of the phrase-based translation. Taking a look at the hierarchical rules used for this sentence shown in Table 4 and the phrase-based counterpart in Table 5, the reason becomes clear. Rules 2,4 and 6 account for long-distance relationships, which the standard phrase-based paradigm is unable to capture. Rule 2 moves the verb 関する to the correct position after the sentence topic / subject. The phrase-based system on the other hand has learned to overgenerate the verb with phrase 10. Rule 4 switches the order of the two adjoining clauses separated by the particle の. The phrase-based decoder keeps the original word order, which is incorrect in English in this case. Finally, rule 6 again performs a reordering of the auxiliary subclause を研磨するため meaning *for polishing* before its object 半導体ウエハ, meaning *semiconductor wafer*, which is the correct English word order. The phrase-based system again fails to reorder correctly.

source	本発明は、半導体ウェハなどを研磨するための研磨方法に関する。
phrase-based	the present invention relates to a semiconductor wafer or for polishing relates to a method of polishing .
hierarchical	the present invention relates to a polishing method for polishing a semiconductor wafer or the like .
reference	The present invention relates to a method for polishing a semiconductor wafer or the like .

Figure 2: Example sentence from *test*, comparing the hierarchical and the phrase-based translation system.

No.	Japanese	English
9	.	.
10	本発明は、半導体	the present invention
11	ウェハなど	relates to a semiconductor
12	を研磨するため	wafer or
13	の研磨方法に関する	for polishing
		relates to a method of polishing

Table 5: Phrases used for translating the example sentence from Figure 2 with the phrase-based paradigm.

bilingual corpora	Chinese	English
Sentences	992,519	
Running Words	41,249,103	42,651,202
Vocabulary	95,320	315,953

Table 6: Corpus statistics of the preprocessed bilingual training data for the RWTH systems for the NTCIR-9 Chinese-English subtask.

## 4. CHINESE-ENGLISH

### 4.1 System Setup

**Preprocessing** The preprocessing mainly consists of tokenization and categorization. The tokenization cleans up the data and separates punctuations from neighboring words so that they are individual tokens. For Chinese, tokenization also includes Chinese word segmentation. We use the LDC segmenter<sup>3</sup>. The categorization was done as described in Section 2.5.

**Corpus** Table 6 shows the statistics of the bilingual data used. We filtered out a small fraction with a mismatching source/target sentence length. The LM is built on the target side of the bilingual corpora. Table 7 shows the monolingual corpus statistics. We combine this monolingual data with the English side of the bilingual data to build a *big LM* ( we refer to the LM that only uses the English side of the bilingual corpora as *small LM* ). For the phrase-based decoder, we use a 6-gram LM, for the hierarchical system a 4-gram LM.

The organizer provided a development corpus with 2000 sentences. To speed up the system tuning, we randomly split it into two parts and use them as *development* and *test* corpora.

**Additional models** We utilize the following additional models in the log linear framework: The triplet lexicon model and the discriminative lexicon model [10], which take a wider context into account, and the discriminative reordering model [20] as well as the source decoding sequence model [2] which capture phrase order information.

### 4.2 System combination of bidirectional translation systems

<sup>3</sup>[http://projects ldc.upenn.edu/Chinese/LDC\\_ch.htm](http://projects ldc.upenn.edu/Chinese/LDC_ch.htm)

monolingual corpora	English running words
us2003	1,486,878,644
us2004	1,465,846,627
us2005	1,295,478,799

Table 7: Corpus statistics of the preprocessed monolingual training data for the RWTH systems for the NTCIR-9 Chinese-English subtask.

Generally speaking, system combination is used to combine hypotheses generated by several different translation systems. Ideally, these systems should utilize different translation mechanisms. For example, combination of a phrase-based SMT system, a hierarchical SMT system and a rule-based system usually leads to some improvements in translation quality. For the NTCIR-9 Patent MT Chinese-English task, the system combination was done as follows. We use both a phrase-based (see Section 2.1) and a hierarchical phrase-based decoder (see Section 2.2). For each of the decoders we do a bi-directional translation, which means the system performs standard direction decoding (left-to-right) and inverse direction decoding (right-to-left). We thereby obtain a total of four different translations.

To build the inverse direction system, we used exactly the same data as the standard direction system and simply reversed the word order of the bilingual corpora. For example, the bilingual sentence pair “今天<sub>1</sub>是<sub>2</sub>星期天<sub>3</sub>。||Today<sub>1</sub>is<sub>2</sub>Sunday<sub>3</sub>.” is now transformed to “。星期天<sub>3</sub>是<sub>2</sub>今天<sub>1</sub>|| Sunday<sub>3</sub>is<sub>2</sub>Today<sub>1</sub>”. With the inversed corpora, we then trained the alignment, the language model and our translation systems in the exactly same way as the normal direction system. For decoding, the test corpus is also reversed.

The idea of utilizing right-to-left decoding has been proposed by [19] and [3] where they try to combine the advantages of both of the left-to-right and right-to-left decoding with a bidirectional decoding method. We also try to reap benefits from two-direction decoding, however, we use a system combination to achieve this goal.

### 4.3 Experimental Results

The results are shown in Tables 8 and 9. According to the rules of this evaluation, each team must submit at least one translation using only the bilingual data. We therefore split the results into two tables: Table 8 shows the results using only the bilingual data, and Table 9 presents the system results when also using the monolingual data for LM training. From the scores we can see that the monolingual training data definitely helps the translation with around 1.5 points BLEU improvement and a decrease in TER of 1 point. The results also show that the inverse hypotheses differs a lot from the normal baseline systems. With the help of our in-house system combination approach (see Section 2.3), we combined these four different hypotheses. For the big language model we achieved an improvement of 0.2 points in BLEU and 0.5 points in TER compared to the best single system. For the small language model, the improvement was 0.5 points in BLEU compared to the best single

Chinese→English	opt criterion	dev		test	
		BLEU	TER	BLEU	TER
Jane	BLEU	35.4	51.1	33.8	52.1
Jane inverse	BLEU	35.4	49.6	34.4	50.4
PBT	BLEU	34.6	51.1	33.0	52.3
PBT inverse	BLEU	34.7	51.0	32.8	52.3
system combination	BLEU	36.4	48.6	34.9	50.4

Table 8: Systems for the Chinese-English patent task using a small language model (Truecase results, BLEU and TER results are in percentage)

Chinese→English	opt criterion	dev		test	
		BLEU	TER	BLEU	TER
Jane	BLEU	37.3	48.2	35.7	49.8
Jane inverse	BLEU	37.2	48.1	36.3	48.9
PBT	BLEU	36.1	49.7	34.9	50.4
PBT inverse	BLEU	35.7	50.1	34.3	51.2
system combination	BLEU	37.2	47.9	36.5	48.4

Table 9: Systems for the Chinese-English patent task using a big language model (Truecase results, BLEU and TER results are in percentage)

system.

## 5. CONCLUSION

RWTH Aachen participated in the Japanese-to-English and the Chinese-to-English track of the NTCIR-9 PatentMT [4] task. Both the hierarchical and the phrase-based translation paradigm were used. Several different techniques were utilized to improve the respective baseline systems. Among them are *merged endings* and *KatakanaSplit* for the Japanese preprocessing, using additional monolingual data to build LMs, syntactic models for the hierarchical system and a system combination of bidirectional systems for the Chinese-English subtask. In this way, RWTH was able to achieve the 2nd place in the Japanese-English and the 3rd place in Chinese-English task with regard to the automatic BLEU measure.

## Acknowledgments

This work was achieved as part of the Quaero Programme, funded by OSEO, French State agency for innovation.

## 6. REFERENCES

- [1] D. Chiang. Hierarchical Phrase-Based Translation. *Computational Linguistics*, 33(2):201–228, 2007.
- [2] M. Feng, A. Mauser, and H. Ney. A source-side decoding sequence model for statistical machine translation. In *Proceedings of the Conference of the Association for Machine Translation in the Americas 2010 (AMTA 2010)*, Denver, Colorado, USA, Oct. 2010.
- [3] A. Finch and E. Sumita. Bidirectional phrase-based statistical machine translation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1124–1132, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [4] I. Goto, B. Lu, K. P. Chow, E. Sumita, and B. K. Tsou. Overview of the patent machine translation task at the ntcir-9 workshop. In *Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, NTCIR-9, 2011.
- [5] L. Huang and D. Chiang. Forest rescoring: Faster decoding with integrated language models. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 144–151, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [6] M. Huck, J. Wuebker, C. Schmidt, M. Freitag, S. Peitz, D. Stein, A. Dagnelies, S. Mansour, G. Leusch, and H. Ney. The rwth aachen machine translation system for wmt 2011. In *EMNLP 2011 Sixth Workshop on Statistical Machine Translation*, pages 405–412, Edinburgh, UK, July 2011.
- [7] P. Koehn and K. Knight. Empirical Methods for Compound Splitting. In *Proceedings of European Chapter of the ACL (EACL 2009)*, pages 187–194, 2003.
- [8] E. Matusov, G. Leusch, R. Banchs, N. Bertoldi, D. Dechelotte, M. Federico, M. Kolss, Y.-S. Lee, J. Marino, M. Paulik, S. Roukos, H. Schwenk, and H. Ney. System Combination for Machine Translation of Spoken and Written Language. *IEEE Transactions on Audio, Speech and Language Processing*, 16(7):1222–1237, 2008.
- [9] E. Matusov, N. Ueffing, and H. Ney. Computing Consensus Translation from Multiple Machine Translation Systems Using Enhanced Hypotheses Alignment. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 33–40, 2006.
- [10] A. Mauser, S. Hasan, and H. Ney. Extending Statistical Machine Translation with Discriminative and Trigger-Based Lexicon Models. In *Conference on Empirical Methods in Natural Language Processing*, pages 210–217, 2009.
- [11] J. Nelder and R. Mead. The Downhill Simplex Method. *Computer Journal*, 7:308, 1965.
- [12] F. Och. Minimum Error Rate Training for Statistical Machine Translation. In *Proc. Annual Meeting of the Association for Computational Linguistics*, pages 160–167, Sapporo, Japan, July 2003.

- [13] F. Och and H. Ney. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51, 2003.
- [14] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002.
- [15] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, pages 223–231, Cambridge, Massachusetts, USA, August 2006.
- [16] D. Stein, S. Peitz, D. Vilar, and H. Ney. A Cocktail of Deep Syntactic Features for Hierarchical Machine Translation. In *Conference of the Association for Machine Translation in the Americas 2010*, page 9, Denver, USA, Oct. 2010.
- [17] A. Stolcke. SRILM - an extensible language modeling toolkit. In *Proc. Int. Conf. on Spoken Language Processing*, volume 2, pages 901 – 904, Denver, Colorado, USA, Sept. 2002.
- [18] D. Vilar, S. Stein, M. Huck, and H. Ney. Jane: Open Source Hierarchical Translation, Extended with Reordering and Lexicon Models. In *ACL 2010 Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*, pages 262–270, Uppsala, Sweden, July 2010.
- [19] T. Watanabe and E. Sumita. Bidirectional decoding for statistical machine translation. In *Proceedings of the 19th international conference on Computational linguistics - Volume 1, COLING '02*, pages 1–7, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- [20] R. Zens and H. Ney. Discriminative Reordering Models for Statistical Machine Translation. In *Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics Annual Meeting (HLT-NAACL), Workshop on Statistical Machine Translation*, pages 55–63, New York City, June 2006.
- [21] R. Zens and H. Ney. Improvements in Dynamic Programming Beam Search for Phrase-based Statistical Machine Translation. In *Proc. of the Int. Workshop on Spoken Language Translation (IWSLT)*, Honolulu, Hawaii, Oct. 2008.