The RWTH Aachen System for NTCIR-10 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

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 10th NTCIR Workshop. Both phrase-based and hierarchical SMT systems were trained for the Japanese-English and Chinese-English tasks. Experiments were conducted to compare standard and inverse direction decoding, the performance of several additional models and the addition of monolingual training data. Moreover, 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

MeCab

1. INTRODUCTION

This is the RWTH Aachen University system paper for the Patent Translation Task of the 10th NTCIR Workshop [7]. We submitted results for the two subtasks: Japanese-English and Chinese-English. For Japanese-English, the submission is a standard phrase-based translation system, which is augmented with a word class language model and a hierarchical reordering model (HRM). Adding the HRM leads to very substantial improvements of 2.9% BLEU and 3.5% RIBES. Further, the preprocessing was improved compared to last year's submission by reducing inconsistent categories. The Chinese-English submission is a system combination of four systems: left-to-right hierarchical, right-to-left hierarchical, left-to-right phrase-based and right-to-left phrase-based systems. Here, we rebuilt the preprocessing pipeline from scratch to match the requirements of the Patent Translation Task.

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-10 Patent Translation evaluation RWTH's state-of-the-art phrase-based and hierarchical translation systems as well as our in-house system combination framework are utilized. We employ GIZA++ [14] to train word alignments. All systems are evaluated using the automatic BLEU [15] and TER [16] metrics. For Japanese-English we also score with RIBES [9].

2.1 Phrase-Based System

RWTH uses two different phrase-based translation systems. One is an in-house system similar to the one described in [23], which we denote as PBT, and one is the phrase-based decoder implemented in the open source toolkit Jane [20], which we denote as SCSS. In both systems, 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 sets further include an n-gram language model (LM), phrase-level IBM-1 and word-, phrase- and distortion-penalties. In PBT, parameters are optimized with the downhill simplex algorithm [12] on the word graphs. The SCSS decoder, on the other hand, optimizes with MERT [13] on n-best lists. It contains some extended features, including a word class language model, a language model look-ahead [21] which speeds up decoding and a hierarchical lexicalized reordering model [6].

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 is carried out using the cube pruning algorithm [8].

The standard models integrated into the hierarchical 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 and 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 [13] on 100-best lists.

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 is described in [2], which details RWTH's joint submission with SYSTRAN to the 10th NTCIR Workshop. 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 the constrained track of both language pairs, we trained a language model on the target side of the bilingual data. For the unconstrained Japanese-English task, RWTH performed data selection based on [11] on the United States Patent and Trademark Office data and the data published by the Japan Patent Office. The best $\frac{1}{8}$ of the *us2003*, *us2004* and *us2005* as well as the best $\frac{1}{20}$ of the Japan Patent office data of the years 1993 through 2002 were selected. Further, RWTH trained a 7-gram word class language model with 500 word classes on the target side of the bilingual training data.

For the Chinese-English task, the best $\frac{1}{2}$ of data sets *us2003*, *us2004* and *us2005* of the United States Patent and Trademark Office corpus as well as the best $\frac{1}{10}$ of the Japan Patent office data of the years 1993 through 2002 were selected.

2.5 Categorization

To reduce the sparseness of the training data in both tasks, three different categories (written numbers, digital numbers and ordinals) are introduced. Each word in the training data fitting into one of the categories is replaced by a unique category symbol using rule-based scripts. The advantage of this method is that the phrase-based system can learn more general rules containing these category symbols instead of phrases containing concrete numbers. When translating the test set, the symbol is again replaced by the original value at the end of the translation process. Here, Chinese numerals are converted into Arabic numerals with a rule-based script, and Japanese numbers (e.g. "三つ") are replaced by written English numbers ("three"). An example of the applied categories for Japanese and English can be seen in Table 1.

Japanese	English
$number { 305 }	$number { 305 }
$written { 三つ }	$written { three }
$ordinal { 第二 }	$ordinal { second }

Table 1: Categorization of written numbers, digital numbers and ordinals

3. JAPANESE-ENGLISH

3.1 Preprocessing

The Japanese text was segmented into words using the publicly available MeCab toolkit[1].

One problem of the categorization method described in Section 2.5 are inconsistent categorizations. They occur when the Japanese categorization does not generate the same categories as the English categorization. The two most frequent inconsistencies are the following:

- a number is written as a digit in Japanese and as a written word in English (e.g. "１種以上" vs. "more than one").
- one language uses a number, while the other uses a different expression (e.g. "六角" vs. "hexagonal", "一本のアイソレータ" vs. "a single isolator").

We applied a simple rule-based harmonization of the categories in both languages and removed inconsistent categorizations. The harmonization scheme removes a category in one sentence if it does not occur in its translation. Moreover, it replaces a number on the Japanese source side by a written number if the written form exists in the English sentence.

For some examples of the changes made see Table 2. The harmonized categories brought an improvement of 1.3 BLEU and 2.5 RIBES over last year's best constrained system.

Before harmonization	After harmonization
$written{六}角ナット hexagonal nut	六角ナット hexagonal nut
もう$written{一つ} another	もう一つ another
$number{１}以上の層 at least $written{ one } layer	$written {one}以上の層 at least $written{ one } layer
初回 $ordinal { first } time	初回 first time

Table 2: A monolingual categorization can lead to inconsistent categories between Japanese and English. A bilingual harmonization removes these inconsistencies.

[1] http://mecab.sourceforge.net/

Similar to our system for the NTCIR-9 evaluation, we used a frequency-based compound splitting method to split words written in katakana. This is especially helpful to reduce the number of unknown loanwords in the test and evaluation set. For a detailed description of this technique refer to our submission to the NTCIR-9 Workshop [4]. Statistics of the training data after preprocessing are given in Table 3.

bilingual corpus	Japanese	English
Sentences	3,172,273	
Running Words	113,684,101	108,438,989
Vocabulary	156,638	117,960

Table 3: Corpus statistics for the Japanese-English bilingual training data.

3.2 System setup

We use both the standard phrase-based (see Section 2.1) and the hierarchical system (see Section 2.2) implemented in the Jane toolkit. 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 LM trained only on the bilingual data. An additional language model, denoted as *unconstrainedLM*, is a 4-gram LM trained on the bilingual data and a selection of the monolingual data sets as described in Section 2.4. The selected data has a total size of 739M running words. We experimented with both standard direction decoding (left-to-right) and *inverse* direction decoding (right-to-left). The phrase-based decoder is extended with a 7-gram word class language model (*wcLM*) and a hierarchical reordering model (*HRM*).

3.3 Experimental Results

The experimental results are shown in Table 4. We observe a strong improvement compared to our submission to last year's NTCIR-9 Workshop. Further, we can see that the inverse translation direction leads to better translations than the standard direction, both for the phrase-based and the hierarchical paradigm. Adding the word class language model to the phrase-based decoder also yields a small improvement. The most substantial improvement is reached with the hierarchical reordering model, increasing the scores by 2.9% BLEU, 3.0% TER and 3.5% RIBES on the test set. While the hierarchical phrase-based decoder is in its baseline setting superior to the phrase-based decoder, with the addition of this reordering model the latter now clearly outperforms the former. Using the language model trained on additional monolingual data yields further gains on all metrics. Two examples of the improvements gained by the hierarchical reordering can be found in Table 5. Although the baseline system does not contain a hard distortion limit and thus allows for long-range reorderings, these reorderings are highly penalized. With the addition of the hierarchical reordering model, the scaling factor of this penalty was optimized to a value close to zero. Consequently, the hierarchical reordering method is able perform long range reorderings which the standard phrase-based system would not choose, as can also be seen in the two example sentences.

4. CHINESE-ENGLISH

4.1 New Preprocessing Pipeline

RWTH participated in the Chinese-English subtask of the NTCIR-9 PatentMT evaluation. The preprocessing toolkit used for NTCIR-9 was an old toolkit which was designed and optimized for the GALE project and NIST evaluation. After the NTCIR-9 evaluation we found lots of errors in the preprocessed corpus which possibly decreased the performance of the system. For NTCIR-10 we built a new Chinese-English preprocessing pipeline from scratch. The preprocessing framework is designed for the corpus provided by the PatentMT organizer.

For the Chinese side of the preprocessing, the following steps are performed:

1. Delete all spaces in the corpus
2. Separate English words
3. Change Chinese symbols into their English form
4. Generate the number category for arabic numbers
5. Chinese word segmentation
6. Generate number category for numbers written with Chinese characters
7. Correct some Chinese word segmentation
8. Translate the content of the category

For the Chinese word segmentation, Step 5 uses the longest word match strategy. First of all, we do a left-to-right longest word match. This algorithm reads the Chinese input sentence from left to right and looks up the longest dictionary entry that matches the input and then increments the start position before searching for the next longest matched dictionary entry. We build a lexicon from multiple resources and manually filter the dictionary in different ways. The final dictionary consists of 220K Chinese words. Secondly, we do a right-to-left longest word match using the same lexicon. We compare the segmentation results between left-to-right and right-to-left algorithms. The error patterns can be derived from the highly frequent inconsistent segmentation between left-to-right and right-to-left. In Step 7, we correct the left-to-right segmentation results based on the error patterns.

For the English side of the preprocessing, the following steps are performed:

1. Tokenization
2. Process hyphens
3. Make number category
4. Recase the first word of every sentence

For hyphens, we first treat all hyphens as single tokens. In the patent documents, there are large amount of chemical formulars such as CH2-CH4. In this case, the hyphen will not be processed as a single token. In other words, the hyphen is part of the token CH2-CH4. For the English side of the corpus, the patent documents contain large amounts of words which are capitalized or all capitalized. Hence, we only recase the first word of every sentence based on frequency.

Japanese→English	dev			test		
	BLEU	TER	RIBES	BLEU	TER	RIBES
SCSS (2011 constrained submission)	25.2	64.9	66.4	27.7	63.7	66.9
SCSS	27.5	62.3	68.5	29.0	60.0	69.4
SCSS inverse	27.6	62.3	68.4	29.5	59.3	70.1
SCSS inverse +wcLM	27.9	62.2	68.7	29.9	59.1	70.6
SCSS inverse +wcLM +HRM	30.6	59.4	71.9	32.8	56.1	74.1
SCSS inverse +wcLM +HRM +unconstrainedLM	31.5	58.5	72.5	33.3	55.7	74.6
HPBT	28.9	62.2	67.8	30.3	60.2	68.5
HPBT inverse	29.2	61.7	69.1	30.7	59.1	70.6

Table 4: RWTH systems for the NTCIR-10 Japanese-English Patent translation task (truecase). SCSS is the phrase-based system, HPBT the hierarchical system, which are both part of the Jane toolkit. All results are in percentage. The word class language model is denoted with *wcLM*, the hierarchical reordering model with *HRM* and the language model trained on additional monolingual data with *unconstrainedLM*.

source	駆動 ギア 偏心 成分 補正 プロ ファ イル 502 の 具体 的 な 生成 方法 を 説明 する .
without hier. reordering	the drive gear 502 of the eccentric component correction profile generating method will be described .
with hier. reordering	a specific method for generating a driving gear eccentric component correction profile 502 will be described .
reference	a specific method of generating the driving gear eccentricity component correction profile 502 will be described .
source	検出 パラ メータ は , 検査 レシピ 候補 毎 に 格納 さ れる .
without hier. reordering	detection parameters for each candidate is stored in the inspection recipe .
with hier. reordering	the detection parameters are stored for each candidate inspection recipe .
reference	the detection parameters are stored for each recipe candidate .

Table 5: Two example sentences from the *test* set translated with the SCSS decoder, illustrating the effect of the hierarchical reordering model.

bilingual corpora	Chinese	English
Sentences	1,000,000	
Running Words	38,901,617	44,521,254
Vocabulary	131,755	227,968

Table 6: Corpus statistics of the preprocessed bilingual training data for the RWTH systems for the NTCIR-10 Chinese-English subtask.

monolingual corpora	English running words
us2003	2,717,100,542
us2004	2,708,798,538
us2005	2,115,055,043
tcdata	2,423,893,417

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

4.2 System Setup

Table 6 shows the statistics of the bilingual data used. We used all bilingual data provided by the PatentMT organizer. The LM is built on the target side of the bilingual corpora. Data selection based on [11] is applied. Table 7 shows statistics of the selected monolingual corpus. We combine this selected monolingual data with the English side of the bilingual data to build a big LM. We build a 6-gram LM with modified Kneser-Ney discounting using SRILM [17] for the Chinese-English systems. The distortion limit for phrase-based SMT systems is set to 30.

We tune our systems on the development corpus provided by the organizers, which has 2000 sentences with single references. The test corpus is the evaluation corpus used for NTCIR-9 Chinese-English PatentMT task which also has 2000 sentences with single references.

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 [22] as well as the source decoding sequence model [3] which capture phrase order information.

4.3 System combination of bidirectional translation systems

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, a 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-10 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 a standard direction decoding (left-to-right) and an 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␣是␣星期天␣。‖Today␣is␣Sunday␣.” is now transformed to “。␣星期天␣是␣今天 ‖ .␣Sunday␣is␣Today”. With the inversed corpora, we then trained the alignment, the language model and our translation systems in the exact 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 [5], where the authors try to combine the advantages of both 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.

4.4 Experimental Results

The results are shown in Table 8. From the scores we can see that the difference between hierarchical phrase-based decoder (HPBT) and phrase-based decoder (PBT) is quite small. For the test corpus, PBT is even 0.1 BLEU and 0.1 TER better than HPBT. The results also show that the inverse hypotheses differs a lot from the normal baseline systems. For hierarchical phrase-based system, the inverse HPBT is 1.1 BLEU and 1.7 TER better than the standard HPBT for the test corpus. For phrase-based system, the inverse PBT is 0.6 BLEU and 0.5 TER worse than the standard PBT for the test corpus. The best single system is the inverse hierarchical phrase-based system (inverse HBPT). With the help of our in-house system combination framework described in [2], we combined these four different hypotheses. The last row in Table 8 shows an improvement of 0.8 points in BLEU and 0.2 points in TER on test corpus compared to the best single system.

Systems	dev		test	
	BLEU	TER	BLEU	TER
HPBT	43.0	44.1	39.8	45.3
HPBT inverse	43.3	42.8	40.9	43.6
PBT	42.3	44.3	39.9	45.2
PBT inverse	42.1	44.7	39.3	45.7
system combination	44.2	42.5	41.7	43.4

Table 8: Experimental results for the Chinese-English PatentMT task. HPBT is the hierarchical phrase-based translation system and PBT means the phrase-based translation system.

5. CONCLUSION

RWTH Aachen participated in the Japanese-to-English and the Chinese-to-English track of the NTCIR-10 PatentMT task. For both tasks, we improved the preprocessing pipeline as compared to the 2011 submissions [4]. We experimented with both the hierarchical and the phrase-based translation paradigm and tested several different techniques in order to improve the respective baseline systems. In Japanese-English, the baseline is already substantially better than our 2011 submission. Further, the hierarchical reordering model led to strong improvements in all three metrics. For the Chinese-English subtask, we improve the baseline by a system combination of bidirectional systems.

In this way, RWTH was able to achieve the 1st 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, M. Freitag, H. Ney, B. Buschbeck, J. Senellart, and J. Yang. The System Combination RWTH Aachen - SYSTRAN for the NTCIR-10 PatentMT Evaluation. In *Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, NTCIR-10, 2013.

[3] 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.

[4] M. Feng, C. Schmidt, J. Wuebker, S. Peitz, M. Freitag, and H. Ney. The RWTH Aachen System for NTCIR-9 PatentMT. In *Ninth NTCIR Workshop Meeting*, pages 600–605, Tokyo, Japan, Dec. 2011.

[5] 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.

[6] M. Galley and C. D. Manning. A Simple and Effective Hierarchical Phrase Reordering Model. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 848–856, Stroudsburg, PA, USA, 2008.

[7] I. Goto, K. P. Chow, B. Lu, E. Sumita, and B. K. Tsou. Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop. In *Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*, NTCIR-10, 2013.

[8] 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.

[9] H. Isozaki, T. Hirao, K. Duh, K. Sudoh, and H. Tsukada. Automatic evaluation of translation quality for distant language pairs. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 944–952, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[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] R. Moore and W. Lewis. Intelligent Selection of Language Model Training Data. In *ACL (Short Papers)*, pages 220–224, Uppsala, Sweden, July 2010.

[12] J. Nelder and R. Mead. The Downhill Simplex Method. *Computer Journal*, 7:308, 1965.

[13] 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.

[14] F. Och and H. Ney. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51, 2003.

[15] 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.

[16] 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.

[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] J. Wuebker, M. Huck, S. Peitz, M. Nuhn, M. Freitag, J.-T. Peter, S. Mansour, and H. Ney. Jane 2: Open Source Phrase-based and Hierarchical Statistical Machine Translation. In *International Conference on Computational Linguistics*, pages 483–491, Mumbai, India, Dec. 2012.

[21] J. Wuebker, H. Ney, and R. Zens. Fast and Scalable Decoding with Language Model Look-Ahead for Phrase-based Statistical Machine Translation. In *Annual Meeting of the Assoc. for Computational Linguistics*, pages 28–32, Jeju, Republic of Korea, July 2012.

[22] 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.

[23] 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.