Measure Word Generation for English-Chinese SMT Systems

Dongdong Zhang[1], Mu Li[1], Nan Duan[2], Chi-Ho Li[1], Ming Zhou[1]

[1]Microsoft Research Asia   [2]Tianjin University
Beijing, China   Tianjin, China
`{dozhang,muli,v-naduan,chl,mingzhou}@microsoft.com`

Abstract

Measure words in Chinese are used to indicate the count of nouns. Conventional statistical machine translation (SMT) systems do not perform well on measure word generation due to data sparseness and the potential long distance dependency between measure words and their corresponding head words. In this paper, we propose a statistical model to generate appropriate measure words of nouns for an English-to-Chinese SMT system. We model the probability of measure word generation by utilizing lexical and syntactic knowledge from both source and target sentences. Our model works as a post-processing procedure over output of statistical machine translation systems, and can work with any SMT system. Experimental results show our method can achieve high precision and recall in measure word generation.

1 Introduction

In linguistics, *measure words* (MW) are words or morphemes used in combination with numerals or demonstrative pronouns to indicate the count of nouns[1], which are often referred to as *head words* (HW).

Chinese measure words are grammatical units and occur quite often in real text. According to our survey on the measure word distribution in the Chinese Penn Treebank and the test datasets distributed by Linguistic Data Consortium (LDC) for Chinese-to-English machine translation evaluation, the average occurrence is 0.505 and 0.319 measure words per sentence respectively. Unlike in Chinese, there is no special set of measure words in English. Measure words are usually used for mass nouns and any semantically appropriate nouns can function as the measure words. For example, in the phrase *three bottles of water*, the word *bottles* acts as a measure word. Countable nouns are almost never modified by measure words[2]. Numerals and indefinite articles are directly followed by countable nouns to denote the quantity of objects.

Therefore, in the English-to-Chinese machine translation task we need to take additional efforts to generate the missing measure words in Chinese. For example, when translating the English phrase *three books* into the Chinese phrases "三本书", where *three* corresponds to the numeral "三" and *books* corresponds to the noun "书", the Chinese measure word "本" should be generated between the numeral and the noun.

In most statistical machine translation (SMT) models (Och et al., 2004; Koehn et al., 2003; Chiang, 2005), some of measure words can be generated without modification or additional processing. For example, in above translation, the phrase translation table may suggest the word *three* be translated into "三", "三本", "三只", etc, and the word *books* into "书", "书本", "名册" (scroll), etc. Then the SMT model selects the most likely combination "三本书" as the final translation result. In this example, a measure word candidate set consisting of "本" and "只" can be generated by bilingual phrases (or synchronous translation rules), and the best measure word "本" from the measure

---

[1] The uncommon cases of verbs are not considered.

[2] There are some exceptional cases, such as "*100 head of cattle*". But they are very uncommon.

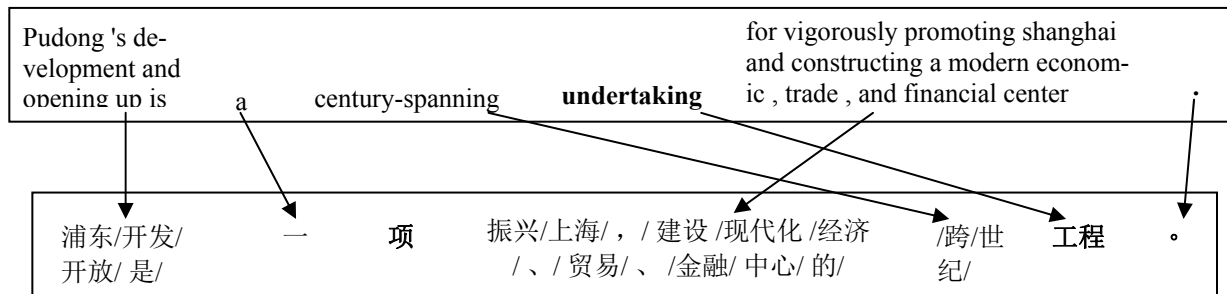浦东/开发/开放/ 是/ | 一 | 项 | 振兴/上海/，/ / 建设 /现代化 /经济 / 、/ 贸易/ 、/金融/ 中心 的/ | /跨/世纪/ | 工程 | 。

Figure 1. Example of long distance dependency between MW and its modified HW

word candidate set can be selected by the SMT decoder. However, as we will show below, existing SMT systems do not deal well with the measure word generation in general due to data sparseness and long distance dependencies between measure words and their corresponding head words.

Due to the limited size of bilingual corpora, many measure words, as well as the collocations between a measure and its head word, cannot be well covered by the phrase translation table in an SMT system. Moreover, Chinese measure words often have a long distance dependency to their head words which makes language model ineffective in selecting the correct measure words from the measure word candidate set. For example, in Figure 1 the distance between the measure word "项" and its head word "工程" (undertaking) is 15. In this case, an n-gram language model with $n<15$ cannot capture the MW-HW collocation. Table 1 shows the relative position's distribution of head words around measure words in the Chinese Penn Treebank, where a negative position indicates that the head word is to the left of the measure word and a positive position indicates that the head word is to the right of the measure word. Although lots of measure words are close to the head words they modify, more than sixteen percent of measure words are far away from their corresponding head words (the absolute distance is more than 5).

To overcome the disadvantage of measure word generation in a general SMT system, this paper proposes a dedicated statistical model to generate measure words for English-to-Chinese translation.

We model the probability of measure word generation by utilizing rich lexical and syntactic knowledge from both source and target sentences. Three steps are involved in our method to generate measure words: Identifying the positions to gener-

ate measure words, collecting the measure word candidate set and selecting the best measure word. Our method is performed as a post-processing procedure of the output of SMT systems. The advantage is that it can be easily integrated into any SMT system. Experimental results show our method can significantly improve the quality of measure word generation. We also compared the performance of our model based on different contextual information, and show that both large-scale monolingual data and parallel bilingual data can be helpful to generate correct measure words.

Position	Occurrence	Position	Occurrence
1	39.5%	-1	0
2	15.7%	-2	0
3	4.7%	-3	8.7%
4	1.4%	-4	6.8%
5	2.1%	-5	4.3%
>5	8.8%	<-5	8.0%

Table 1. Position distribution of head words

2 Our Method

2.1 Measure word generation in Chinese

In Chinese, measure words are obligatory in certain contexts, and the choice of measure word usually depends on the head word's semantics (e.g., shape or material). The set of Chinese measure words is a relatively close set and can be classified into two categories based on whether they have a corresponding English translation. Those not having an English counterpart need to be generated during translation. For those having English translations, such as "米" (meter), "吨" (ton), we just use the translation produced by the SMT system itself. According to our survey, about 70.4% of measure words in the Chinese Penn Treebank need

to be explicitly generated during the translation process.

In Chinese, there are generally stable linguistic collocations between measure words and their head words. Once the head word is determined, the collocated measure word can usually be selected accordingly. However, there is no easy way to identify head words in target Chinese sentences since for most of the time an SMT output is not a well formed sentence due to translation errors. Mistake of head word identification may cause low quality of measure word generation. In addition, sometimes the head word itself is not enough to determine the measure word. For example, in Chinese sentences "他家有 5 口人" (there are five people in his family) and "总共有 5 个人参加了会议" (a total of five people attended the meeting), where "人" (people) is the head word collocated with two different measure words "口" and "个", we cannot determine the measure word just based on the head word "人".

2.2 Framework

In our framework, a statistical model is used to generate measure words. The model is applied to SMT system outputs as a post-processing procedure. Given an English source sentence, an SMT decoder produces a target Chinese translation, in which positions for measure word generation are identified. Based on contextual information contained in both input source sentence and SMT system's output translation, a measure word candidate set M is constructed. Then a measure word selection model is used to select the best one from M. Finally, the selected measure word is inserted into previously determined measure word slot in the SMT system's output, yielding the final translation result.

2.3 Measure word position identification

To identify where to generate measure words in the SMT outputs, all positions after numerals are marked at first since measure words often follow numerals. For other cases in which measure words do not follow numerals (e.g., "许多/台/电脑" (many computers), where "台" is a measure word and "电脑" (computers) is its head word), we just mine the set of words which can be followed by measure words from training corpus. Most of

words in the set are pronouns such as "该" (this), "那" (that) and "若干" (several). In the SMT output, the positions after these words are also identified as candidate positions to generate measure words.

2.4 Candidate measure word generation

To avoid high computation cost, the measure word candidate set only consists of those measure words which can form valid MW-HW collocations with their head words. We assume that all the surrounding words within a certain window size centered on the given position to generate a measure word are potential head words, and require that a measure word candidate must collocate with at least one of the surrounding words. Valid MW-HW collocations are mined from the training corpus and a separate lexicon resource.

There is a possibility that the real head word is outside the window of given size. To address this problem, we also use a source window centered on the position p_s, which is aligned to the target measure word position p_t. The link between p_s and p_t can be inferred from SMT decoding result. Thus, the chance of capturing the best measure word increases with the aid of words located in the source window. For example, given the window size of 10, although the target head word "工程" (undertaking) in Figure 1 is located outside the target window, its corresponding source head word *undertaking* can be found in the source window. Based on this source head word, the best measure word "项" will be included into the candidate measure word set. This example shows how bilingual information can enrich the measure word candidate set.

Another special word {NULL} is always included in the measure word candidate set. {NULL} represents those measure words having a corresponding English translation as mentioned in Section 2.1. If {NULL} is selected, it means that we need not generate any measure word at the current position. Thus, no matter what kinds of measure words they are, we can handle the issue of measure word generation in a unified framework.

2.5 Measure word selection model

After obtaining the measure word candidate set M, a measure word selection model is employed to select the best one from M. Given the contextual information C in both source window and target

window, we model the measure word selection as finding the measure word $m*$ with highest posterior probability given C:

$$m^* = \text{argmax}_{m \in M} P(m|C) \quad (1)$$

To leverage the collocation knowledge between measure words and head words, we extend (1) by introducing a hidden variable h where H represents all candidate head words located within the target window:

$$m^* = \text{argmax}_{m \in M} \sum_{h \in H} P(m, h|C)$$
$$= \text{argmax}_{m \in M} \sum_{h \in H} P(h|C)P(m|h, C) \quad (2)$$

In (2), $P(h|C)$ is the head word selection probability and is empirically estimated according to the position distribution of head words in Table 1. $P(m|h, C)$ is the conditional probability of m given both h and C. We use maximum entropy model to compute $P(m|h, C)$:

$$P(m|h, C) = \frac{\exp(\sum_i \lambda_i f_i(m,C))}{\sum_{m' \in M} \exp(\sum_i \lambda_i f_i(m',C))} \quad (3)$$

Based on the different features used in the computation of $P(m|h, C)$, we can train two sub-models – a monolingual model (*Mo-ME*) which only uses monolingual (Chinese) features and a bilingual model (*Bi-ME*) which integrates bilingual features. The advantage of the Mo-ME model is that it can employ an unlimited monolingual target training corpora, while the Bi-ME model leverages rich features including both the source and target information and may improve the precision. Compared to the Mo-ME model, the Bi-ME model suffers from small scale of parallel training data. To leverage advantages of both models, we use a combined model *Co-ME*, by linearly combing the monolingual and bilingual sub-models:

$$m^* = \text{argmax}_{m \in M} \lambda P_{Mo-ME} + (1 - \lambda)P_{Bi-ME}$$

where $\lambda \in [0,1]$ is a free parameter that can be optimized on held-out data and it was set to 0.39 in our experiments.

2.6 Features

The computation of Formula (3) involves the features listed in Table 2 where the Mo-ME model only employs target features and the Bi-ME model leverages both target features and source features.

For target features, *n*-gram language model score is defined as the sum of log *n*-gram probabilities within the target window after the measure

word is filled into the measure word slot. The MW-HW collocation feature is defined to be a function f_1 to capture the collocation between a measure word and a head word. For features of surrounding words, the feature function f_2 is defined as 1 if a certain word exists at a certain position, otherwise 0. For example, $f_2(人,-2)=1$ means the second word on the left is "人". $f_2(书,3)=1$ means the third word on the right is "书". For punctuation position feature function f_3, the feature value is 1 when there is a punctuation following the measure word, which indicates the target head word may appear to the left of measure word. Otherwise, it is 0. In practice, we can also ignore the position part, i.e., a word appears anywhere within the window is viewed as the same feature.

Target features	Source features
n-gram language model score	MW-HW collocation
MW-HW collocation	surrounding words
surrounding words	source head word
punctuation position	POS tags

Table 2. Features used in our model

For source language side features, MW-HW collocation and surrounding words are used in a similar way as does with target features. The source head word feature is defined to be a function f_4 to indicate whether a word e_i is the source head word in English according to a parse tree of the source sentence. Similar to the definition of lexical features, we also use a set of features based on POS tags of source language.

3 Model Training and Application

3.1 Training

We parsed English and Chinese sentences to get training samples for measure word generation model. Based on the source syntax parse tree, for each measure word, we identified its head word by using a toolkit from (Chiang and Bikel, 2002) which can heuristically identify head words for sub-trees. For the bilingual corpus, we also perform word alignment to get correspondences between source and target words. Then, the collocation between measure words and head words and their surrounding contextual information are extracted to train the measure word selection models. According to word alignment results, we classify

measure words into two classes based on whether they have non-null translations. We map Chinese measure words having non-null translations to a unified symbol {NULL} as mentioned in Section 2.4, indicating that we need not generate these kind of measure words since they can be translated from English.

In our work, the Berkeley parser (Petrov and Klein, 2007) was employed to extract syntactic knowledge from the training corpus. We ran GIZA++ (Och and Ney, 2000) on the training corpus in both directions with IBM model 4, and then applied the refinement rule described in (Koehn et al., 2003) to obtain a many-to-many word alignment for each sentence pair. We used the SRI Language Modeling Toolkit (Stolcke, 2002) to train a five-gram model with modified Kneser-Ney smoothing (Chen and Goodman, 1998). The Maximum Entropy training toolkit from (Zhang, 2006) was employed to train the measure word selection model.

3.2 Measure word generation

As mentioned in previous sections, we apply our measure word generation module into SMT output as a post-processing step. Given a translation from an SMT system, we first determine the position p_t at which to generate a Chinese measure word. Centered on p_t, a surrounding word window with specified size is determined. From translation alignments, the corresponding source position p_s aligned to p_t can be referred. In the same way, a source window centered on p_s is determined as well. Then, contextual information within the windows in the source and the target sentence is extracted and fed to the measure word selection model. Meanwhile, the candidate set is obtained based on words in both windows. Finally, each measure word in the candidate set is inserted to the position p_t, and its score is calculated based on the models presented in Section 2.5. The measure word with the highest probability will be chosen.

There are two reasons why we perform measure word generation for SMT systems as a post-processing step. One is that in this way our method can be easily applied to any SMT system. The other is that we can leverage both source and target information during the measure word generation process. We do not integrate our measure word generation module into the SMT decoder since there is only little target contextual information available during SMT decoding. Moreover, as we will show in experiment section, a pre-processing method does not work well when only source information is available.

4 Experiments

4.1 Data

In the experiments, the language model is a Chinese 5-gram language model trained with the Chinese part of the LDC parallel corpus and the Xinhua part of the Chinese Gigaword corpus with about 27 million words. We used an SMT system similar to Chiang (2005), in which FBIS corpus is used as the bilingual training data. The training corpus for Mo-ME model consists of the Chinese Peen Treebank and the Chinese part of the LDC parallel corpus with about 2 million sentences. The Bi-ME model is trained with FBIS corpus, whose size is smaller than that used in Mo-ME model training.

We extracted both development and test data set from years of NIST Chinese-to-English evaluation data by filtering out sentence pairs not containing measure words. The development set is extracted from NIST evaluation data from 2002 to 2004, and the test set consists of sentence pairs from NIST evaluation data from 2005 to 2006. There are 759 testing cases for measure word generation in our test data consisting of 2746 sentence pairs. We use the English sentences in the data sets as input to the SMT decoder, and apply our proposed method to generate measure words for the output from the decoder. Measure words in Chinese sentences of the development and test sets are used as references. When there are more than one measure words acceptable at some places, we manually augment the references with multiple acceptable measure words.

4.2 Baseline

Our baseline is the SMT output where measure words are generated by a Hiero-like SMT decoder as discussed in Section 1. Due to noises in the Chinese translations introduced by the SMT system, we cannot correctly identify all the positions to generate measure words. Therefore, besides precision we examine recall in our experiments.

4.3 Evaluation over SMT output

Table 3 and Table 4 show the precision and recall of our measure word generation method. From the

experimental results, the Mo-ME, Bi-ME and Co-ME models all outperform the baseline. Compared with the baseline, the Mo-ME method takes advantage of a large size monolingual training corpus and reduces the data sparseness problem. The advantage of the Bi-ME model is being able to make full use of rich knowledge from both source and target sentences. Also as shown in Table 3 and Table 4, the Co-ME model always achieve the best results when using the same window size since it leverages the advantage of both the Mo-ME and the Bi-ME models.

Wsize	Baseline	Mo-ME	Bi-ME	Co-ME
6		64.29%	67.15%	67.66%
8		64.93%	68.50%	69.00%
10	54.82%	64.72%	69.40%	69.58%
12		65.46%	69.40%	69.76%
14		65.61%	69.69%	70.03%

Table 3. Precision over SMT output

Wsize	Baseline	Mo-ME	Bi-ME	Co-ME
6		51.48%	53.69%	54.09%
8		51.98%	54.75%	55.14%
10	45.61%	51.81%	55.44%	55.58%
12		52.38%	55.44%	55.72%
14		52.50%	55.67%	55.93%

Table 4. Recall over SMT output

We can see that the Bi-ME model can achieve better results than the Mo-ME model in both recall and precision metrics although only a small sized bilingual corpus is used for Bi-ME model training. The reason is that the Mo-ME model cannot correctly handle the cases where head words are located outside the target window. However, due to word order differences between English and Chinese, when target head words are outside the target window, their corresponding source head words might be within the source window. The capacity of capturing head words is improved when both source and target windows are used, which demonstrates that bilingual knowledge is useful for measure word generation.

We compare the results for each model with different window sizes. Larger window size can lead to better results as shown in Table 3 and Table 4 since more contextual knowledge is used to model measure word generation. However, enlarging the window size does not bring significant improvements, The major reason is that even a small win-

dow size is already able to cover most of measure word collocations, as indicated by the position distribution of head words in Table 1.

The quality of the SMT output also affects the quality of measure word generation since our method is performed in a post-processing step over the SMT output. Although translation errors degrade the measure word generation accuracy, we achieve about 15% improvement in precision and a 10% increase in recall over baseline. We notice that the recall is relatively lower. Part of the reason is some positions to generate measure words are not successfully identified due to translation errors. In addition to precision and recall, we also evaluate the Bleu score (Papineni et al., 2002) changes before and after applying our measure word generation method to the SMT output. For our test data, we only consider sentences containing measure words for Bleu score evaluation. Our measure word generation step leads to a Bleu score improvement of 0.32 where the window size is set to 10, which shows that it can improve the translation quality of an English-to-Chinese SMT system.

4.4 Evaluation over reference data

To isolate the impact of the translation errors in SMT output on the performance of our measure word generation model, we conducted another experiment with reference bilingual sentences in which measure words in Chinese sentences are manually removed. This experiment can show the performance upper bound of our method without interference from an SMT system. Table 5 shows the results. Compared to the results in Table 3, the precision improvement in the Mo-ME model is larger than that in the Bi-ME model, which shows that noisy translation of the SMT system has more serious influence on the Mo-ME model than the Bi-ME model. This also indicates that source information without noises is helpful for measure word generation.

Wsize	Mo-ME	Bi-ME	Co-ME
6	71.63%	74.92%	75.72%
8	73.80%	75.48%	76.20%
10	73.80%	74.76%	75.48%
12	73.80%	75.24%	75.96%
14	73.56%	75.48%	76.44%

Table 5. Results over reference data

4.5 Impacts of features

In this section, we examine the contribution of both target language based features and source language based features in our model. Table 6 and Table 7 show the precision and recall when using different features. The window size is set to 10. In the tables, *Lm* denotes the *n*-gram language model feature, *Tmh* denotes the feature of collocation between target head words and the candidate measure word, *Smh* denotes the feature of collocation between source head words and the candidate measure word, *Hs* denotes the feature of source head word selection, *Punc* denotes the feature of target punctuation position, *Tlex* denotes surrounding word features in translation, *Slex* denotes surrounding word features in source sentence, and *Pos* denotes Part-Of-Speech feature.

Feature setting	Precision	Recall
Baseline	54.82%	45.61%
Lm	51.11%	41.24%
+Tmh	61.43%	49.22%
+Punc	62.54%	50.08%
+Tlex	64.80%	51.87%

Table 6. Feature contribution in Mo-ME model

Feature setting	Precision	Recall
Baseline	54.82%	45.61%
Lm	51.11%	41.24%
+Tmh+Smh	64.50%	51.64%
+Hs	65.32%	52.26%
+Punc	66.29%	53.10%
+Pos	66.53%	53.25%
+Tlex	67.50%	54.02%
+Slex	69.52%	55.54%

Table 7. Feature contribution in Bi-ME model

The experimental results show that all the features can bring incremental improvements. The method with only *Lm* feature performs worse than the baseline. However, with more features integrated, our method outperforms the baseline, which indicates each kind of features we selected is useful for measure word generation. According to the results, the feature of MW-HW collocation has much contribution to reducing the selection error of measure words given head words. The contribution of *Slex* feature explains that other surrounding words in source sentence are also helpful since head word determination in source language

might be incorrect due to errors in English parse trees. Meanwhile, the contribution from *Smh*, *Hs* and *Slex* features demonstrates that bilingual knowledge can play an important role for measure word generation. Compared with lexicalized features, we do not get much benefit from the *Pos* features.

4.6 Error analysis

We conducted an error analysis on 100 randomly selected sentences from the test data. There are four major kinds of errors as listed in Table 8. Most errors are caused by failures in finding positions to generate measure words. The main reason for this is some hint information used to identify measure word positions is missing in the noisy output of SMT systems. Two kinds of errors are introduced by incomplete head word and MW-HW collocation coverage, which can be solved by enlarging the size of training corpus. There are also head word selection errors due to incorrect syntax parsing.

Error type	Ratio
unseen head word	32.14%
unseen MW-HW collocation	10.71%
missing MW position	39.29%
incorrect HW selection	10.71%
others	7.14%

Table 8. Error distribution

4.7 Comparison with other methods

In this section we compare our statistical methods with the pre-processing method and the rule-based methods for measure word generation in a translation task.

In pre-processing method, only source language information is available. Given a source sentence, the corresponding syntax parse tree T_s is first constructed with an English parser. Then the pre-processing method chooses the source head word h_s based on T_s. The candidate measure word with the highest probability collocated with h_s is selected as the best result, where the measure word candidate set corresponding to each head word is mined over a bilingual training corpus in advance. We achieved precision 58.62% and recall 49.25%, which are worse than the results of our post-processing based methods. The weakness of the pre-processing method is twofold. One problem is data sparseness with respect to collocations be-

tween English head words and Chinese measure words. The other problem comes from the English head word selection error introduced by using source parse trees.

We also compared our method with a well-known rule-based machine translation system – SYSTRAN[3]. We translated our test data with SYSTRAN's English-to-Chinese translation engine. The precision and recall are 63.82% and 51.09% respectively, which are also lower than our method.

5   Related Work

Most existing rule-based English-to-Chinese MT systems have a dedicated module handling measure word generation. In general a rule-based method uses manually constructed rule patterns to predict measure words. Like most rule based approaches, this kind of system requires lots of human efforts of experienced linguists and usually cannot easily be adapted to a new domain. The most relevant work based on statistical methods to our research might be statistical technologies employed to model issues such as morphology generation (Minkov et al., 2007).

6   Conclusion and Future Work

In this paper we propose a statistical model for measure word generation for English-to-Chinese SMT systems, in which contextual knowledge from both source and target sentences is involved. Experimental results show that our method not only achieves high precision and recall for generating measure words, but also improves the quality of English-to-Chinese SMT systems.

In the future, we plan to investigate more features and enlarge coverage to improve the quality of measure word generation, especially reduce the errors found in our experiments.

Acknowledgements

Special thanks to David Chiang, Stephan Stiller and the anonymous reviewers for their feedback and insightful comments.

References

Stanley F. Chen and Joshua Goodman. 1998. An Empirical study of smoothing techniques for language modeling. *Technical Report TR-10-98*, *Harvard University Center for Research in Computing Technology*, 1998.

David Chiang and Daniel M. Bikel. 2002. Recovering latent information in treebanks. *Proceedings of COLING '02*, 2002.

David Chiang. 2005. *A hierarchical phrase-based model for statistical machine translation*. In *Proceedings of ACL 2005*, pages 263–270.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT-NAACL 2003*, pages 127-133.

Einat Minkov, Kristina Toutanova, and Hisami Suzuki. 2007. Generating complex morphology for machine translation. In *Proceedings of 45th Annual Meeting of the ACL*, pages 128-135.

Franz J. Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of 38th Annual Meeting of the ACL*, pages 440-447.

Franz J. Och and Hermann Ney. 2004. *The alignment template approach to statistical machine translation*. Computational Linguistics, 30:417–449.

Kishore Papineni, Salim Roukos, ToddWard, and WeiJing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the ACL*, pages 311-318.

Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *Proceedings of HLT-NAACL*, 2007.

Andreas Stolcke. 2002. SRILM - an extensible language modeling toolkit. In *Proceedings of International Conference on Spoken Language Processing*, volume 2, pages 901-904.

Le Zhang. MaxEnt toolkit. 2006. http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html

---

[3] http://www.systransoft.com/