

# Head Finalization Reordering for Chinese-to-Japanese Machine Translation

Han Dan<sup>+</sup>   Katsuhito Sudoh<sup>‡</sup>   Xianchao Wu<sup>‡\*</sup>  
Kevin Duh<sup>‡†</sup>   Hajime Tsukada<sup>‡</sup>   Masaaki Nagata<sup>‡</sup>

<sup>+</sup>The Graduate University For Advanced Studies, Tokyo, Japan

<sup>‡</sup>NTT Communication Science Laboratories, NTT Corporation

<sup>+</sup>handan@nii.ac.jp, <sup>\*</sup>wuxianchao@baidu.com, <sup>†</sup>kevinduh@is.naist.jp

<sup>‡</sup>{sudoh.katsuhito, tsukada.hajime, nagata.masaaki}@lab.ntt.co.jp

## Abstract

In Statistical Machine Translation, reordering rules have proved useful in extracting bilingual phrases and in decoding during translation between languages that are structurally different. Linguistically motivated rules have been incorporated into Chinese-to-English (Wang et al., 2007) and English-to-Japanese (Isozaki et al., 2010b) translation with significant gains to the statistical translation system. Here, we carry out a linguistic analysis of the Chinese-to-Japanese translation problem and propose one of the first reordering rules for this language pair. Experimental results show substantial improvements (from 20.70 to 23.17 BLEU) when head-finalization rules based on HPSG parses are used, and further gains (to 24.14 BLEU) were obtained using more refined rules.

## 1 Introduction

In state-of-the-art Statistical Machine Translation (SMT) systems, bilingual phrases are the main building blocks for constructing a translation given a sentence from a source language. To extract those bilingual phrases from a parallel corpus, the first step is to discover the implicit word-to-word correspondences between bilingual sentences (Brown et al., 1993). Then, a symmetrization matrix is built (Och and Ney, 2004) by using word-to-word alignments, and a wide variety

of heuristics can be used to extract the bilingual phrases (Zens et al., 2002; Koehn et al., 2003).

This method performs relatively well when the source and the target languages have similar word order, as in the case of French, Spanish, and English. However, when translating between languages with very different structures, as in the case of English and Japanese, or Japanese and Chinese, the quality of extracted bilingual phrases and the overall translation quality diminishes.

In the latter scenario, a simple but effective strategy to cope with this problem is to reorder the words of sentences in one language so that it resembles the word order of another language (Wu et al., 2011; Isozaki et al., 2010b). The advantages of this strategy are two fold. The first advantage is at the decoding stage, since it enables the translation to be constructed almost monotonically. The second advantage is at the training stage, since automatically estimated word-to-word alignments are likely to be more accurate and symmetrization matrices reveal more evident bilingual phrases, leading to the extraction of better quality bilingual phrases and cleaner phrase tables.

In this work, we focus on Chinese-to-Japanese translation, motivated by the increasing interaction between these two countries and the need to improve direct machine translation without using a pivot language. Despite the countries' close cultural relationship, their languages significantly differ in terms of syntax, which poses a severe difficulty in statistical machine translation. The syntactic relationship of this language pair has not been carefully studied before in the machine translation

<sup>\*</sup>Now at Baidu Japan Inc.

<sup>†</sup> Now at Nara Institute of Science and Technology (NAIST)

field, and our work aims to contribute in this direction as follows:

- We present a detailed syntactic analysis of several reordering issues in Chinese-Japanese translation using the information provided by an HPSG-based deep parser.
- We introduce novel reordering rules based on head-finalization and linguistically inspired refinements to make words in Chinese sentences resemble Japanese word order. We empirically show its effectiveness (e.g. 20.70 to 24.23 BLEU improvement).

The paper is structured as follows. Section 2 introduces the background and gives an overview of similar techniques related to this work. Section 3 describes the proposed method in detail. Experimental evaluation of the performance of the proposed method is described in section 4. There is an error analysis on the obtained results in section 5. Conclusions and a short description on future work derived from this research are given in the final section.

## 2 Background

### 2.1 Head Finalization

The structure of languages can be characterized by phrase structures. The head of a phrase is the word that determines the syntactic category of the phrase, and its modifiers (also called dependents) are the rest of the words within the phrase. In English, the head of a phrase can be usually found before its modifiers. For that reason, English is called a head-initial language (Cook and Newson, 1988). Japanese, on the other hand, is head-final language (Fukui, 1992), since the head of a phrase always appears after its modifiers.

In certain applications, as in the case of machine translation, word reordering can be a promising strategy to ease the task when working with languages with different phrase structures like English and Japanese. Head Finalization is a successful syntax-based reordering method designed to reorder sentences from a head-initial language to resemble the word order in sentences from a head-final language (Isozaki et al., 2010b). The essence

of this rule is to move the syntactic heads to the end of its dependency by swapping child nodes in a phrase structure tree when the head child appears before the dependent child.

Isozaki et al. (2010b) proposed a simple method of Head Finalization, by using an HPSG-based deep parser for English (Miyao and Tsujii, 2008) to obtain phrase structures and head information. The score results from several mainstream evaluation methods indicated that the translation quality had been improved; the scores of Word Error Rate (WER) and Translation Edit Rate (TER) (Snover et al., 2006) had especially been greatly reduced.

### 2.2 Chinese Deep Parsing

Syntax-based reordering methods need parsed sentences as input. Isozaki et al. (2010b) used *Enju*, an HPSG-based deep parser for English, but they also discussed using other types of parsers, such as word dependency parsers and Penn Treebank-style parsers. However, to use word dependency parsers, they needed an additional heuristic rule to recover phrase structures, and Penn Treebank-style parsers are problematic because they output flat phrase structures (i.e. a phrase may have multiple dependents, which causes a problem of reordering within a phrase). Consequently, compared to different types of parsers, *Head-Final English* performs the best on the basis of English *Enju*'s parsing result.

In this paper, we follow their observation, and use the HPSG-based parser for Chinese (*Chinese Enju*) (Yu et al., 2011) for Chinese syntactic parsing. Since Chinese *Enju* is based on the same parsing model as English *Enju*, it provides rich syntactic information including phrase structures and syntactic/semantic heads.

Figure 1 shows an example of an XML output from Chinese *Enju* for the sentence “wo (I) qu (go to) dongjing (Tokyo) he (and) jingdu (Kyoto).” The label `<cons>` and `<tok>` represent the non-terminal nodes and terminal nodes, respectively. Each node is identified by a unique “id” and has several attributes. The attribute “head” indicates which child node is the syntactic head. In this figure, `<head=“c4” id=“c3”>` means that the node that has `id=“c4”` is the syntactic head of the node that has `id=“c3”`.

```

<cons cat="N" head="t0" id="c1">
  <tok id="t0"> wo (I) </tok>
</cons>
<cons cat="V" head="c3" id="c2" schema="head_mod">
  <cons cat="V" head="c4" id="c3">
    <cons cat="V" head="t1" id="c4">
      <tok id="t1"> qu (go to) </tok>
    </cons>
  </cons>
  <cons cat="N" head="c6" id="c5" schema="coord_left">
    <cons cat="N" head="t2" id="c6">
      <tok id="t2"> dongjing (Tokyo) </tok>
    </cons>
    <cons cat="COORD" head="c8" id="c7">
      <cons cat="CONJ" head="t3" id="c8">
        <tok id="t3"> he (and) </tok>
      </cons>
    </cons>
    <cons cat="N" head="t4" id="c9">
      <tok id="t4"> jingdu (Kyoto) </tok>
    </cons>
  </cons>
</cons>
<cons cat="PU" head="t5" id="c10">
  <tok id="t5"> . </tok>
</cons>
</cons>

```

Figure 1: An XML output for a Chinese sentence from Chinese Enju. For clarity, we only draw information related to the phrase structure and the heads.

### 2.3 Related Work

Reordering is a popular strategy for improving machine translation quality when source and target languages are structurally very different. Researchers have approached the reordering problem in multiple ways. The most basic idea is pre-ordering (Xia and McCord, 2004; Collins et al., 2005), that is, to do reordering during preprocessing time, where the source side of the training and development data and sentences from a source language that have to be translated are first reordered to ease the training and the translation, respectively. In (Xu et al., 2009), authors used a dependency parser to introduce manually created pre-ordering rules to reorder English sentences when translating into five different SOV (Subject-Object-Verb) languages. Other authors (Genzel, 2010; Wu et al., 2011) use automatically generated rules induced from parallel data. Tillmann (2004) used a lexical reordering model, and Galley et al. (2004) followed a syntactic-based model.

In this work, however, we are centered in the design of manual rules inspired by the Head Finalization (HF) reordering (Isozaki et al., 2010b). HF reordering is one of the simplest methods for pre-ordering that significantly improves word alignments and leads to a better translation quality. Al-

though the method is limited to translation where the target language is head-final, it requires neither training data nor fine-tuning. To our knowledge, HF is the best method to reorder languages when translating into head-final languages like Japanese.

The implementation of HF method for English-to-Japanese translation appears to work well. A reasonable explanation for this is the close match between the concept of “head” in this language pair. However, for Chinese-to-Japanese, there are differences in the definitions of numbers of important syntactic concepts, including the definition of the syntactic head. We concluded that the difficulties we encountered in using HF to Chinese-to-Japanese translation were the result of these differences in the definition of “head”. As we believe that such differences are also likely to be observed in other language pairs, the present work is generally important for head-initial to head-final translation as it shows a systematic linguistic analysis that consistently improves the effectivity of the HF method.

## 3 Syntax-based Reordering Rules

This section describes our method for syntax-based reordering for Chinese-to-Japanese translation. We start by introducing Head Finalization for Chinese (HFC), which is a simple adaptation of Isozaki et al. (2010b)’s method for English-to-Japanese translation. However, we found that this simple method has problems when applied to Chinese, due to peculiarities in Chinese syntax. In Section 3.2, we analyze several distinctive cases of the problem in detail. And following this analysis, Section 3.3 proposes a refinement of the original HFC, with a couple of exception rules for reordering.

### 3.1 Head Finalization for Chinese (HFC)

Since Chinese and English are both known to be head-initial languages<sup>1</sup>, the reordering rule introduced in (Isozaki et al., 2010b) ideally would reorder Chinese sentences to follow the word order

<sup>1</sup>As Gao (2008) summarized, whether Chinese is a head-initial or a head-final language is open for debate. Nevertheless, we take the view that most Chinese sentence structures are head-initial since the written form of Chinese mainly behaves as an head-initial language.

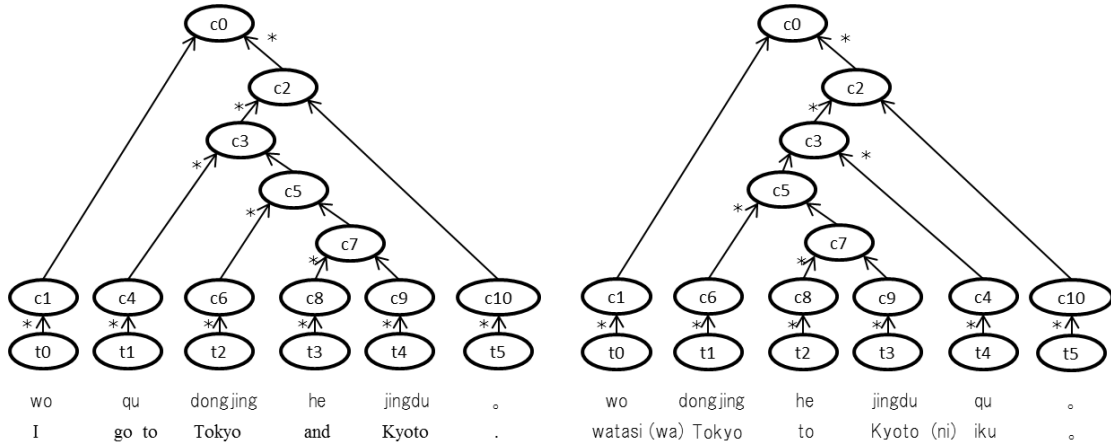


Figure 2: Simple example for Head-Final Chinese. The left figure shows the parsing tree of the original sentence and its English translation. The right figure shows the reordered sentence along with its Japanese translation. (“\*” indicate the syntactic head).

of their Japanese counterparts.

Figure 2 shows an example of a head finalized Chinese sentence based on the output from Chinese Enju shown in Figure 1. Notice that the coordination exception rule described in (Isozaki et al., 2010b) also applies to Chinese reordering. This exception rule says that child nodes are not swapped if the node is a coordination<sup>2</sup>. Another exception rule is for punctuation symbols, which are also preserved in their original order. In this case, as can be seen in the example in Figure 2, the nodes of **c3**, **c6**, and **c8** had not been swapped with their dependency. In this account, only the verb “qu” had been moved to the end of the sentence, following the same word order as its Japanese translation.

### 3.2 Discrepancies in Head Definition

Head Finalization relies on the idea that head-dependent relations are largely consistent among different languages while word orders are different. However, in Chinese, there has been much debate on the definition of head<sup>3</sup>, possibly because Chinese has fewer surface syntactic features than other languages like English and Japanese. This causes some discrepancies between the definitions

<sup>2</sup>Coordination is easily detected in the output of Enju; it is marked by the attributes `xcat="COORD"` or `schema="coord-left/right"` as shown in Figure 1.

<sup>3</sup>In this paper, we only consider the syntactic head.

of the head in Chinese and Japanese, which leads to undesirable reordering of Chinese sentences. Specifically, in preliminary experiments we observed unexpected reorderings that are caused by the differences in the head definitions, which we describe below.

#### 3.2.1 Aspect Particle

Although Chinese has no syntactic tense marker, three aspect particles following verbs can be used to identify the tense semantically. They are “le0” (did), “zhe0” (doing), and “guo4” (done), and their counterparts in Japanese are “ta”, “teiru”, and “ta”, respectively. Both the first word and third word can represent the past tense, but the third one is more often used in the past perfect.

The Chinese parser<sup>4</sup> treated aspect particles as dependents of verbs, whereas their Japanese counterparts are identified as the head. For example in Table 1<sup>5</sup>, “qu” (go) and “guo” (done) aligned with “i” and “tta”, respectively. However, since “guo” is treated as a dependent of “qu”, by directly implementing the Head Final Chinese (HFC), the sentence will be reordered like

<sup>4</sup>The discussions in this section presuppose the syntactic analysis done by Chinese Enju, but most of the analysis is consistent with the common explanation for Chinese syntax.

<sup>5</sup>English translation (En); Chinese original sentence (Ch); reordered Chinese by Head-Final Chinese (HFC); reordered Chinese by Refined Head-Final Chinese (R-HFC) and Japanese translation (Ja).

HFC in Table 1, which does not follow the word order of the Japanese (Ja) translation. In contrast, the reordered sentence from refined-HFC (R-HFC) can be translated monotonically.

|       |  |
|-------|--|
| En    | I have been to Tokyo.                  |
| Ch    | wo <b>qu guo</b> dongjing.             |
| HFC   | wo dongjing <b>guo qu</b> .            |
| R-HFC | wo dongjing <b>qu guo</b> .            |
| Ja    | watashi (wa) Tokyo (ni) <b>i tta</b> . |

Table 1: An example for Aspect Particle. Best word alignment Ja-Ch (En): “watashi” – “wo”(I); “Tokyo” – “dongjing” (Tokyo); “i” – “qu” (been); “tta” – “guo” (have).

### 3.2.2 Adverbial Modifier ‘bu4’

Both in Chinese and Japanese, verb phrase modifiers typically occur in pre-verbal positions, especially when the modifiers are adverbs. Since adverbial modifiers are dependents in both Chinese and Japanese, head finalization works perfectly for them. However, there is an exceptional adverb, “bu4”, which means negation and is usually translated into “nai”, which is always at the end of the sentence in Japanese and thus is the head. For example in Table 2, the word “kan” (watch) will be identified as the head and the word “bu” is its dependent; on the contrary, in the Japanese translation (Ja), the word “nai”, which is aligned with “bu”, will be identified as the head. Therefore, the Head Final Chinese is not in the same order, but the reordered sentence by R-HFC obtained the same order with the Japanese translation.

|       |  |
|-------|--|
| En    | I do not watch TV.                       |
| Ch    | wo <b>bu kan</b> dianshi.                |
| HFC   | wo dianshi <b>bu kan</b> .               |
| R-HFC | wo dianshi <b>kan bu</b> .               |
| Ja    | watashi (wa) terebi (wo) <b>mi nai</b> . |

Table 2: An example for Adverbial Modifier bu4. Best word alignment Ja-Ch (En): “watashi” – “wo” (I); “terebi” – “dianshi” (TV); “mi” – “kan” (watch); “nai” – “bu” (do not).

### 3.2.3 Sentence-final Particle

Sentence-final particles often appear at the end of a sentence to express a speaker’s attitude: e.g. “ba0, a0” in Chinese, and “naa, nee” in Japanese. Although they appear in the same position in both Chinese and Japanese, in accordance with the differences of head definition, they are identified as the dependent in Chinese while they are the head in Japanese. For example in Table 3, since “a0” was identified as the dependent, it had been reordered to the beginning of the sentence while its Japanese translation “nee” is at the end of the sentence as the head. Likewise, by refining the HFC, we can improve the word alignment.

|       |                             |
|-------|-----------------------------|
| En    | It is good weather.         |
| Ch    | tianqi zhenhao <b>a</b> .   |
| HFC   | <b>a</b> tianqi zhenhao.    |
| R-HFC | tianqi zhenhao <b>a</b> .   |
| Ja    | ii tennki desu <b>nee</b> . |

Table 3: An example for Sentence-final Particle. Best word alignment Ja-Ch (En): “tennki” – “tianqi” (weather); “ii” – “zhenhao” (good); “nee” – “a” (None).

### 3.2.4 Et cetera

In Chinese, there are two expressions for representing the meaning of “and other things” with one Chinese character: “deng3” and “deng3 deng3”, which are both identified as dependent of a noun. In contrast, in Japanese, “nado” is always the head because it appears as the right-most word in a noun phrase. Table 4 shows an example.

|       |   |
|-------|---|
| En    | Fruits include apples, etc.                       |
| Ch    | shuiguo baokuo pingguo <b>deng</b> .              |
| HFC   | shuiguo <b>deng</b> pingguo baokuo.               |
| R-HFC | shuiguo pingguo <b>deng</b> baokuo.               |
| Ja    | kudamono (wa) ringo <b>nado</b> (wo) fukunde iru. |

Table 4: An example for Et cetera. Best word alignment Ja-Ch (En): “kudamono” – “shuiguo” (Fruits); “ringo” – “pingguo” (apples); “nado” – “deng” (etc.); “fukunde iru” – “baokuo” (include).

|     |   |
|-----|---|
| AS  | Aspect particle                               |
| SP  | Sentence-final particle                       |
| ETC | <i>et cetera</i> (i.e. deng3 and deng3 deng3) |
| IJ  | Interjection                                  |
| PU  | Punctuation                                   |
| CC  | Coordinating conjunction                      |

Table 5: The list of POSs for exception reordering rules

### 3.3 Refinement of HFC

In the preceding sections, we have discussed syntactic constructions that cause wrong application of Head Finalization to Chinese sentences. Following the observations, we propose a method to improve the original Head Finalization reordering rule to obtain better alignment with Japanese.

The idea is simple: we define a list of POSs, and when we find one of them as a dependent child of the node, we do not apply reordering. Table 5 shows the list of POSs we define in the current implementation<sup>6</sup>. While interjections are not discussed in detail, we should obviously not reorder to interjections because they are position-independent. The rules for PU and CC are basically equivalent to the exception rules proposed by (Isozaki et al., 2010b).

## 4 Experiments

The corpus we used as training data comes from the China Workshop on Machine Translation (CWMT) (Zhao et al., 2011). This is a Japanese-Chinese parallel corpus in the news domain, containing 281,322 sentence pairs. We also collected another Japanese-Chinese parallel corpus from news containing 529,769 sentences and merged it with the CWMT corpus to create an extended version of the CWMT corpus. We will refer to this corpus as “CWMT ext.” We split an inverted multi-reference set into a development and a test set containing 1,000 sentences each. In these two sets, the Chinese input was different, but the Japanese reference was identical. We think that this split does not pose any severe problem to the comparison fairness of the experiment, since no new phrases are added during tuning and the experimental conditions remain equal for all tested

<sup>6</sup>The POSs are from Penn Chinese Treebank.

|           |                      | Ch    | Ja    |
|-----------|----------------------|-------|-------|
| CWMT      | Sentences            | 282K  |       |
|           | Run. words           | 2.5M  | 3.2M  |
|           | Avg. sent. leng.     | 8.8   | 11.5  |
|           | Vocabulary           | 102K  | 42K   |
| CWMT ext. | Sentences            | 811K  |       |
|           | Run. words           | 14.7M | 17M   |
|           | Avg. sent. leng.     | 18.1  | 20.9  |
|           | Vocabulary           | 249K  | 95K   |
| Dev.      | Sentences            | 1000  |       |
|           | Run. words           | 29.9K | 35.7K |
|           | Avg. sent. leng.     | 29.9  | 35.7  |
|           | OoV w.r.t. CWMT      | 485   | 106   |
|           | OoV w.r.t. CWMT ext. | 244   | 53    |
| Test      | Sentences            | 1000  |       |
|           | Run. words           | 25.8K | 35.7K |
|           | Avg. sent. leng.     | 25.8  | 35.7  |
|           | OoV w.r.t. CWMT      | 456   | 106   |
|           | OoV w.r.t. CWMT ext. | 228   | 53    |

Table 6: Characteristics of CWMT and extended CWMT Chinese-Japanese corpus. Dev. stands for Development, OoV for “Out of Vocabulary” words, K for thousands of elements, and M for millions of elements. Data statistics were collected after tokenizing.

methods. Detailed Corpus statistics can be found in Table 6.

To parse Chinese sentences, we used Chinese Enju (Yu et al., 2010), an HPSG-based parser trained with the Chinese HPSG treebank converted from Penn Chinese Treebank. Chinese Enju requires segmented and POS-tagged sentences to do parsing. We used the Stanford Chinese segmenter (Chang et al., 2008) and Stanford POS-tagger (Toutanova et al., 2003) to obtain the segmentation and POS-tagging of the Chinese side of the training, development, and test sets.

The baseline system was trained following the instructions of recent SMT evaluation campaigns (Callison-Burch et al., 2010) by using the MT toolkit Moses (Koehn et al., 2007) in its default configuration. Phrase pairs were extracted from symmetrized word alignments and distortions generated by GIZA++ (Och and Ney, 2003) using the combination of heuristics “grow-diag-final-and” and “msd-bidirectional-fe”. The language model was a 5-gram language model estimated on the target side of the parallel corpora by using the modified Kneser-Ney smoothing (Chen and Goodman, 1999) implemented in

the SRILM (Stolcke, 2002) toolkit. The weights of the log-linear combination of feature functions were estimated by using MERT (Och, 2003) on the development set described in Table 6.

The effectiveness of the reorderings proposed in Section 3.3 was assessed by using two precision metrics and two error metrics on translation quality. The first evaluation metric is BLEU (Papineni et al., 2002), a very common accuracy metric in SMT that measures  $N$ -gram precision, with a penalty for too short sentences. The second evaluation metric was RIBES (Isozaki et al., 2010a), a recent precision metric used to evaluate translation quality between structurally different languages. It uses notions on rank correlation coefficients and precision measures. The third evaluation metric is TER (Snover et al., 2006), another error metric that computes the minimum number of edits required to convert translated sentences into its corresponding references. Possible edits include insertion, deletion, substitution of single words, and shifts of word sequences. The fourth evaluation metric is WER, an error metric inspired in the Levenshtein distance at word level. BLEU, WER, and TER were used to provide a sense of comparison but they do not significantly penalize long-range word order errors. For this reason, RIBES was used to account for this aspect of translation quality.

The baseline system was trained and tuned using the same configuration setup described in this section, but no reordering rule was implemented at the preprocessing stage.

Three systems have been run to translate the test set for comparison when the systems were trained using the two training data sets. They are the baseline system, the system consisting in the naïve implementation of HF reordering, and the system with refined HFC reordering rules. Assessment of translation quality can be found in Table 7.

As can be observed in Table 7, the translation quality, as measured by precision and error metrics, was consistently and significantly increased when the HFC reordering rule was used and was significantly improved further when the refinement proposed in this work was used. Specifically, the BLEU score increased from 19.94 to 20.79 when the CWMT corpus was used, and from 23.17 to 24.14 when the extended CWMT corpus was used.

| AS   | SP   | ETC  | IJ    | PU    | COOD  |
|------|------|------|-------|-------|-------|
| 3.8% | 0.8% | 1.3% | 0.0%* | 21.0% | 38.3% |

Table 8: Weighted recall of each exception rule during reordering on CWMT ext. training data, dev data, and test data. (\* actual value 0.0016%.)

Table 8 shows the recall of each exception rule listed in Section 3, and was computed by counting the times an exception rule was triggered divided by the number of times the head finalization rule applied. Data was collected for CWMT ext. training, dev and test sets. Although the exception rules related to aspect particles, *Et cetera*, sentence-final particles and interjections have a comparatively lower frequency of application than punctuation or coordination exception rules, the improvements they led to are significant.

## 5 Error Analysis

In Section 3 we have analyzed syntactic differences between Chinese and Japanese that led to the design of an effective refinement. A manual error analysis of the results of our refined reordering rules showed that some more reordering issues remain and, although they are not side effects of our proposed rule, they are worth mentioning in this separate section.

### 5.1 Serial Verb Construction

Serial verb construction is a phenomenon occurring in Chinese, where several verbs are put together as one unit without any conjunction between them. The relationship between these verbs can be progressive or parallel. Apparently, Japanese has a largely corresponding construction, which indicates that no reordering should be applied. An example to illustrate this fact in Chinese is “**weishi** (maintain) **shenhua** (deepen) zhongriguanxi (Japan-China relations) de (of) gaishan (improvement) jidiao (basic tone).”<sup>7</sup> The two verbs “weishi” (in Japanese, *iji*) and “shenhua” (in Japanese, *shinka*) are used together, and they follow the same order as in Japanese: “*nicchukankei* (Japan-China re-

<sup>7</sup>English translation: Maintain and deepen the improved basic tone of Japan-China relations.

|             | CWMT         |              |              |              | CWMT ext.    |              |              |              |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|             | BLEU         | RIBES        | TER          | WER          | BLEU         | RIBES        | TER          | WER          |
| baseline    | 16.74        | 71.24        | 70.86        | 77.45        | 20.70        | 74.21        | 66.10        | 72.36        |
| HFC         | 19.94        | 73.49        | 65.19        | 71.39        | 23.17        | 75.35        | 61.38        | 67.74        |
| refined HFC | <b>20.79</b> | <b>75.09</b> | <b>64.91</b> | <b>70.39</b> | <b>24.14</b> | <b>77.17</b> | <b>59.67</b> | <b>65.31</b> |

Table 7: Evaluation of translation quality of a test set when CWMT and CWMT extended corpus were used for training. Results are given in terms of BLEU, RIBES, TER, and WER for baseline, head finalization, and proposed refinement of head finalization reordering rules.

lations) no (of) kaizan (improvement) kityo (basic tone) wo **iji** (maintain) **shinka** (deepen) suru (do).”

## 5.2 Complementizer

A “complementizer” is a particle used to introduce a complement. In English, a very common complementizer is the word “that” when making a clausal complement, while in Chinese it can denote other types of word, such as verbs, adjectives or quantifiers. The complementizer is identified as the dependent of the verb that it modifies. For instance, a Chinese sentence: “wo (I) **mang wan** le (have finished the work).” This can be translated into Japanese: “watashi (I) wa shigoto (work) wo **owa tta** (have finished).” In Chinese, the verb “mang” is the head while “wan” is the complementizer, and its Japanese counterpart “owa tta” has the same word order.

However, during the reordering, “mang” will be placed at the end of the sentence and “wan” in the beginning, leading to an inconsistency with respect to the Japanese translation where the complementizer “tta” is the head.

## 5.3 Verbal Nominalization and Noun Verbalization

As discussed by Guo (2009), compared to English and Japanese, Chinese has little inflectional morphology, that is, no inflection to denote tense, case, etc. Thus, words are extremely flexible, making verb nominalization and noun verbalization appear frequently and commonly without any conjugation or declension. As a result, it is difficult to do disambiguation during POS tagging and parsing. For example, the Chinese word “kaifa” may have two syntactic functions: verb (develop) and noun (development). Thus, it is difficult to reliably tag

without considering the context. In contrast, in Japanese, “suru” can be used to identify verbs. For example, “kaihatu suru” (develop) is a verb and “kaihatu” (development) is a noun. This ambiguity is prone to not only POS tagging error but also parsing error, and thus affects the identification of heads, which may lead to incorrect reordering.

## 5.4 Adverbial Modifier

Unlike the adverb “bu4” we discussed in Section 3.2, the ordinary adverbial modifier comes directly before the verb it modifies both in Chinese and Japanese, but not in English. Nevertheless, in accordance with the principle of identifying the head for Chinese, the adverb will be treated as the dependent and it will not be reordered following the verb it modified. As a result, the alignment between adverbs and verbs is non-monotonic. This can be observed in the Chinese sentence “guojia (country) **yanli** (severely) chufa (penalize) jiage (price) weifa (violation) xingwei (behavior)”<sup>8</sup>, and its Japanese translation: “kuni (country) wa kakaku (price) no ihou (violation) kouji (behavior) wo **kibisiku** (severely) syobatu (penalize).” Both in Chinese and Japanese, the adverbial modifier “yanli” and “kibisiku” are directly in front of the verb “chufa” and “syobatu”, respectively. However, the verb in Chinese is identified as the head and will be reordered to the end of the sentence without the adverb.

<sup>8</sup>English translation: The country severely penalizes violations of price restrictions.



## 5.5 POS tagging and Parsing Errors

There were word reordering issues not caused solely by differences in syntactic structures. Here we summarize two that are difficult to remedy during reordering and that are hard to avoid since reordering rules are highly dependent on the tagger and parser.

- POS tagging errors

In Chinese, for example, the word “Iran” was tagged as “VV” or “JJ” instead of “NR”. This led to identifying “Iran” as a head in accordance with the head definition in Chinese, and it was reordered undesirably.

- Parsing errors

For example, in the Chinese verb phrase “touzhi (invest) 20 yi (200 million) meiyuan (dollars)”, “20” and “yi” were identified as dependent of “touzhi” and “meiyuan”, respectively, which led to an unsuitable reordering for posterior word alignment.

## 6 Conclusion and Future Work

In the present work, we have proposed novel Chinese-to-Japanese reordering rules inspired in (Isozaki et al., 2010b) based on linguistic analysis on Chinese HPSG and differences among Chinese and Japanese. Although a simple implementation of HF to reorder Chinese sentences performs well, translation quality was substantially improved further by including linguistic knowledge into the refinement of the reordering rules.

In Section 5, we found more patterns on reordering issues when reordering Chinese sentences to resemble Japanese word order. The extraction of those patterns and their effective implementation may lead to further improvements in translation quality, so we are planning to explore this possibility.

In this work, syntactic information from a deep parser has been used to reorder words better. We believe that using semantic information can further increase the expressive power of reordering rules. With that objective, Chinese Enju can be used since it provides the semantic head of nodes

and can interpret sentences by using their semantic dependency.

## Acknowledgments

This work was mainly developed during an internship at NTT Communication Science Laboratories. We would like to thank Prof. Yusuke Miyao for his invaluable support on this work.

## References

- P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, and R.L. Mercer. 1993. The mathematics of machine translation. In *Computational Linguistics*, volume 19, pages 263–311, June.
- Chris Callison-Burch, Philipp Koehn, Christof Monz, Kay Peterson, and Omar Zaidan, editors. 2010. *Proceedings of the joint 5th workshop on Statistical Machine Translation and MetricsMATR*. Association for Computational Linguistics, July.
- Pi-Chuan Chang, Michel Galley, and Christopher D. Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the 3rd Workshop on SMT*, pages 224–232, Columbus, Ohio. Association for Computational Linguistics.
- Stanley F. Chen and Joshua Goodman. 1999. An empirical study of smoothing techniques for language modeling. *Computer Speech and Language*, 4(13):359–393.
- Michael Collins, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL ’05*, pages 531–540, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Vivian James Cook and Mark Newson. 1988. *Chomsky’s Universal Grammar: An introduction*. Oxford: Basil Blackwell.
- Naoki Fukui. 1992. *Theory of Projection in Syntax*. CSLI Publisher and Kuroshio Publisher.
- Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. Whats in a translation rule? In *Proceedings of HLT-NAACL*.
- Qian Gao. 2008. Word order in mandarin: Reading and speaking. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, volume 2, pages 611–626.
- Dmitriy Genzel. 2010. Automatically learning source-side reordering rules for large scale machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics, COLING ’10*,

- pages 376–384, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Yuqing Guo. 2009. *Treebank-based acquisition of Chinese LFG resources for parsing and generation*. Ph.D. thesis, Dublin City University.
- Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010a. Automatic evaluation of translation quality for distant language pairs. In *Proceedings of Empirical Methods on Natural Language Processing (EMNLP)*.
- Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010b. Head finalization: A simple re-ordering rule for sov languages. In *Proceedings of WMTMetricsMATR*, pages 244–251.
- P. Koehn, F. J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proceedings HLT/NAACL’03*, pages 48–54.
- Philipp Koehn et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the ACL Demo and Poster Sessions, 2007*, pages 177–180, June 25–27.
- Yusuke Miyao and Jun’ichi Tsujii. 2008. Feature forest models for probabilistic hpsg parsing. *Computational Linguistics*, 34:35–80, March.
- F. J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Franz Josef Och and Hermann Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics*.
- Franz J. Och. 2003. Minimum error rate training for statistical machine translation. In *Proceedings of the 41st annual conference of the Association for Computational Linguistics, 2003*, pages 160–167, July 7–12.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual conference of the Association for Computational Linguistics, 2002*, pages 311–318, July 6–12.
- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of Association for Machine Translation in the Americas*, pages 223–231.
- Andreas Stolcke. 2002. SRILM – an extensible language modeling toolkit. In *Proceedings of the 7th international conference on Spoken Language Processing, 2002*, pages 901–904, September 16–20.
- Christoph Tillmann. 2004. A unigram orientation model for statistical machine translation. In *Proceedings of HLT-NAACL 2004: Short Papers, HLT-NAACL-Short ’04*, pages 101–104, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings OF HLT-NAACL*, pages 252–259.
- Chao Wang, Michael Collins, and Philipp Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 737–745, Prague, Czech Republic, June. Association for Computational Linguistics.
- Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, and Masaaki Nagata. 2011. Extracting pre-ordering rules from predicate-argument structures. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 29–37, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.
- Fei Xia and Michael McCord. 2004. Improving a statistical mt system with automatically learned rewrite patterns. In *Proceedings of the 20th international conference on Computational Linguistics, COLING ’04*, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Peng Xu, Jaeho Kang, Michael Ringgaard, and Franz Och. 2009. Using a dependency parser to improve smt for subject-object-verb languages. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL ’09*, pages 245–253, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kun Yu, Yusuke Miyao, Xiangli Wang, Takuya Matsuzaki, and Jun ichi Tsujii. 2010. Semi-automatically developing chinese hpsg grammar from the penn chinese treebank for deep parsing. In *COLING (Posters)’10*, pages 1417–1425.
- Kun Yu, Yusuke Miyao, Takuya Matsuzaki, Xiangli Wang, and Junichi Tsujii. 2011. Analysis of the difficulties in chinese deep parsing. In *Proceedings of the 12th International Conference on Parsing Technologies*, pages 48–57.
- R. Zens, F.J. Och, and H. Ney. 2002. Phrase-based statistical machine translation. In *Proceedings of KI’02*, pages 18–32.
- Hong-Mei Zhao, Ya-Juan Lv, Guo-Sheng Ben, Yun Huang, and Qun Liu. 2011. Evaluation report for the 7th china workshop on machine translation (cwmt2011). *The 7th China Workshop on Machine Translation (CWMT2011)*.