Example-based Machine Translation using Structural Translation Examples

Eiji Aramaki and Sadao Kurohashi

Graduate School of Information Science and Tech.
University of Tokyo
{aramaki, kuro}@kc.t.u-tokyo.ac.jp

Abstract

This paper proposes an example-based machine translation system which handles structural translation examples. The structural translation examples have the potential advantage of high-usability. However, technologies which build such translation examples are still being developed. In such a situation, the comparison of the proposed system and the other approach systems is meaningful. This paper presents the system algorithm and its performance on the IWSLT04 Japanese-English unrestricted task.

1. Introduction

We are developing an example-based machine translation (EBMT)[1] system using structural translation examples, which is potentially suitable to deal with the infinite productivity of languages. Structural translation examples have the advantage of high-usability, and the system under this approach needs only a reasonable scale of corpus.

However, building the structural translation examples requires many technologies, e.g., parsing and tree-alignment and so on, which are still being developed. So a naive method without such technologies may be efficient in a limited domain.

In such a situation, we believe that the comparison of the proposed system and the other approach systems is meaningful.

The proposed system challenged the "Japanese-English unrestricted" task, but it utilized no extra bilingual corpus of the domain; it used only a training corpus given in the IWSLT04, Japanese and English parsers, a Japanese thesaurus and translation dictionaries. Figure 1 shows the system outline. It consists of two modules, (1) an alignment module and (2) a translation module.

The alignment module estimates correspondences in the corpus using translation dictionaries. Then, the alignment results are stored in a translation memory which is a database of translation examples. The translation module selects plausible translation examples for each parts of an input sentence. Finally, the selected examples are combined to generate an output sentence.

This paper is organized as follows. The next section presents our system algorithm. Section 3 reports experimental results. Then, Section 4 presents our conclusions.
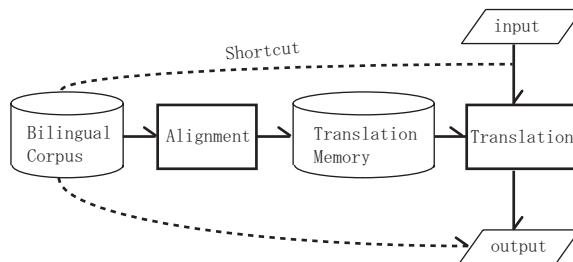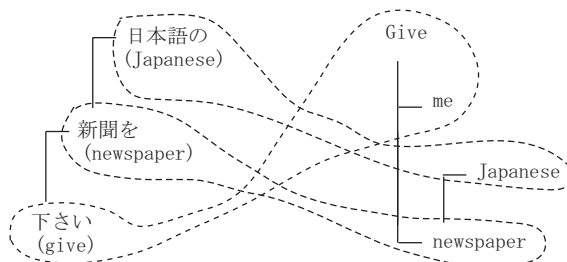

Figure 1: System Outline.

Figure 2: Aligned Sentence Pair.
* In this paper, a sentence structure is illustrated by locating its root node at the left.

2. Algorithm

2.1. Alignment Module

An EBMT system needs a large set of translation examples. In order to build them, we use the dictionary-based alignment method presented in [2].

First, sentence pairs are parsed by the Japanese parser KNP [3] and the English nl-parser [4]. The English parser outputs a phrase structure. Then, it is converted into a dependency structures by rules which decide on a head word in a phrase. A Japanese phrase consists of sequential context words and their following function words. An English phrase is a base-NP or a base-VP.

Then, correspondences are estimated by using translation dictionaries. We used four dictionaries: EDR, EDICT,
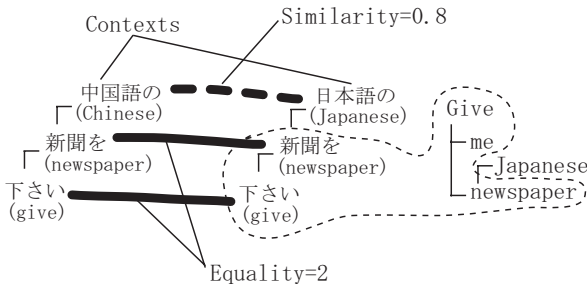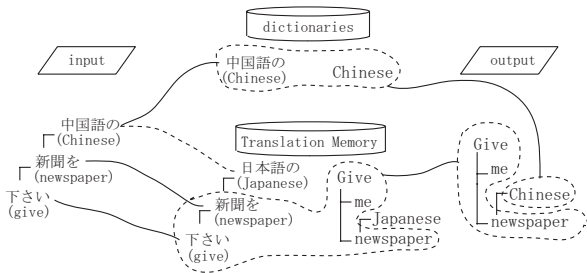
Figure 4: Equality and Similarity.

Figure 5: Example of Translation Flow.

ENAMDICT, and EIJIRO. These dictionaries have about two million entries in total. If there are out of dictionary phrases, they are merged into their parent correspondence. A sample alignment result is shown in Figure 2.

After alignment, the system generates all combinations of correspondences which are connected to each other. We call such a combination of correspondences a translation example. As a result, the 6 translation examples shown in Figure 3 are generated from the aligned sentence pair shown in Figure 2.

Finally, these translation examples are stored in the translation memory. In this operation, surrounding phrases (its parent and its children phrases) are also preserved as the contexts (mentioned in the next Section).

2.2. Translation Module

First, an input sentence is analyzed by the parser[3]. Then, for each phrase of the input sentence, the system selects plausible translation examples from the translation memory by using the following three measures.

1. **Equality**: If large parts of a translation example are equal to the input, we regard it as a reliable example. The equality is the number of translation example phrases which are equal to the input. The system conducts the equal check in content words and some function words which express a sentence mood. The differences of the other function words are disregarded. In Figure 4, the translation example has equality 2.

2. **Similarity**: Context is an important clue for word selection. We regard the context as the surrounding phrases of the equal part. The similarity score between the surrounding phrases and their corresponding input phrases is calculated with a Japanese thesaurus (max=1.0).

3. **Confidence**: We also take into account the alignment confidence. We define the alignment confidence as the ratio of content words which can be found in dictionaries (max =1.0).

The detailed definitions of those measures are presented in [5]. These measures are weighted by a parameter λ as follows[1], and the system selects the translation examples which have the highest score for each parts of the input:

$$(Equality + Similarity) \times (\lambda + Confidence).$$

If there is no translation example, the system uses the translation dictionaries and acquires target expressions. If the translation dictionaries have no entry, the system stops the following procedures and goes to a shortcut pass (mentioned in Section 2.3).

After the selection of translation examples, the target expressions in the examples are combined into a target dependency tree and its word order is decided. In this operation, the dependency relations and the word order are decided by the following principles.

1. The dependency relations and the word order in a translation example are preserved.

2. The dependency relations between the translation examples are equal to the relations of their corresponding input phrases.

3. The word order between translation examples is decided by the rules governing both the dependency relation and the word order.

Figure 5 shows an example for a Japanese input which means "give me a Chinese newspaper" with selected examples and its target dependency tree.

2.3. Shortcut

Yet there are no perfect alignment and parsing technologies, so the proposed system has a risk of pre-processing errors. In view of this, we also prepare another translation method without such pre-processing   We call this method a **shortcut**. The shortcut method searches the most similar translation examples by using a character-based DP matching method, and outputs its target parts as it is.

The shortcut is used in the following three situations.

---

[1] λ was determined by a preliminary experiment not to deteriorate the accuracy of the system. In preliminary experiments, we set λ as 1.
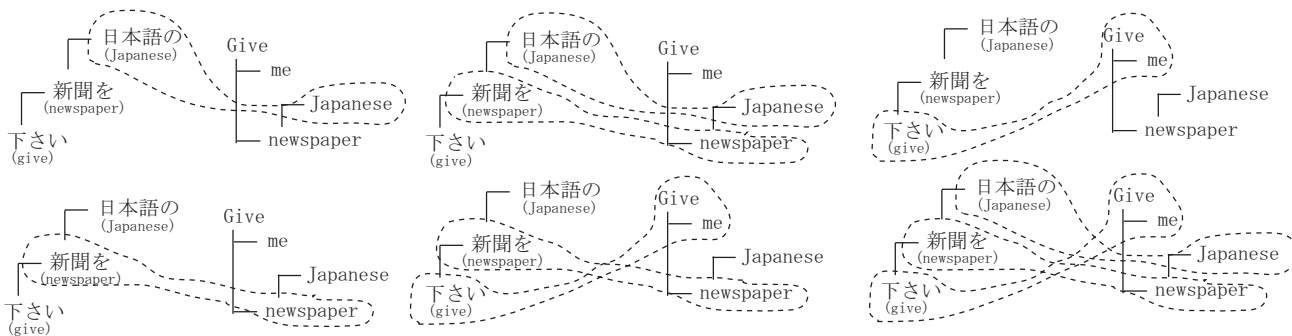
Figure 3: Translation Examples.

Almost Equal: An input has more than 90% similarity which is calculated by a character-based DP matching method.

No expression: The system can not acquire any target expressions from either the translation memory or the dictionaries.

Un-grammatical: The system generates un-grammatical expressions, e.g., the same word sequence.

3. Experiments

3.1. Experimental Condition

We built translation examples from a training-set for the IWSLT04. The training-set consists of 20,000 Japanese and English sentence pairs. The evaluation was conducted using a dev-set and a test-set for the IWSLT04, which consist of about 500 Japanese sentences with 16 references.

3.2. Result

The following five automatic evaluation results are shown in Table 1 and some translation samples are shown in Table 2.

BLEU: The geometric mean of n-gram precision by the system output with respect to the reference translations[6].

NIST: A variant of BLEU using the arithmetic mean of weighted n-gram precision values.

WER (word error rate): The edit distance between the system output and the closest reference translation.

PER (position-independent WER): A variant of mWER which disregards word ordering.

GTM (general text matcher): Harmonic mean of precision and recall measures for maximum matchings of aligned words in a bitext grid.

Table 1: Result.

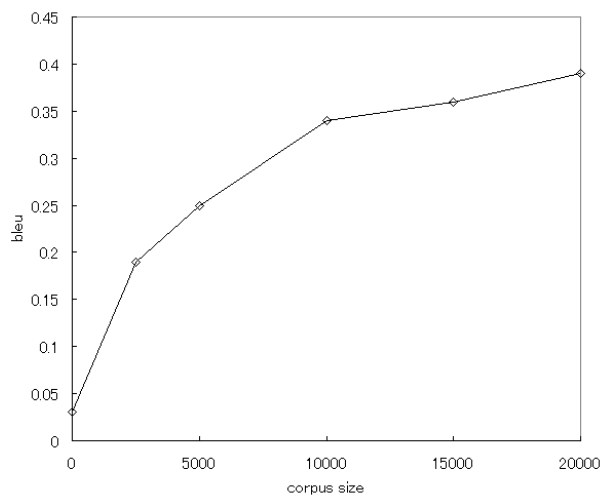|	bleu	nist	wer	per	gtm
dev-set	0.38	7.86	0.52	0.45	0.66
test-set	0.39	7.89	0.49	0.42	0.67

Figure 6: Corpus Size and Performance (BLEU).
* The system without a corpus can generate translations using only the translation dictionaries.

93

Table 2: Result Samples

	ズキズキ痛みます。
input	zuki zuki itamimasu.
output	it is a throbbing pain
ref	I am suffering from a throbbing pain .
	市役所行きのバス停はどこですか。
input	siyakusyo iki no basu tei ha doko desuka.
output	where is the bus stop for the city hall
ref	Where is the bus stop for buses going to city hall ?
	この綿のセーターを試着してもよろしいですか。
input	kono men no seta- wo sityaku sitemo yorosii desuka
output	i would like to try this sweater for an cotton
ref	Is it alright if I try on this cotton sweater ?
	搭乗ゲートはどこですか。
input	touzyou ge-to ha doko desuka
output	where is the gate
ref	Where is the passenger boarding gate ?
	日本へ送っていただけますか。
input	nihon ni okutte itadake masuka
output	could you send it to this japan
ref	Could you send this to Japan ?

The dev-set and test-set scores are similar because the system has no tuning metrics for the dev-set.

Then, we investigated the relation between the corpus size (the number of sentence pairs) and its performance (bleu) . The result is shown in Figure 6. The score is not saturated at the point of x=20,000. Therefore, the system will achieve a higher performance if we obtain more corpora.

3.3. Error Analysis

Most of the errors are classified into the following three problems.

1. **Function Words**: Because the system selects translation examples using mainly content words, it sometimes generates un-natural function words, especially in determiners and prepositions. For example, the system generates the output "i 'd like to contact **my** japanese embassy" using a translation example "I 'd like to contact **my** bank".

   In the future, the system should deal with translation examples more carefully.

2. **Word Order**: The word order between translation examples is decided by the heuristic rules. The lack of rules leads to the wrong word order. For example, "is there anything a like local cuisine?"

A target language model may be helpful for this problem.

3. **Lack of a Subject**: The proposed system sometimes generates an output without a subject, for example, "has a bad headache". It is because the input sentence often includes a zero-pronoun.

   In the future, we are planning to incorporate the zero-pronoun resolution technology.

4. Conclusion

In this paper, we described an EBMT system which handles structural translation examples. The experimental result shows the basic feasibility of this approach. In the future, as the amount of corpora increases, the system will achieve a higher performance.

5. References

[1] M. Nagao, "A framework of a mechanical translation between Japanese and English by analogy principle," in *In Artificial and Human Intelligence*, 1984, pp. 173–180.

[2] E. Aramaki, S. Kurohashi, S. Sato, and H. Watanabe, "Finding translation correspondences from parallel parsed corpus for example-based translation," in *Proceedings of MT Summit VIII*, 2001, pp. 27–32.

[3] S. Kurohashi and M. Nagao, "A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures," *Computational Linguistics*, vol. 20, no. 4, 1994.

[4] E. Charniak, "A maximum-entropy-inspired parser," in *In Proceedings of NAACL 2000*, 2000, pp. 132–139.

[5] E. Aramaki, S. Kurohashi, H. Kashioka, and H. Tanaka, "Word selection for ebmt based on monolingual similarity and translation confidence," in *Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, 2003, pp. 57–64.

[6] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of ACL 2002*, 2002, pp. 311–318.