Towards Automatic Translation from Japanese into Japanese Sign Language*

TOKUDA Masaaki , OKUMURA Manabu
School of Information Science,
Japan Advanced Institute of Science and Technology
(Tatsunokuchi, Ishikawa 923-12 Japan)
tokuda@jaist.ac.jp, oku@jaist.ac.jp

Abstract

In this paper, we present a prototype MT system named SYUWAN which can translate Japanese into Japanese sign language. One of the most important problems in this translation is that there are very few entries in a sign language dictionary, compared with Japanese one. To solve this problem, when the original input word does not exist in a sign language dictionary, SYUWAN applies some techniques to find a similar word from a Japanese dictionary and substitutes this word for the original word. As the result, SYUWAN can translate up to 95% of words which are morphologically analyzed.

Introduction

The deaf communicate with each other by using sign language which is composed of hands, arms and face expression. Japanese deafs use Japanese sign language (JSL) which is different from both phonetic language (Japanese language) and other sign language (American sign language (ASL)). According to recent linguistic researches, JSL has peculiar syntax. However, there are few researches on JSL, and no practical machine translation (MT) system between sign language and phonetics language exists.

In the past, Adachi (Adachi92) analyzed a pair of daily news sentences which include Japanese and sign language, and studied the method to translate from Japanese into sign language, but currently the sign language expressions of daily news is improved. Therefore, it is necessary to reconsider a translation method. Nishikawa and Terauchi (Nishikawa95) studied expression and translation of sign language using computer graphics. They designed and implemented a translation system, but they assumed that input words of their system have already been analyzed and attached completely with necessary information (part of speech, word sense and so on).

Besides sign language has a problem with its special characteristic, it is very difficult to be processed on a computer because it is a visual language. Recently, there have been many researches on recognition or generation of sign language (Kurokawa92) (Lee94) (Ohira95), but practical input/output devices or computerized systems of sign language are not available even now. The processing load of sign language is very heavy. Moreover, they studied mainly on computer graphics, and used a complex description for sign language which can be easily processed by computers but hard to be understood by human. Therefore, a description for sign language which is simple for human to understand and suitable for being processed by computers, is needed.

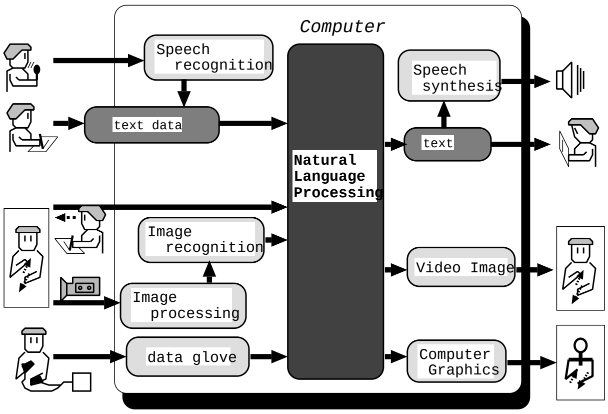


Figure 1: An outline of MT system for sign language

In this paper, we introduce "Sign Language Description Method (SLDM)" which is a description method of sign language using Japanese words as labels for expressing a word sequence. We propose a translation method referring to Adachi's work, and implement a translation system named "SYUWAN" which processes raw input data of Japanese sentences and outputs the corresponding symbols based on SLDM.

We take a transfer approach as a translation method for SYUWAN which needs a source and target language dictionaries. In this work, we use a Japanese dictionary as the source language dictionary and a sign

---

*This work is partly supported by the Inamura Foundation.

language dictionary as the target language dictionary, respectively.

At present, we can use several large Japanese machine readable dictionaries (MRD), but such JSL dictionaries are very small. For instance, the Japanese Sign Language Dictionary (JSLD94) contains 634 words. It is insufficient for translating Japanese into sign language. Towards this problem, we construct a new JSL MRD which contains 3162 words with several advises from deafs and sign translators.

However, the number of entries in a sign language dictionary is smaller than that of Japanese word dictionaries. It is difficult to realize a JSL dictionary including enough words in the near future. Furthermore, JSL vocabulary seems smaller than Japanese vocabulary because a JSL word may have a large number of senses (in comparison with a Japanese word), but its meaning will be determined by the context and nonverbal expression. Therefore, even if a JSL dictionary is perfectly constructed, it would have a smaller number of head words than Japanese word dictionaries would.

Towards the solution of this problem, we propose a method to get similar words from a Japanese word dictionary. If the input word does not exist in a sign language dictionary, SYUWAN substitutes similar words for the input word. This process is useful for finding an alternative word which exists in a sign language dictionary in order to enable the system to translate the input word into sign language.

We also make an experiment using daily news sentences. These are taken from "NHK Syuwa News" which consist of a Japanese sentence and a JSL. We analyze these Japanese sentences and translate into SLDM expressions. If these words are not entries in a JSL dictionary, SYUWAN tries to get similar words and translates them to the JSL words.

In section 2, a description method of sign language is explained. The sign language dictionary of MT system is described in section 3. The translation method of SYUWAN is illustrated in section 4. Section 5 describes some experimental results and problems of SYUWAN. Finally, conclusions and further works are shown in section 6 and 7, respectively.

Description of Sign Language

Many description methods of sign language are proposed, but they are not suitable to be used in a computer because they mix pictures or special characters together (Kanda94) (Honna90).

In this paper, we propose a simple but expressive description method named "Sign Language Description Method (SLDM)" which uses Japanese words and a few symbol characters as labels. In this description, Japanese words and sign language words are related in one-to-one manner (McCoy94) (Rouren90). For example, "Kyou, Hon wo katta (I bought a book today.)" is denoted by "Kyou / Hon / Kau (today / book / buy)". Although we have to keep records of a link information between Japanese labels and Japanese sign language (JSL) expressions, we determine the label which associates these relations in easier way and can be read and written easily by a human. The symbol characters express direction, time and finger spelling (the Japanese syllabary). For example, character "<" expresses a point something or hand direction. A SLDM sentence, "Kare <" (he <) expresses the action that the left hand takes the shape of "Kare" (he), and the right hand points to the left hand. The other SLDM sentence, "Kare < Kiku" (he < ask) expresses the left hand takes the shape of "Kare" (he), the right hand takes the shape of "Kiku" (ask) and moves to the left hand.

We translate Japanese into sign language based on SLDM. In this way, there is no need to process heavy task like image processing to show sign language using computer graphics. Moreover it is easy to map between input/output sign language and that of SLDM. In future we will connect SYUWAN to other input/output devices.

Sign Language Dictionary

Our system is a so-called transfer method which translates an input sentence into target language using some predefined rules. This seems to work well, because our sign language SLDM is very similar to Japanese language structure. In transfer methods, a dictionary of source and target languages as well as a set of structure transfer rules are needed. In this work, the Japanese EDR[1] (EDR95a) (EDR95b) dictionary and a JSL dictionary are used as the source language dictionary and the target language dictionary, respectively. The EDR dictionary is a machine readable dictionaries (MRD) which include enough words for Japanese language, but there exist only small JSL word dictionaries, such as the one used in (JSLD94) contains only 634 words. Although there are other dictionaries which contain up to 3000-6000 words (Rouren87) (ISLD84) (NSLD92), they include many useless words (dialect, obsolete, and so on). Towards this problem, we collected Japanese head words from several JSL dictionaries, re-edited them and constructed a larger sign language word dictionary which includes 3162 words. These words are common entries in several dictionary and used in real situation. We think these words are sufficient for daily conversation.

Translation from Japanese into Sign Language

We make a prototype Machine Translation system named "SYUWAN". It translates Japanese into JSL in 4-step processing as follows.

1. Japanese morphological analysis and removal of needless words.

---

[1] Japan Electronic Dictionary Research Institute, Ltd.

2. Application of translation rules and direct translation to sign language word.
3. Translation to sign language word using similar words.
4. Translation to finger spelling.

Currently, our translation method neglects dealing with the problem of word sense ambiguity and this problem is left as one of our further works.

Morphological Analysis

In Japanese, words are not bound by space. Therefore, it is necessary to determine a word pause, part of speech and inflection. We use JUMAN (Matsumoto94) for morphological analysis, which processes successfully 98 % words. After the analysis, SYUWAN removes needless words (auxiliary verb, etc.) from the output of JUMAN. We use news sentences in the experiment of SYUWAN this time, because the structures of these sentences in Japanese and JSL look similar, where an auxiliary verb does not influence a structure of sentence.

Direct Translation into Sign Language Words

As the next step, SYUWAN checks each word whether it exists in the JSL word dictionary. At this time, if it does then translate it to JSL word. SYUWAN applies the phrase translation rule for date expression, and termination translation rules for the end of word's terms at the same time. The former rule is a special expression in JSL, where the left hand expresses month and the right hand expresses day at the same time. For example, if there is a date expression like "9 gatsu 18 nichi"[2](September 18), SYUWAN applies a phrase rule, and output "/ 9 gatsu 18 nichi /"[3]. The latter rule applies to the ends of the word which is translated to symbol characters. Because this translation is difficult, we think processing them by the rule is more suitable.

Finding a Sign Language Word using Japanese Dictionary

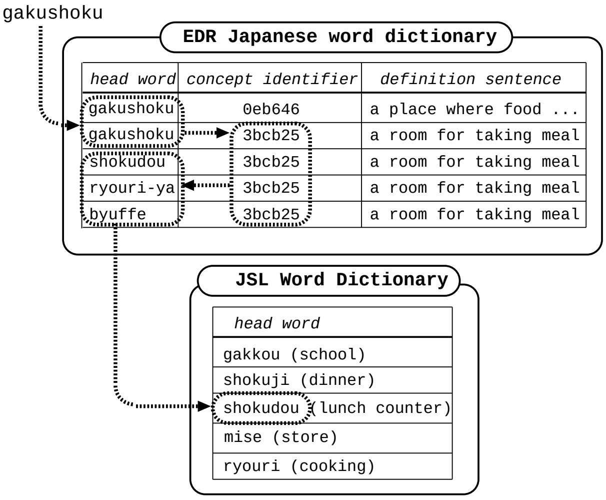
One of the most important problems in translation Japanese and JSL is the small size of Japanese label vocabulary in the JSL word dictionary. In the following experiment, half of the words in daily news sentences are not entries in the JSL dictionary. These words are failed in the direct translation step. Therefore, SYUWAN tries to find JSL words from the Japanese word dictionary. These words are resemble to a original Japanese word.

In the future, although the dictionary may be revised and some more entries may be added, there are still fewer words in the JSL dictionary than those in the Japanese dictionary. This is caused by the fact that we cannot distinct so many sign language words due to the limited combination of arms and hand shapes. Moreover, it is possible to apply our proposed techniques to learn a new JSL word and extend the JSL dictionary.

If the input word does not exist in the JSL word dictionary, SYUWAN tries to get similar words from the machine readable Japanese word dictionary, translates them to JSL words. SYUWAN uses EDR Electronic dictionary as its Japanese word dictionary. To derive similar words, three following methods are applied.

Using the concept identifier SYUWAN tries to get similar words which have the same concept identifiers with the input word from EDR Japanese Word Dictionary (EDR95a). This dictionary includes nearly 400,000 words entries. Each entry is composed of the head word, the concept identifier, the definition sentence of a concept (Japanese and English), and so on. The head words which have the same concept identifiers are similar in the sense than any other words in dictionary. Therefore, SYUWAN tries to get some head words which have the same concept identifiers with an original word, and tries to translate them to sign language word as the first and second steps.

For example, when SYUWAN translates the word "gakushoku (the refectory)" which is not an entry in the JSL dictionary, gets a concept identifier of gakushoku, and gets head words which have the same concept identifiers with gakushoku. The results are "shokudou (a lunch counter)", "byuffe (a buffet)", "ryouri-ya (a restaurant)" and so on. The word "shokudou" is only an entry in the JSL word dictionary. Finally, SYUWAN translates the word "gakushoku" to "shokudou" (Figure 2).



head word	concept identifier	definition sentence
gakushoku	0eb646	a place where food ...
gakushoku	3bcb25	a room for taking meal
shokudou	3bcb25	a room for taking meal
ryouri-ya	3bcb25	a room for taking meal
byuffe	3bcb25	a room for taking meal

head word
gakkou (school)
shokuji (dinner)
shokudou (lunch counter)
mise (store)
ryouri (cooking)

Figure 2: Using the concept identifier

---

[2] It is a Japanese sentence.

[3] It is a SLDM sentence.

Using the definition sentence of a concept As the second method, the definition sentence of a concept is used. The EDR Japanese word dictionary has a concept explanation which is definition sentence of a concept and a head word. SYUWAN extracts a list of words from a definition sentence using some extraction rules. These words are processed later in the same manner as the input is. These extraction rules are of two kinds, called "overlap-remove" rule and "toiu" rule. The former rule removes an overlap head word from the definition sentence. If the definition sentence contains the head word, SYUWAN processes recursively that word without termination. The overlap-remove rule avoid these situation. The latter rule depends on the expression of the EDR Japanese word dictionary. The definition sentence of a peculiar noun, is usually in a fixed form like a "Nippon toiu kuni" (A country called Japan). The "toiu" rule translates this explanation to "kuni / namae / Nippon" (country / name / Japan).

After the application of these rules, the derived explanation is processed by morphological analysis, a word list is extracted (Figure 3).

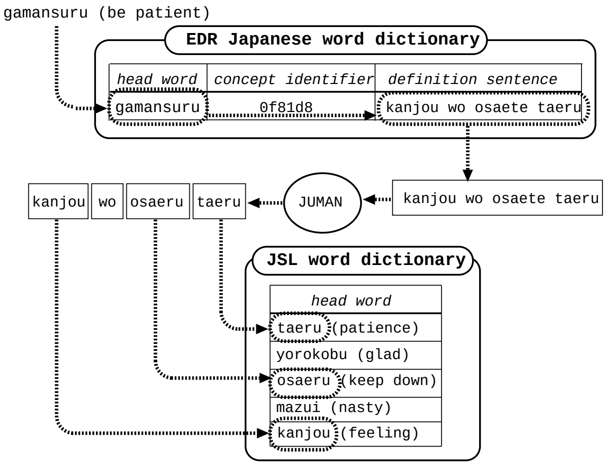


Figure 3: Using concept explanation

Using the concept hierarchy The last method is to get a super-concept word from EDR Concept Dictionary (EDR95b). In this dictionary, 510,000 concept identifiers are connected hierarchically. Each entry is related with some super-concepts and sub-concepts. It seems a super-concept is a abstraction of sub-concept, therefore, SYUWAN substitutes a super-concept's head word for a original word. For example, a "fish" concept connects "salmon", "sardine", "mackerel" as its sub-concepts, and "organism" as its super-concept (Figure 4). We think that the head word of super-concept is similar to the original input word.

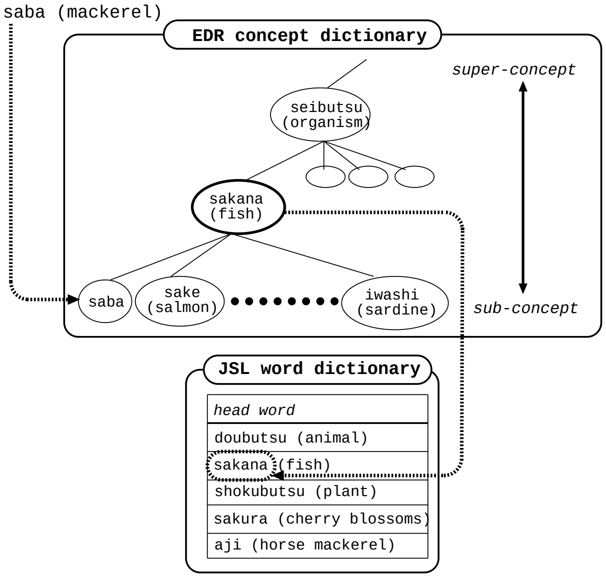


Figure 4: Using hierarchical concept

Translation into Finger Spelling

As the last step, SYUWAN translates the remaining words to finger spelling if they are nouns. The finger spelling expresses Japanese syllabaries (Kana). Each character of Japanese syllabaries corresponds to a finger shape uniquely. This method is used frequently in translation from Japanese noun words which do not exist in JSL word dictionary, such as technical terms. The finger spelling has one-to-one correspondence with Japanese syllabary. Therefore the translation into JSL always succeeds. However, this method does not consider word meaning and is applied for a noun at the last step of our system.

Experiments and Results

In the experiment of SYUWAN, we used daily news sentences as the input data. News sentences are taken from NHK[4]news for hard of hearing people. These news sentences consist of Japanese speech and sign language which are suitable for our work due to the following characteristics.

- It is easy to perform translation between Japanese and sign language, because the newscaster (not a deaf person) translates the news sentence, and there are many similar constructions between both structures.
- We can get unusual sign language words (politics and economy terms, etc.).

---

[4] Nippon Housou Kyoukai (Japan Broadcasting Corporation)

We prepared 91 sentences including 1303 terms, and we revised the result of morphological analysis processed by JUMAN beforehand.

SYUWAN processes an input in four steps, but cannot automatically select a best word which is obtained from EDR dictionary by using the similar-word method. We select and confirm the output by hand.

As the result of our experiment, 423 terms (32%) were needless for SLDM at the first step of morphological analysis. 297 terms (22%) were successfully at the step of direct translation to sign language word, 165 terms (12%) were successfully translated by translation rules. (Table 1)

In the next step, SYUWAN tried to get similar words from the EDR dictionary. At first SYUWAN got words which have the same concept identifiers, and 58 terms were found. All of these terms were translated successfully. Secondly SYUWAN got word lists from a concept explanation, and 269 terms were found. After applying the extraction rules, there were 254 terms which could be translated successfully, and 15 terms were useless. Lastly SYUWAN got super concepts of head words, as the result, and 15 terms were found. We confirmed the result by hand, 13 terms were translated successfully, but 2 terms were useless (disagreement a head word and a input word). In this step, 325 terms (24%) could be successfully translated.

In the last step, among the rest terms, 38 terms (3%) were translated to finger spelling, 55 terms (4%) were fail to be translated to sign language.

In conclusion, SYUWAN could succeed up to 95 % of translation from sentences. Among 880 terms which were needed to translate, 297 terms (34%) were directly translated, 165 terms (19%) were translated by rule, and 325 terms (37%) were translated by finding similar words.

Some failures, mainly in step 3, are caused by the following reasons.

- There are some useless definition sentence of a concept in the EDR Japanese word dictionary. For example, the head word of "saigen" (reappearance)'s definition sentence is a "saigen-suru" (reappear), and "mujitsu" (not guilty) 's definition sentence is a "tsumi ga mujitsu na koto" (It is not guilty). Even if the extraction rules applied, no useful words are obtained from these sentences, then it is impossible to find some similar words by this method. The EDR dictionary is going to be revised in the near future. We plan to make an experiment using the new version EDR dictionary and expect to get a better result.
- Some definition sentence of a concept does not include an appropriate words and too long for SLDM. The head word "tagaku (a lot of money)"s definition sentence is "suuryou ya teido ga takusan de aru koto. (a large quantity or a large number)". We prefer to get two words "money" and "many", but they are not included in the definition sentence.

  It is possible to solve this problem by the way that SYUWAN processes word explanation recursively. But it is not practical, because there are some noises added when SYUWAN processes a definition sentence of a concept. It is necessary to improve the processing precision of SYUWAN.
- In this experiment we do not consider a word sense, there are many cases that the head words disagree with correct senses. For example, when the word "ton" is processed, SYUWAN matches to a "ton" (tonnage), however, that word means a "ton chan" (Mr. Ton) the in the input sentence.

Another experimental failures are caused by some reasons as shown below.

No entry in dictionaries Some words like "gokigenyou" (How are you), "dewa" (as for) are not registered in both the EDR dictionary and the JSL word dictionary. We think that registering more words in the JSL dictionary, will solve this problem.

Kana expression If the input sentence is notated by kana, SYUWAN cannot recognize words in the sentence correctly. To solve the problem, we plan to use reading information in the EDR dictionary and if possible, we try to describe the input with kanji.

Processing of the ending of a word
When the passive sentence is translated, SYUWAN makes a mistake in recognizing the term and fails to translate.

This is caused by the fact that, JUMAN cannot process all sentences completely. We can improve this by adjusting the JUMAN system.

Conclusions

We proposed a new description for sign language named Sign Language Description Method which uses Japanese words as its labels. We constructed a large JSL dictionary including 3162 head words, and implemented a MT system SYUWAN which translates Japanese into JSL. When a word does not exist in the sign language dictionary, we proposed the following techniques to get a target word based on similarity from the EDR dictionary

- Using the concept identifier
- Using the definition sentence of a concept
- Using the concept hierarchy

Our system can succeed up to 95 % of translation.

Future works

We did not deal with word sense ambiguity in this work. It is not always a case that there is a one-to-one correspondence between a JSL word sense and a Japanese word sense. The way to automatically determine word sense will be taken into account as our further work.

all of terms in sentences : 1303					
needless terms	other terms : 880				
	Direct Translation	Apply Translation rules	Use similar word	finger spelling	failure
423 (32%)	297 (22%)	165 (12%)	325 (24%)	38 (3%)	55 (4%)

Table 1: Result

In this experiment, SYUWAN can translate 95% terms. This seems a good result. However, the result is evaluated based on words, not sentences. We plan to improve SYUWAN to output sign language sentences completely.

In this paper, we propose a translation method of term-to-term correspondence which tends to do a literal translation like a Manually Coded Japanese (MCJ). SYUWAN is a prototype which can be used for a learning purpose or beginners. In future, we will consider in more detail about the grammar of JSL and implement a more practical version of SYUWAN.

Acknowledgments

We are grateful to member of Okumura laboratory, especially Dr. Thanaruk Theeramunkong, for many comments and advises.

When we constructed our dictionary, we received many helps and many ideas from the deaf and sign translators in Ishikawa. Many cooperations and advises are obtained from a sign language circle named TENOHIRA in Nonoichi.

References

Hisahiro ADACHI, Shouzou YOSIZAWA, Masako FUJITA, Takashi MATSUMOTO, Kazuo KAMATA (1992). Analysis of News Sentences with Sign Language and Sign Translation Processing. Information Processing Society of Japan, Natural Language Processing, Vol.91-3. 17–24. (in Japanese)

Makoto Nishikawa, Mutsuhiro Terauchi. (1995) A Study on Japanese - Sign Language Translation Considering Placement of Sign Words. IEICE, Human Communication Science, Vol.94-94, 55–62. (in Japanese)

Takao KUROKAWA. Gesture Coding and a Gesture Dictionary for a Nonverbal Interface (1992). IEICE Trans. Fundamentals, Vol.E75-A, No.2. 112–121. February.

Jintae Lee, T.L.Kunii (1994). Generation and Recognition of Sign Language Using Graphic Models. IISF/ACM Japan International Symposium: Computers as Our Better Partners. 96–103. March.

Eiji OHIRA, Hirohiko SAGAWA, Tomoko SAKIYAMA, Masaru OHKI (1995). A Segmentation Method for Sign Language Recognition. IEICE Trans. Inf. & Syst. E78 D, 49–57, January.

KANDA Kazuyuki (1994). Syuwagaku kougi. Fukumura Shuppan. (in Japanese)

HONNA Nobuyuki, KATO Mihoko (1990). Syuwa no Hyouki-hou ni tsuite. Sign Language Communication Studies of The Quarterly of The Japan Institute for Sign Language Studies, No.4. pages 2–9. (in Japanese)

McCoy K.F., McKnitt W.M., Peischl D.M., Pennington C.A., Vanderheyden P.B., Demasco P.W. . AAC-user therapist interactions : Preliminary linguistic observations and implications for companision. Proc. of the RESNA '94 Annual Conference. 129–131.

Japanese Federation of the Deaf (1990). Sign Language for the Intermediate Class. (in Japanese)

KANDA Kazuyuki(1994). Japanese Sign Language dictionary. Alpha Media Co.,Ltd.

Syuwa communication Kenkyukai (1992). Shin Syuwa Jiten. Chuuouhouki.

MARUYAMA Kouji (1984). Irasuto Syuwa Jiten. Dynamic Sellers Publishing Co.,Ltd.

Japanese Federation of the Deaf (1987). The glossaries of the Japanese Signs (Watashi tachi no SYUWA in Japanese).

Japan Electronic Dictionary Research Institute LTD (1995). Japanese Word Dictionary Version 1.0.

Japan Electronic Dictionary Research Institute LTD (1995). Concept Dictionary Version 1.0.

NAGAO Lab., Kyoto University (1994). Japanese Morphological Analysis System JUMAN Manual version 2.0.