Machine Translation of Japanese into Sign Language

TOKUDA Masaaki
School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 1995

Abstract

Japan Sign Language (JSL) is very useful for deafs in Japan. JSL expressions are composed hands, arms and face expression, there are some differences between phonetics language like Japanese and English. But account of linguistics research in recent years, JSL has peculiar construction rule of sentences like other language, Japanese and English, too.

Sign Language is visual language. Therefore, it is very difficult to process in a computer. Recently, there have been many researches of recognition or generation of Sign Language words, but practical input/output device or system of Sign Language with computers are not available even now. Moreover processing are very heavy load for a computer. Leaving that question, it is necessary to express Sign Language by symbols in a computer for Natural Language Processing. The description method of Sign Language for record on a paper and Machine Translation (MT) system between Sign Language and Japanese are needed for the public and all of the people (of course Deaf and not Deaf), it is useful to put into practice there. In this paper, we propose "Sign Language Description Method (SLDM)" which is a description method of Sign Language using a Japanese words for label to express a word sequence, a hands direction and times clearly. SLDM is one to one correspondence between Japanese words and Sign Language words. For example, "Kyou, Hon wo katta (I bought a book today.)" is denoted by "Kyou/ Hon/ Kau (buy/ book/ today)". Although we have to keep on record of link information between Japanese labels and Sign Language expression, but we determine the label which associate these relations in easier way and can read/written by human. We try to take an approach to translate easly between input/output Sign Language and SLDM, dealing with a Japanese and SLDM using computer is not a heavy task like dealing with a image processing.

SYUWAN which translates from Japanese to Sign Language is transferred to a MT system. The Transfer method is analysis an input sentence, and translate to output of the target language. We think it is the best method because our Sign Language's SLDM resembles to Japanese on the notation. In the transfer method, we need a dictionary of source and target language and structure rules. In this case, Japanese and Sign Language dictionary are needed. Many Machine Readable Dictionaries (MRD) of Japanese exist thoroughly, but the existing Sign Language Dictionary has about 700 words, it is insufficient. We collect Japanese head words from several Sign Language Dictionary, re-edited and formed a MRD of Sign Language word Dictionary which includes 3162 words. When we choose words into a dictionary we received many helps and many ideas from deafs and Sign translators. In our experiment, we use daily news sentences, which resemble to Japanese and SLDM on a notation.

We make a prototype MT system named "SYUWAN" has 4 steps of processes as below, when translating from Japanese into Sign Language.

1. Japanese morphological analysis and removal of needless terms.
2. Application of translation rules and direct translation to Sign Language word.
3. Translation to Sign Language using resemble words.
4. Translation to finger spelling.

In the morphological analysis step, we use a morphological analysis system JUMAN. After Analysis, SYUWAN removes needless terms from output of JUMAN.

Next step, SYUWAN checks each terms whether it exists in the Sign Language Word Dictionary, if it does then translate it to Sign Language word. And SYUWAN matches the phase translation rule to date expression, and termination translation rules to the end of word's terms at the same time.

A most important problem in translation Japanese and Sign Language is the small number of vocabulary in Sign Language's Japanese label. If the input term does not exist in Sign Word Dictionary, SYUWAN tries to get a resemble word from machine readable Japanese dictionaries, translate it to Sign Word. SYUWAN using EDR (Electronic Dictionary Research) as machine readable dictionary, to get a word three following ways.

1. SYUWAN tries to get a resemble word which has the same concept identifier with input word from EDR Japanese Word Dictionary. This dictionary includes 400,000 word entries. Each entry is composed of head word, concept identifier, concept explication (Japanese and English), etc. We think that head words which have the same identifier are similar in the sense with each other.
2. The second method uses a concept explanation. We think concept explanation is a explanation sentence of head word. SYUWAN extracts a word lists from a explanation sentence will be transformed by extract rules and later be processed in the same way as input process. These extract rules are of two kinds, overlap remove rule and "toiu" rule. The former is overlap words in explication and head word. The latter is called description syntactic rule in JSL.
3. The last method is to get a super-concept word from EDR Concept Dictionary. EDR Concept Dictionary has 510,000 concept. Each entry is related sub-concept and super-concept which is abstraction. For example, fish's sub-concept is mackerel, and super-concept is organism. We think that head word of super-concept is more resemble than any other word except follow words.

Now SYUWAN is not automatic when many word are get as resemble word, to choice a best one word. In this experiment, that judge by hand.

The last step, SYUWAN translate the remaining words to finger spelling if they are nouns. This method is in use frequently at translation from foreign word which does not exist for Sign Word obviously.

In the experiment of SYUWAN, we used daily news sentences as input data. News sentences are NHK news for hard of hearing people. These news sentences consist of Japanese speech and Sign Language, we think that it is easy to perform translation between Japanese and Sign Language. We prepare 91 sentences as 1303 terms, and we check that result be mixed by JUMAN, revision of inputs terms pause. Result of experiment, 297 terms (22%) are successfully translated, 165 terms (12%) are successfully translated by translation rules. 423 terms (32%) are needless for SLDM from the beginning. 325 terms

(25%) are obtained by finding some resemble words from EDR dictionary. Among the rest terms, 38 terms (3%) are translated to finger spelling, 55 terms (55%) are failure to be translated to Sign Language. These failure are caused by the reasons as below.

- EDR Japanese word dictionary's concept explanations are useless. For example, the concept explanations is the same as head word that "saigen-suru (reappearance)" is "saigen"s explanation.
- The concept explanation is too long for SLDM. For example, head word "tagaku (a lot of money)"s explication is "syuuryou ya teido ga takusan de aru koto. (a large quantity or a large number)". We desire to get two words "money" and "many".

In conclusion, SYUWAN can succeed up to 95 % of translation.

In future, we study a method of automatic to choice a best one words when many resemble word are selected.

Keywords : Sign Language, machine translation, machine readable dictionary, resemble word