ARTIFICIAL AND HUMAN INTELLIGENCE (A. Elithorn and R. Banerji, editors).
Elsevier Science Publishers. B.V.

Chapter 11

A FRAMEWORK OF A MECHANICAL TRANSLATION BETWEEN JAPANESE AND ENGLISH BY ANALOGY PRINCIPLE

MAKOTO NAGAO
Department of Electrical Engineering, Kyoto University, Kyoto, Japan

Summary

Problems inherent in current machine translation systems have been reviewed and have been shown to be inherently inconsistent. The present paper defines a model based on a series of human language processing and in particular the use of analogical thinking.

Machine translation systems developed so far have a kind of inherent contradiction in themselves. The more detailed a system has become by the additional improvements, the clearer the limitation and the boundary will be for the translation ability. To break through this difficulty we have to think about the mechanism of human translation, and have to build a model based on the fundamental function of language processing in the human brain. The following is an attempt to do this based on the ability of analogy finding in human beings.

1. Prototypical consideration

Let us reflect about the mechanism of human translation of elementary sentences at the beginning of foreign language learning. A student memorizes the elementary English sentences with the corresponding Japanese sentences. The first stage is completely a drill of memorizing lots of similar sentences and words in English, and the corresponding Japanese. Here we have no translation theory at all to give to the student. He has to get the translation mechanism through his own instinct. He has to compare several different English sentences with the corresponding Japanese. He has to guess, make inferences about the structure of sentences from a lot of examples.

Along the same lines as this learning process, we shall start the consideration of our machine translation system, by giving lots of example sentences with their corresponding translations. The system must be able to recognize the similarity and the difference of the given example sentences. Initially a pair of sentences are given, a simple English sentence and the corresponding Japanese sentence. The next step is to give another pair of sentences (English and Japanese), which is different from the first only by one word.

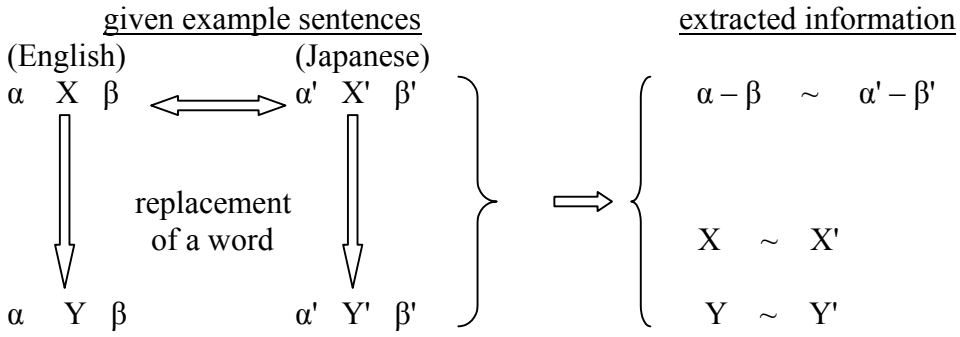


Fig. 1

This word replacement operation is done one word at a time in the subject, object, and complement positions of a sentence with lots of different words. For each replacement man must give the information to the system of whether the sentence is acceptable or non-acceptable. Then the system will obtain at least the following information from this experiment:

(1) Certain facts about the structure of a sentence;
(2) Correspondence between English and Japanese words.
These are expressed symbolically in Table 1.

These results indicate that we can formulate a word dictionary between English and Japanese, and a set of noun groups by a sentential context. If this experiment is done for different kinds of verbs the noun grouping will become much more fine and complex, and more reliable. Then certain kinds of relations will be established between word groups in a very complicated network structure. A noun may belong to several different groups with many different relations to other nouns. This is a kind of extensional representation of word meanings.

The same experiment can be done to verbs by replacing a verb in the same contextual environment. However, this is not so easy as noun replacement, because each verb has certain specific features as to the sentential structure, and no good grouping of verbs can be expected. So the sentential structure abstraction is done for each verb, and the structures are memorised in the verb dictionary entry for individual verb basis in such forms as (1).

This is a procedure, of finding the case frames for each verb mechanically. But to get a good and reliable result we have to have a huge amount of sample sentences which are carefully prepared. To distinguish word usages of similar nature, we sometimes have to prepare near miss sentences. The data preparation of this kind is very difficult, and the speed of learning of the linguistic structures by the system is very slow.

2. A modified approach

To improve this simple language learning process, we can think of the utilization of ordinary word dictionaries and thesauri. In an ordinary word dictionary a verb has, in the explanation part, typical usages of the verb in example sentences rather than by grammatical explanations. That is, typical sentential structures which the verb is governed by are given by examples. These dictionary examples give us, human beings, plenty of information as to the usage of verbs in constructing sentences. Man is guided by these examples, makes inferences, and generates varieties of sentences.

We want to incorporate this human process into our mechanical translation system. And for this purpose we need varieties of knowledge in our system. The knowledge the machine can utilize at the moment, however, is an ordinary word dictionary and thesaurus, which is of course not comparable to the human knowledge about the word and the sentences. A thesaurus is a system of word groupings of similar nature. It has the information about synonyms, antonyms, upper/lower concept relations, part/whole relations and so on. The thesauri available at present are all very old, and they are not satisfactory from our standpoint, but we can use them properly.

The most important function in the utilization of example sentences in an ordinary dictionary is how to find out the similarity of the given input sentence and an example sentence, which can be a guide for the translation of the input sentence. First the global sytactic similarity between the input and example sentences must be checked. Then the replaceability of the corresponding words is tested by tracing the thesaurus relations. If the the replaceability for every word is sufficiently sure, then the translation sentence of the example sentence is changed by replacing the words to the translation words of the input sentence. In this way the translation can be obtained.

For example, we are given an example sentence (2) in the table for the verb eat from an English-Japanese dictionary, and its translation as sentence (3). Suppose sentence (4) is given for translation. The system checks the replaceability (~) of the words (5) by tracing the synonym and upper/lower concept relations in a thesaurus. Because these are similar word pairs, the system determines that the translated example (3) can be used for the the translation of (4). From the dictionary the translation of the words (5) is (6) in the table, and the replaced result is (7) which is a good translation of the sentence (4).

When sentence (8) is given, the similarity check of (9) fails in the thesaurus, and no translation comes out. If this is an example sentence in an entry of eat, and has the Japanese translation (10), then the input sentence (11) can be translatable as (12).

The important point in this process is the recognition of the similarity between the input sentence and an example sentence in a dictionary. This completely depends on the structure of the thesaurus. Typical examples of YABURERU (be defeated, or be broken) are sentences (13) and (15), and the corresponding translations as (14) and (16).

Suppose we are given a sentence (17). To know which usage of YABURERU fits to this sentence, we check the words, president and vote in a thesaurus, and find out the relations (18). We can determine from this information that (17) is more related to (13) than to (15), and the translation is obtained as (19).

To do an experiment along these lines, we stored all the contents of an ordinary Japanese dictionary, and an ordinary English-Japanese dictionary and an English-English dictionary (Longman's) into computer files. We will have a Japanese thesaurus very soon. We want to have a good English thesaurus in computer usable form.

3. Machine translation by analogy

Our fundamental ideas about the translation are:

(1) Man does not translate a simple sentence by doing deep linguistic analysis, rather,

(2) Man does the translation, first, by properly decomposing an input sentence into certain fragmental phrases (very often, into case frame units), then, by translating these fragmental phrases into other language phrases, and finally by properly composing these fragmental translations into one long sentence. The translation

of each fragmental phrase will be done by the analogy translation principle with proper examples as its reference, which is illustrated above.

European languages have a certain common basis among them, and the mutual translation between these languages will be possible without great structural changes in sentential expressions. But the translation between two languages which are totally different, like English and Japanese, has a lot of difficult problems. Sometimes the same contents are expressed by completely different sentential structures, and there is no good structural correspondence between each part of the sentences of the two languages.

For example, a Japanese sentence (20) corresponds to such a different English sentence as (21) ~ (24). Another example is (25), which will literally correspond to such sentences as (26) ~ (28). But, it simply means (29).

A translation of this kind cannot be achieved by a mere detailed syntactic analysis of the original sentence. If we pick up each word and look for the corresponding translation word, the synthesis of a target language sentence becomes almost impossible. The choice of a proper translation from many candidates of a source language word is also very difficult without seeing the wider sentential context.

Therefore we adopted the method which may be called machine translation by example-guided inference, or machine translation by the analogy principle, and whose fundamental idea has been introduced already in the above. One of the strong reasons for this approach has been that the detailed analysis of a source language sentence is of no use for the translation between languages of completely different structure like English and Japanese. We have to see as wide a scope as possible in a sentence, and the translation must be from a block of words to a block of words. To realize this we have to store varieties of example sentences in the dictionary and to have a mechanism to find out analogical example sentences for the given one.

It is very important to point out that, if we want to construct a system of learning, we have to be able to give the system the data which is not very much processed. In our system the augmentation of the knowledge is very simple and easy. It requires only the addition of new words and new usage examples and their translations. It does not require the information which is deeply analyzed and well arranged. Linguistic theories change rapidly to and fro, and sometimes a model must be thrown away in a few years. On the contrary, language data and its usage do not change for a long time. We will rely on the primary data rather than analysed data which may change sometimes because of changes in the theory.

4. A practical approach

The process of mechanical translation by analogy is again very time consuming in its primary structure. So we divide the process into a few substages and give all the available information we have to the system, in the initial system construction. The learning comes in only at the augmentation stage of the system, which is mainly the increase of example sentences and the improvement of the thesaurus.

The following substages have been distinguished in our Japanese English translation system which is being constructed.

(a) Reduction of redundant expressions, and supplement of eliminated expressions in a Japanese input sentence, and getting an essential sentential structure. Sentence (30) has almost the same meaning as sentence (31).

(b) Analysis of sentential structure by case grammar. Phrase structure grammar is not suitable for the analysis of Japanese, because the word order in Japanese is almost free except that the final predicate verb comes at the end.

(c) Retrieval of target language words, and example phrases which are stored in the word entries from the dictionary. The dictionary contains varieties of examples besides grammatical information, meaning and, for verbs, the case frames.

(d) Recognition of the similarity between the input sentential phrases and example phrases in the dictionary. The word thesaurus is used for the similarity finding.

(e) Choice of a global sentential form for translation. For example, sentence (32) has such translations as (33) and (34). These can only be derived from the examples for the word result in the dictionary.

(f) The choice of local phrase structure is determined by the requirements of the global sentential structure.

It is very difficult to clarify what factors contribute to the determination of the stages (e) and (f). These remain to be solved.

Table 1

(1) S・verb・O・C ⟷ S′は・O′を・C′に・verb′,

S, S′ ∈ W_X, O, O′ ∈ W_Y, C, C′ ∈ W_Z

where W_X, W_Y, and W_Z are semantic groups of words X, Y, Z.

(2) A man eats vegetables.

(3) 人は 野菜を たべる。
(man) (vegetable) (eat)

(4) He eats potatoes.

(5) man ～ he

vegetable ～ potato

(6) 人 ～ 彼
(man) (he)

野菜 ～ じゃがいも
(vegetable) (potato)

(7) 彼は じゃがいもを たべる。

(8) Acid eats metal.

(9) acid ～ man

metal ～ vegetable

Table 1 (continued)

（１０） 酸は　金属を　侵す。
(acid) (metal) (eat / invade / attack)

（１１） Sulphuric acid eats iron.

（１２） 硫酸は　鉄を　侵す。
(sulphuric acid) (iron) (eat)

（１３） 彼は　選挙に　破れた。
(he) (election) (be defeated)

（１４） He was defeated by the election.

（１５） 紙袋は　重みで　破れた。
(paper bag) (weight) (be broken)

（１６） The paper bag was broken by the weight.

（１７） 大統領は　投票に　破れた。
(president) (vote)

（１８） 大統領 ～ 人
(president) (man)

投票 ～ 選挙
(vote) (election)

（１９） The president was defeated by the vote.

（２０） 残念　ながら　明日は　行け　ません。
(regret / disappointment) (though / inspite of / while / with) (tomorrow) (go / visit / attend) (not)

（２１） To my regret I cannot go tomorrow.

（２２） I am sorry I cannot visit tomorrow.

（２３） It is a pity that I cannot go tomorrow.

（２４） Sorry, tomorrow I will not be available.

Table 1 (continued)

（２５） 国際政治 の 事 について 書いた 本。
(international politics) (of) (matter / thing / affair / situation / event) (about / of / on / with) (write / draw) (book / volume / work)

（２６） a book in which the affairs of international politics is written.

（２７） a book in which (someone) wrote about the events of international politics.

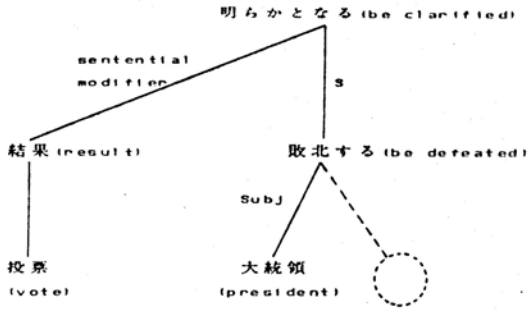
（２８） a book written about the events of international politics.

（２９） a book on international politics.

（３０） 日本語の 翻訳の 場合に ついては， 難しい 問題が ある。
(Japanese language) (translation) (case) (about) (difficult) (problem) (exist)

（３１） 日本語の 翻訳 には 難しい 問題が ある。

（３２） 投票 の 結果 大統領 の 敗北が 明らかと なった。
(vote) (of) (result) (president) (of) (defeat) (clear / evident) (become)



（３３） As the result of the vote the defeat of the president becomes definite.

（３４） The result of the vote revealed that the president was defeated.