A CONSIDERATION ON THE CONCEPTS STRUCTURE AND LANGUAGE
— IN RELATION TO SELECTIONS OF TRANSLATION EQUIVALENTS OF VERBS IN MACHINE TRANSLATION SYSTEMS —

Sho Yoshida

Department of Electronics, Kyushu University 36,
Fukuoka 812, Japan

ABSTRACT

To give appropriate translation equivalents for target words is one of the most fundamental problems in machine translation systrms. Especially, when the MT systems handle languages that have completely different structures like Japanese and European languages as source and target languages. In this report, we discuss about the data strucutre that enables appropriate selections of translation equivalents for verbs in the target language. This structure is based on the concepts strucutre with associated information relating source and target languages. Discussion have been made from the standpoint of realizability of the structure (*e.g.* from the standpoint of easiness of data collection and arrangement, easiness of realization and compactness of the size of storage space).

1. Selection of Translation Equivalent

Selection of translation equivalent of a verb becomes necessary when,
(1) the verb has multiple meanings, or
(2) the meaning of the verb is modified under different contexts (though it cannot be thought as multiple meanigns).

For example, those words 'する', '演奏する', '弾く', '遊ぶ', 'ごっこをする ', 'ゆれ動く ', ••• are selectively used as translation equivalents of an English verb 'play' according as its context.

1. play tennis : テニスをする
2. play in the ground : グランドで遊ぶ
3. The children were playing ball (with each other) : 子供達はボールごっこをしていた
4. play piano : ピアノを弾く
5. Lightning palyed across the sky as the storm began : 嵐が始まるとゆれ動いた

In the above examples, they are not essentially due to multiple meanigns of 'play' but need to assign different translation euqivalents according as the differences of contexts in the case of 1. to 3., and due to multiple meanings in the cases of 4. or 5.

A typical idea for selecting translation euqivalents so far is shown in the following example.

Lets take a verb 'play'. If the object words of the verb belong to a category $C_{obj}^{play:\text{する}}$, we give a verb 'する'(=do) as its appropriate translation equivalent. If the object words belong to a category $C_{obj}^{play:\text{弾く}}$, we give '弾く' as an appropriate translation equivalent of 'play'.

Thus, we categories words (in the target language) that are agent, object, ••• of a given verb (in the source language) according as differences of its appropriate translation equivalents.

In other words, these words are categorized according as "such expression as a verb with its case filled with these words be afforded in the target language or not", and are by no means categorized by their concepts (meaning) alone.

For example, for tennis, baseball, ••• $\in C_{obj}^{play:\text{する}}$ ={tennis, baseball, card, •••}, translation of 'play' are given as follows.

play tennis : テニスをする
play baseball : 野球をする
play card : カードをする

To the words belonging to $C_{obj}^{play:\text{弾く}}$ = {piano, violine, harp, ••• }, the translation equivalent of 'play' is given as follows.

play piano : ピアノを弾く
play violine : バイオリンを弾く
play harp : ハープを弾く

Categories given in this way have a problem that not a small part of them do not coincide with natural categories of concepts. For example, members 'テニス(tennid)' and '野球(baseball)' of a category $C_{obj}^{play:\text{する}}$ belong to a natural category of concepts 球戯(ball game), but 'カード(card)' does'nt. Instead it belongs to a conceptual category 遊戯(game in general). 球戯 is considered as a sub-category of 遊戯. Therefore, if we regard $C_{obj}^{play:\text{する}}$ as 遊戯, then テニス(tennis), カード(card), フットボール (football), ゴルフ (golf), ••• can be members of it, but 碁(go), 将棋(shogi) which also belong to the conceptual category 遊戯, are not appropriate as members of $C_{obj}^{play:\text{する}}$. ('play go : 碁をする', 'play shogi : 将棋をする' are not appropriate, instead we say 'play go : 碁を指す', 'play shogi : 将棋を指す')

Therefore, $C_{obj}^{play:\text{する}}$ should be devided into two categories $C_{obj}^{play:\text{する}}$ and $C_{obj}^{play:\text{指す}}$.

The problem here is that, such division of categories do not necessarily coincide with natural division of conceptual categories. For

example, translation equivalent '指す' cannot be assigned to a verb 'play' when object word of it is チェス (chess), which is a game similar to 碁 or 将棋. Moreover, if the verb differs from 'play', then the corresponding structure of categories of nouns also differs from that of play. Thus we have to prepare different structure of categories for each verb.

This is by no means preferable from both considerations of space size and realizability on actual data, because we have to check all the combinations of several ten thousands nouns with each verb.

2. Concepts Structure with Associated Information

So we turn our standpoint and take natural categories of nouns (concepts) as a base and associate to it through case relation pairs of a verb and its translation equivalent.

Let a structure of natural categories of nouns were given (independently of verbs).

A part of the categories (concepts) structure and associated information (such as a verb and its translation equivalent pair through case relation etc.) is given in Fig.1.

In Fig.1, verbs associated are limited to a few ones such as Do (obj = musical instrument) ⇒ Play (obj = musical instrument). Becsuse, from the definition of musical instrument : "an object which is played to give musical sound (such as a piano, a horn, etc.)", we can easily recall a verb 'play' as the most closely related verb in this case.

It can generally be said that the more the noun's relation to human becomes closer and the more the level of abstract of the noun becomes lower the numbers of verbs that are closely related to them and therefore have to associate to them (nouns) become large. And that the numbers of associated ideoms or ideom like parases become large. Therefore, the division of categories must further be done.

The process of constructing this data structure is as follows.

(1) Find a pair of verb and associated translation equivalent (Do⇒ Play : 演奏する) that can be associated in common to a part of the structure of the categories as in Fig.1, and then find appropriate translation equivalents in detail at the lower level categories.
(2) To each verb found in the process of the association, consults ordinary dictionary of translation equivalents and word usage of verbs and obtain the set of all the translation euqivalents for the verb.
(3) Then find nouns (categories) related through case relation to each translation equivalent verb thus obtained by consulting word usage dictionary. Then check all the nouns belonging to nearby categories in the given concepts structure and find a nouns group to which we associate the translation equivalent.

In this manner, we can find pairs of verb and its translation equivalent for any noun belonging to a given category. To summarize the advantage of the latter method, (1) to (4) follows.

(1) The only one natural conceptural categories structure should be given as the basis of this data structure. This categories structure is stable, and will not be changed basically, and is constructed independently from verbs. In other words, it is constructed indepndently from target language expression.
(2) To each noun in a given conceptual category, numbers of associated pairs of verb and its translation equivalent are generally small and can easily be found.
(3) Association of the pair of verb and its translation equivalent through case relation should be given to one category for which the association hold in common for any member of it. In Fig.1, a conceptual category $C_{obj}^{play:弾く}$ is created from two categories 鍵盤楽器 (keyboad musical instrument) and 弦楽器 (string musical instrument) for this purpose. And then associate through case relation specific pair of verb and its translation equivalent to exceptional nouns in the category.
(4) From (1) to (3), it follows that this data structure needs considerably less space and is more practical to construct than the former method.(chapter 1)

3. Concluding Remarks

We proposed a data structure based on concepts structure with associated pairs of verb and its translation equivalent through case relations to enable the appropriate selections of translation equivalents of verbs in MT systems.

Additional information that should be associated to this data structure for the selections of translation equivalents is ideoms or ideom like phrases. The association process is similar to the association process in chapter 2.

Only the selections of translation equivalents for English into Japanese MT have been discussed on the assumption that the translation equivalents for nouns were given.

Though the selection of translation equivalents for nouns are also important, the effect of application domain dependence is so great that we strongly relied on that property at the present circumstances.

There are cases that translation equivalents are determined by pairs of verbs and nouns to each other. So we need to study the problem of selection of translation equivalent also from this point of view.

Reference

(1) Sho Yoshida : Conceptual Taxonomy for Natural Language Processing, Computer Science & Technologies, Japan Annual Reviews in Electronics, Computers & Telecommunications, CHMSHA & North-Holland, 1982.

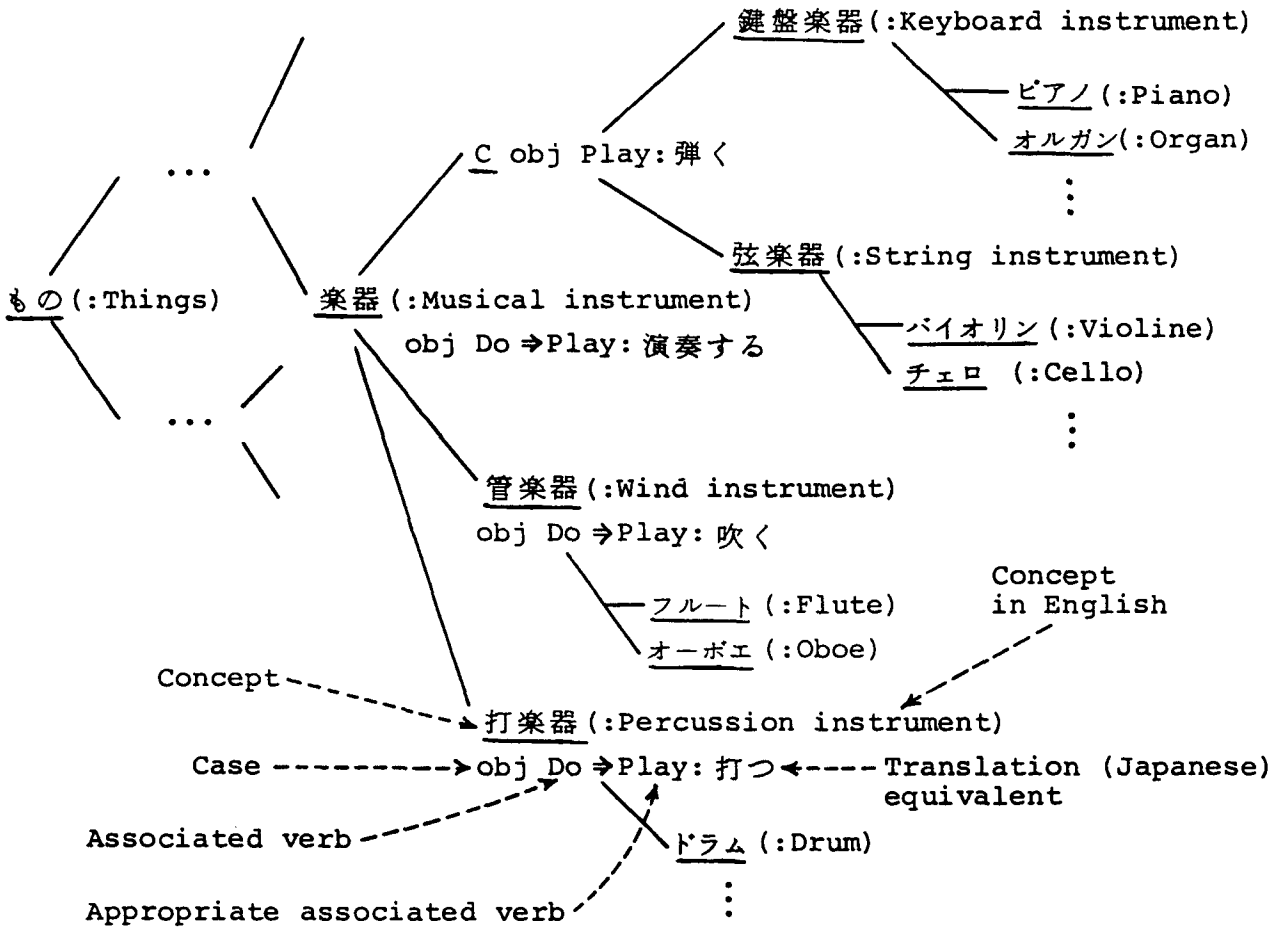

Fig. 1 A Part of Concepts Structure with Associated Information