GENERAL MT AND UNIVERSAL GRAMMAR

Paper read before the Conference on Mechanical Translation, June 17 - 20, 1952 at the M.I.T., Cambridge, Mass., by Dr. Erwin Reifler, Associate Professor of Chinese, University of Washington, Seattle

In a previous manuscript[1] I have already indicated the advisability of distinguishing between SPECIFIC and GENERAL MT. Specific MT is concerned with the translation from one code language into one particular target language. General MT, on the other hand, has to do with the translation from one code language into many target languages. Here we are dealing only with General MT.

General MT poses the following question: Is it possible to solve the problems of MT in such a way that one and the same preparation of the code text may serve for a MT into many languages? In my first paper presented to this Conference[2] I have already outlined how the mechanical determination of all incident meaning may be made possible by a special orthography supplementing the graphic distinctiveness of certain grammatical meanings of the code language. Thus the question whether one and the same preparation of the code text can suffice for General MT will ultimately depend on the extent to which we can correlate grammatical meanings of different languages. This raises a number of problems closely connected with questions of Universal Grammar.

At first sight the situation seems hopeless. Linguistics tells us that such a thing as a Universal Grammar does not exist. Linguists do, however, speak of language universals, Leonard Bloomfield, for instance, says: "A task for linguists of the future will be to compare the categories of different languages and see what features are universal or at least widespread".[3] And again: "..... a form class comparable to our substantive expressions, with a class meaning something like 'object', seems to exist everywhere....."[4] The same can be said of other form classes.

But there is also evidence of other types of language universals. We find, for example, highly interesting cases of parallel development in the evolution of grammatical meaning. The comparative linguists Osthoff and Walde have suggested the possible cognacy of the Indo-European pronominal "*k^{w}i-" or "*k^{w}o-" (ancestor of Latin "quis, qui, quae, quod, etc," and also of English "who, what, why, when, whore, etc.") and the Latin verb "quire", meaning "may" or "can". The same phenomenon is found in the Hebrew indefinite pronoun "kol", meaning "whosoever, every, all" and the verb "ya-kol", meaning "can". The most fascinating parallel, however, is supplied by the Chinese words represented by the characters [Ch*] and [Ch]. The first is the most common classical indefinite, relative and interrogative pronoun, the second is a verb with the meanings "nay, can". You will notice that the character for the second word occurs in the character of the first. This is no mere coincidence. Both words are also phonologically closely related. Moreover, in ancient inscriptions [Ch], that is the unaugmented character, is also found in the sense of [Ch]. These examples would indicate that in the three unrelated families of

---

* [Chinese characters (Ch) are not reproduced in this transcript – *Ed.*]

[1] Studies in Mechanical Translation, No. 1: MT (Jan. 10, 1950), #40.

[2] Studies in MT, No. 3: MT WITH A PRE-EDITOR AND WRITING FOR MT.

[3] LANGUAGE, p. 270.

[4] Ib., pp. 270 and 271.

languages from which they are taken the pronominal concept has developed from an earlier meaning of something like "may" or "can", that is from the concept of "possibility".[5]

This particular semasiological coincidence does not seem to be of any consequence for MT. But take, for instance, the case of Chinese "shih[4]" ([Ch]) which as a free form is used in modern literary Chinese in the sense of "this, those" as well as in the sense of forms of the verb "to be". The last meaning is comparatively late, ancient Chinese does not know it. In another paper soon to be published[6] I have shown that the same phenomenon occurs in a large number of languages. In the same paper I have pointed out the strong probability that also the meaning of English "is" evolved from an earlier demonstrative-pronominal concept. The available phonological information could well support such a semantic derivation. The semantic content of sentences like "John is a child" was in curly times apparently given the form "John this: a child".

As a consequence of such considerations we may, for instance, decide to correlate Chinese "shih[4]" ([Ch]) with English "this, these" also in those cases in which it corresponds to forms of the verb "to be". This would not give us an idiomatic translation, but it would be intelligible and simplify our mechanical correlation problems. I may mention here that this latter approach is on the lines of Dr. Warren Weaver's fourth type of attacks on the semantic difficulties of MT. He speaks of "an approach that goes so deeply into the structure of languages as to come down to the level where they exhibit common traits".[7]

But apart from a large number of universals actually shared by many languages there exists also another kind of universals which in a previous paper[8] I have called "pseudo-universals", namely features in languages to which we arbitrarily impute characteristics of certain well-known universals they actually do not have. The following example will elucidate these universals:

Japanese "hito ga kuru" corresponds in meaning to English "a person comes". "Hito" means "a person", "kuru" means "comes". "Ga" is a particle which is conveniently explained as a formal indicator of the nominative case of the preceding noun. But in fact "ga" is a marker of something like a genitive relation and not a marker of the nominative, although the descriptive linguist may in terms of his descriptive system and within the frame of reference of modern Japanese conveniently describe it as a particle denoting the actor. Nor is "kuru" here a verb, but a noun. The Japanese do in this case actually not express the idea "a person comes" by something like "a person comes", but by something like "the coming of a person". Professor Sansom has compared such forms of Japanese statements with English newspaper head-lines like "Death of Jones" which means nothing else but "Jones is dead".[9] It is as if the Japanese first become aware of something coming and then that it is a person that

---

[5] Cf. Erwin Reifler, ETUDE SUR L'ETYMOLOGIE DES CARACTERES CHINOIS. LA SERIE [Ch], Bulletin de l'Université l'Aurore (BUA). 17, III. Tome 5, No. 1, Shanghai,1944, and LINGUISTIC ANALYSIS, MEANING AND COMPARATIVE SEMANTICS Problem No, 4 (submitted for publication to LINGUA, Haarlem, Holland).

[6] Ib., Problem No. 5.

[7] Warren Weaver, Manuscript on MT dated July 15, 1949, p. 11.

[8] Studies in MT, No. 2: SOME PROBLEMS OF THE MECHANICAL TRANSLATION OF LANGUAGES.

[9] Historical Grammar of Japanese, p. 228.

comes, and, therefore, mentally and grammatically subordinate the actor to the action. For all practical purposes, however, we may attribute the function of a nominative particle to "ga" and then proceed to say that Japanese shares, in sentences of this type, in the fairly widespread feature of a formal indicator of the nominative case or of the actor. A similar situation we find in classical Chinese where, what is usually described as a marker of the possessive, is in fact an anaphoric pronoun. "A person comes" assumes there the form of something like "a person, this come".

We are able to impute the characteristics of more or less general "universals" to a very large number of features in many languages and many language teachers actually are doing this daily for practical purposes. This may cause the linguist to shudder in horror. But this fact is extremely beneficial for mechanical translation. The concept of "pseudo-universals" permits us to greatly extend the territory of language universals. The purely linguistic evaluation of linguistic phenomena is less important for us than the use we can make of them.

There is, however, also another way of artificially increasing the number of language universals than by arbitrarily attributing grammatical meanings to linguistic forms which they, in fact, do not have, namely by changing the structure of a language. We may, for instance, within the limitations of intelligibility, so modify the grammar of a language as to bring it more in line with the grammar of other languages. *I* shall exemplify this procedure further below when discussing the problem of ADJUSTED MODEL TARGET LANGUAGES.

All these treatments of the language problems of MT may prove useful for the mechanization of the translation process. But we have to keep in wind the fundamental difference in the practical requirements of the code and target sides of MT, already outlined in my previous paper on MT.[10] On the code side we should interfere neither with the language nor the conventional "spelling" of the code texts, The only thing with which we may interfere here is their so-called orthography, that is not the conventional alphabetization of words but the form of letters. On the code side we have, furthermore, to ask ourselves whether those types of grammatical meaning whose graphic distinctiveness is essential for the mechanical determination of incident meaning occur in, or can be attributed to, all code languages with MT value. There can be no doubt that, either in fact or via the concept of "pseudo-universals", we can find them in all such code languages. In all such languages we are able to distinguish nouns, principle and auxiliary verbs, adjectives, adverbs, subjects, predicates, direct and indirect objects, present, past and future tense, number, person, etc, etc,

On the target side the situation is different. Here we should neither interfere with the conventional "spelling" nor with the conventional "orthography" as we defined these two terms in our previous paper. But here we can, within the limits of intelligibility, interfere with the language itself. This leads us to the problem of ADJUSTED MODEL TARGET LANGUAGES.

Professor Stuart C. Dodd has in his paper MODEL ENGLISH FOR MECHANICAL TRANSLATION demonstrated a regularized form of English - that is, an English in which, for example, we would not say "brought", but "bringed", not "oxen", but "oxes", not "I am, you

[10] See footnote 2.

are, he is", but "I be, you be, he be", etc, etc. In working out the principles of such an English, Professor Dodd at first thought in terms of a practical world language which can be easily learned by foreigners.

Now it is clear that, if we can put a regularized target language into the dictionary mechanism, we can greatly simplify our MT problems and the engineering work involved. Instead of a correlation between two irregular languages with mostly great differences in their irregularity we would only be faced with the problem of a correlation between one irregular and one regular language. With MT in mind we can, moreover, go a little further than was necessary for a world language. We can construct a different Model English for each of the foreign code languages, namely a Model English adjusted to the peculiarities of each foreign language. The limits to which we may go here are those prescribed by intelligibility for the English reader. For example:

Professor Dodd expresses the past tense of every English verb by a preceding "did", and the future by a preceding "will". Now both German and modern Mandarin have, in terms of "pseudo-universals", an equivalent for English "will" (German "wird", Mandarin "yao^{4}" or "chiang^{1}", both preceding the verb). But neither of these two languages has an equivalent for English "did". However both German and modern Mandarin have an equivalent for English "-ed" in "decided", "bringed", etc, (German "-te" and Mandarin "-la", suffixed or following the verb). Thus in a German-English and Mandarin-English MT unit the English Model Language we would put into the Mechanical dictionary should express the past tense rather by "-ed" than by "did", because then we would have a mechanical one-to-one correlation,

The use of Adjusted Model Target Languages in conjunction with the application of the concept of "pseudo-universals" will go a long way to solve the problems posed by non-grammatical meanings in different grammatical garb. For example one possible Mandarin version of English "he walks quickly", namely "t'a^{1} tsou^{3}-ti k'uai^{4}", contains words also meaning something like "he", "walk" and "quick". But "k'uai^{4}" corresponds here rather to something like English "to be quickness" or "to be quick" whereas "tsou^{3}-ti" means something like "walk's" or "of walk", A literal translation would give something like "he is a quickness of walk" or "he is quick of walk". But in other contexts "tsou^{3}-ti" is often freely translated "walking", which amounts to an arbitrary equation of Chinese "-ti" with "-ing". We may therefore render the Mandarin sentence by "he walk-ing quick". This is bad English, but perfectly intelligible and, because it permits a word-to-word translation, has the great advantage of simplifying the mechanical correlation problem.