TRANSLATION WITHOUT MACHINE

YUEN REN CHAO

If the human organism is viewed as a machine, then all translation is machine translation, though obviously that will be so only in a trivial sense. In the present paper I propose to consider those aspects of translation which are not likely to be taken care of in the near future by non-human machines. Most of these factors, to be sure, have already been considered by speakers at this Congress and writers on the translation problem in recent years or even in the distant past. The point of my paper is to present them with a subjective estimate of their remoteness from or propinquity to machine translation.

The process of translation may be viewed as a space of many dimensions in each of which a translation is good, indifferent, or bad. Since many of the factors will conflict, the total result should be a somewhat complicated function of them. It is indeed premature even to speak of functions, since these dimensions are still pre-systematic, undefined notions, rather than measured quantities. For my present purposes, I shall only consider the various dimensions separately, without attempting to set up any overall evaluation of translation in general or of actual translations in particular. I shall consider in turn 1. Physical makeup of the text; 2. Size of unit to be translated; 3. Style; 4. Grammatical Structure; 5. Subject Matter and cultural categories; and 6. Pragmatics of translation.

1. PHYSICAL MAKE-UP OF THE TEXT

Most translations have to do with written into written text between different languages. However, it may be useful to take a broader perspective by considering other physical forms of texts. In the work of interpreters, the "text" is in the form of live speech. As is well known, the work of written translation and that of the so-called simultaneous translation at the United Nations call for very different kinds of skills and belong to quite different parts of that organization. For good, theoretical reasons, modern linguistics have taken spoken sounds as the proper study of language, but for practical and equally good reasons, machine translation has so far chiefly concerned itself with visual messages for both the original language and the target language, and thus brings back to a more respectable status the expression "written

Language" which we linguists have been looking down upon. However, now that so much work is being done on automatic speech recognition at the acoustic input end as well as on speech synthesis at the acoustic output end, a three stage machine translation from speech to speech will not be many years farther off than graph to graph machine translation. As a rough guess, I would venture to suggest the stages of progress somewhat as follows: (1) before 1965, translation without machine for both writing and speech; (2) translation with machine for writing will begin some time between 1965 and 1970, but still without machine for speech; (3) translation for both speech and writing will begin some time between 1970 and 1975. It may be noted in passing that at stage (3), it may not be necessary to go through the ordinary orthographies of the languages involved, which have to be coded anyway when going through the machine, but may be by-passed through phonemes or their coded equivalents. This will be the case especially with a language in which the writing system is in larger units than phonemes.

2. SIZE OF UNIT TO BE TRANSLATED

The piece to be translated may be of the size of a book, a play, an article, a lecture, a poem, a letter, or a speech, each as a more or less complete piece of discourse, for which there is usually a best translation in a target language. But even here one may have to go beyond the text (cf. 6 below) to decide upon a translation; in fact, otherwise there would be no philology. A unit of the size of a sentence may admit of more than one possible translation and it may need either the linguistic or the situational context to determine in what way the original is to be understood and translated. Such determination, if needed, will still to a large extent be a non-machine part of the translation under present conditions.

When we come down to the sizes of phrases, words, and morphemes, then the absence of one-to-one correspondence between languages becomes even more of a problem, as can be verified by opening any bilingual dictionary. Much of present day research in machine translation, as you are aware, consists in compiling and coding for machine operation units of these sizes, especially at the word level,[1] and in finding automatic ways of decisions on multiple choice by scanning over as little context as possible and as much context as necessary.

Finally, when we come to the size of phonemes, then translation between different languages is reduced to a vacuous case with a correlation of almost zero. In other words, any phoneme in one language translates into any phoneme in another language and the conditions of correspondence will have to be determined by factors from the morpheme level up. The same can be said of distinctive features, if we go beyond

[1] Morris Swadesh puts it more broadly thus: "the unit of translation ... corresponds to the linguist's structural unit", in his article "On the Unit of Translation", *Anthropological Linguistics* 2.2.40 of 39-42 (February 1960).

phonemes to distinctive features. I said that the correlation was "almost zero" because when one translates poetry and song, or wit and humor, even phonetic comparability becomes relevant.

Suprasegmental morphemes (not counting phonemic tones), however, are more translatable, like other morphemes. They are more amenable to machine treatment in that some of them approach cases of language universals, such as rising or higher pitch for suspense and falling or lower pitch for conclusion;[2] on the other hand, they are less amenable to machine translation or even non-machine translation, since they usually are not written at all in conventional orthography and are thus literally lost sight of and neglected, even though they may form an essential part of the message.

3. STYLE

Comparability of style between the original and the target language is of course an important desideratum. If possible, one wants to translate prose into prose, poetry into poetry, archaic into archaic diction, colloquial into colloquial, and slang into slang.[3] On the whole, since machine translation is now fully occupied with problems of multiple choice in lexical units and with rendering of syntactical and morphological interchange, one will have to leave to non-machine translation to take care of most of the problems of comparability of style.

However, one aspect of style, that of the frequency of occurrence of items, seems to permit quantitative treatment and thus partial machine treatment. Everyone is familiar with the special effect produced when a phrase or sentence in one language is rendered word for word into another. Assuming that matters of grammar and vocabulary have been taken care of, the disparity of frequency of occurrence of the correspondence will make the effect either fresh and interesting, or dull and flat, or strange and bizarre, or even unintelligible. I do not say that rendering items into those of comparable frequency of occurrence will necessarily result in a good translation, but its total disregard will tend to contribute to disparity of style. It may, therefore, be useful for dictionary entries, for translation purposes, to contain not only meaning and function, and perhaps style (as some dictionaries already do: *Slang, Arch., etc),* but also the frequency of occurrence.[4]

Related to this factor of frequency is the size of units considered in the preceding discussion, especially at the level of the phoneme and the syllable. Roughly speaking, the variety of kinds of units is a decreasing function of the number of units (i.e. size) needed to carry a given amount of information. For example, with a small inventory

---

[2]   See Dwight L. Bolinger, "Intonation as a Universal", p. 833 of this volume.
[3]   On problems of style see for example J. P. Postgate, *Translation and Translations* (London, 1922); Ronald Knox, *Trials of a Translator* (New York, 1949).
[4] The only dictionary I know of which has that is C. H. Fenn's *The Five Thousand Dictionary,* 5th ed. (Peking, 1940), and Amer. ed. (Cambridge, 1942), in which entries are graded approximately by frequency, for teaching purposes, from A to K.

of phonemes and syllables in Japanese, it takes more syllables to make a morpheme than in languages with a larger inventory. But since length and syllabicity are factors of style, especially in poetry, any disparity in this respect between two languages will complicate the translation between them. That is the reason why European translators of classical Chinese poetry have had to use two or three times as many syllables as the original in order to get in all the original message, and that was also why I had a much easier time of it when I tried to follow the same meters and rhymes in translating Lewis Carroll, because the syllabicity in modern colloquial Chinese is more nearly comparable to that of English.

4. GRAMMATICAL STRUCTURE

The grammatical structure of the languages in translation has been the concern of workers on machine translation even to a greater degree than the problem of vocabulary. The treatment of obligatory items and categories have occupied the attention of all translators,[5] human or other. One often has to choose between overtranslation, as when an inflection is translated as a full word[6] and undertranslation, as when the machine is instructed to "throw out article" or "suppress plural". In common practice one tends to run the risk of overtranslation in order not to lose anything in the original message, though comparability in style will put a limit on that.

One important question with regard to structure is at what level one should set up the equivalences. "Most frequently," as Jakobson says, "translation from one language into another substitutes messages in one language not for separate code-units but for entire messages in some other language.[7] But most of machine translation at the present stage has to be concerned with starting with constructions of a certain type and ending with a similar or a different but regular type in the other language, such as postposed modifying clauses into preposed modifying clauses in a target language which does not allow postposed modifiers. But certain cases of non-correspondence, or at least complicated patterns of correspondence will, at the present stage of the science, have to be left to non-machine translation. I have in mind such cases where one language has one form of structure, say S-V-O, and the other language has a similar structure, for certain instances, but a different structure for other instances, conditioned by non-structural but lexical factors.

A common distinction is often made between literal or word-for-word translation and idiomatic or free translation. But there are more than just two degrees on the scale of literalness and idiomaticity.   If we go below the level of the word, there can also be

[5] For example, Robert E. Longacre, "Items in Context: Their Bearing on Translation Theory", *Language,* 34 (1958), 482-491.
[6] As Roman Jakobson has observed, the meaning of grammatical categories may be expressed by lexical means, if necessary. See *On Translation,* ed. by R. A. Brower (Cambridge, Mass., 1959), 235.
[7] *On Translation,* ed. Brower, 233.

morpheme-by-morpheme translation, while if one tries to translate proverb by prov-
erb, there is often no corresponding internal structure at all. Voegelin developed a
technique of multiple-stage translation in connection with programming electronic
computers.[8] Related to this is Hockett's treatment of intermediate stages of imme-
diate constituents, which one might call IIC.[9] These approaches are useful steps in
clarifying non-machine translation and bringing it closer to machine translation.
Hockett's idea is related to, though not identical with, that of N. D. Andreyev's idea
of intermediary language (IL) in his paper "Linguistic Aspects of Translation" (p.
625 of this volume). The idea of the IL is more ambitious and more intriguing than
that of the IIC, but it is something for the future, while Hockett's IIC can be used
any time now.

   A specially important type of structure is at the level of word formation, either by
way of derivation or by way of compounding. In this area the task of translators of
scientific and journalistic subjects is fairly easy, because, irrespective of whether
complex words are translated by their morpheme components, as for example between
Latin and German, or translated as wholes regardless, the units of translation in
science and politics are for the most part international. In some sense, one might say,
all modern life is of one culture and therefore does not run into difficulties one runs
into for periods and areas remote from that of the target language, where there is
much room for argument for various approaches. Typical cases of this sort are trans-
lations of era names of the Chinese dynasties. The common practice is to transliterate
them, which is to make them comparable as to syllabicity, but is definitely under-
translating, since transliteration is a zero-degree translation, of words which to the
users of the original language do have definite overtones in their constituent morphe-
mes, if not tangible denotata. On the other hand, if one completely translates the
components into full-sounding words, then not only will the syllabicity be increased
three or four times, but much more is said than what a native reader or hearer under-
stands of those words. In problems of this sort even a literary person is faced with a
dilemma, let alone machines.

5. SUBJECT MATTER AND CULTURAL CATEGORIES

Under structure we have noted the ease with which scientific and journalistic terms
can be equated between languages, regardless of their internal structure. That is
because science and current affairs belong on the whole to one contemporary culture.
When dealing with diverse cultures, then the difficulties become serious even for non-
machine translation. We have just noted the dilemma translators face in translating
the dynastic era names. Another interesting case of cultural divergence is that of terms

---

[8]   Charles F. Voegelin, "Multiple Stage Translation", *IJAL, 20* (1954), 271-280.
[9]   Charles F. Hockett, "Translation Via Immediate Constituents", *IJAL,* 20, 313-315.

of address, including kinship terms. What is an everyday short word occurring with high frequency and usable in direct address in one language may have to be equated to a long descriptive phrase, which will make a bad translation from the point of view of most of the other dimensions. You certainly can't very well greet a person with: "Good morning, my female-cousin-on-father's-side-younger-than-myself!"

It would seem that numerals and quantitative notions should be easy to manage by both machine and non-machine, but this so only in the sense that all numbers and quantities can be put in a common code from which it can be translated into the target language. On the other hand, cultural patterns enter into all practical use of numbers and quantities, thus making them as complicated as other disparate cultural items. Some languages have no "dozen" except as a foreign borrowing. The conception of "teenage" is a pure accident of languages which start a special pattern from thirteen on. Units of length, time, money, especially coinage denominations, etc. are also largely cultural. They not only influence the translation of words, they even influence the sizes and prices of things.

Proper names would seem to need no translation and some coding of the phonemics (or the graphemics) of the target language would seem to suffice and is readily translated (or transliterated) mechanically. But even here, especially in the case of names of persons, the translation is sometimes as much of a problem as in the case of terms of address, as one can see by examining the proper names in any bilingual dictionary. National and international committees have been set up to regularize the translation and/or transliteration of proper names and their task is never done.

Another category of cultural material to consider is music. So far as music itself constitutes a presentative rather than representative art, and is not language in the ordinary sense, it would seem that music would not need, nor be capable of translation any more than dance or architecture. But even here there is the same tendency for a person of one cultural background to "read" the pattern of another culture into his own, just as in the case of inexperienced translator from one language into another, or the case of the learner of a foreign language who substitutes for the phonemes of the foreign language with a non-congruent set of phonemes from his own language. A striking example of cross-cultural "translation" of music was from my own experience. Once I heard a piece of Javanese music as consisting of notes *do, re, mi, sol, la,* slightly out of tune to be sure. But I was informed afterwards that it was actually in an equal-tempered pentatonic scale!

6. PRAGMATICS OF TRANSLATION

Finally the pragmatics of translation, or the circumstance of use of the language, is the least amenable to treatment by machine, since here we are considering translation in so far as it is influenced by the situational context. A machine, if enlarged

without limit, could be envisaged to take care of all linguistic context. But to take situational context into account would require that the language had to be lived to be translated, in other words, the question would become: "What would you say under the circumstances?" To a limited extent, one could lexicalize certain typical situations and list the cases under idioms. If the circumstance is that of meeting on the street, the "How goes it?" in German would translate into "Where are you going?" in Chinese. To a praise or compliment, a speaker in one language may say something like "Thank you" but speakers of the target language do not say "Thank you" to a praise but say "No, not at all". One language may record parts of dialogues on stage-direction fashion by saying "laughter" or "sigh", while under the same circumstances a writer in another language may use actual interjections like "Ha ha!" or "Heigh-ho!" for which the original language lacks commonly accepted written forms.[10]

All these problems seem to lead us back again to the matter of literal vs. idiomatic translation. But if I have done anything to justify my going over these already well-known problems of translation, it is to show that there are not only many degrees of literalness and idiomaticity, but also many dimensions in which various degrees of literalness and idiomaticity can be ranged, and, while the initial degrees and the more elementary dimensions can be and are already being handled by machine, much that is interesting and important will remain for some time for translation without machine. Between man and machine, he will have to continue to do as much as he must, though he would like to do as little as he can.

University of California
Berkeley

---

[10] For further examples see discussion by Einar Haugen after the paper by N. D. Andreyev, "Linguistic Aspects of Translation", p. 625 of this volume.