

VICTOR H. YNGVE

## Terminology in the Light of Research on Mechanical Translation

Since the ability to use language with precision is important for many people, anything that improves the exactness with which language can be used is well worth while. The scientist, the sociologist, the lawyer—and the skillful translator of their writings — all rely heavily on language as a tool for precise expression. The efforts to improve terminology by UNESCO, ISO, FIT and others are valuable as a means for developing standard definitions and appropriate terms in various languages. An effort to provide better aids for using existing languages and terminology to the best advantage would be equally valuable. The purpose of this paper is to suggest some of the ways in which this can be done.

Dr. J. E. Holmstrom expressed the ideal<sup>1)</sup> that every technical term should be associated in the minds of all concerned with one definite meaning and one only. If this ideal were reached, technical terms would be tools of precision. From a different point of view <sup>2), 3)</sup>, mechanical translation of these terms would be facilitated, for nothing would be required but a simple mechanical substitution of one word for another. Thousands of technical terms, particularly in chemistry and botany, already conform to the ideal, but other words, carrying several, or many, meanings, are used much more frequently than the unambiguous technical terms even in scientific writing. That writers can attain a fair degree of precision while using words with several meanings is proof that precision does not depend on reaching the goal of one word — one meaning.

To a reader, words with several meanings usually seem unambiguous. He recognizes the intended meaning as if by intuition. In many cases his intuition is based on definite and recognizable clues in the text. If we could find and describe the clues that the reader uses intuitively, we would have a better understanding of precisely how terms with several meanings can be used unambiguously. Workers in mechanical translation have already begun research into the nature of the criteria that resolve ambiguity. Some of the results achieved so far are reported here in the hope that those interested in scientific terminology or the precise use of language in writing or translating may be stimulated to initiate co-operating programs of research.

For the purpose of discussion, the different kinds of clues used by a reader in order to understand a word unambiguously even though it may have several senses can be separated into six partly overlapping types. Clues discussed are those found in certain major European languages, although many will be found in other languages with little modification. This presentation is concerned more with the way in which a reader or translator understands an author's meaning than with the problems of expressing this meaning in a different language.

---

This paper was submitted to *Babel* on August 21, 1956.

<sup>1)</sup> HOLMSTROM, J. E.: How translators can contribute to improving scientific terminology. *Babel*, 1, (1955), 2, p. 73.

<sup>2)</sup> *Mechanical Translation*, a journal published at Mass. Institute of Technology, Cambridge, Massachusetts, U.S.A.

<sup>3)</sup> LOCKE, W. N. and BOOTH, A. D. (editors): *Machine Translation of Languages* — 14 essays, 243 p., 1955 Technology (Cambridge, Mass.) John Wiley & Sons Inc. (New York) jointly with Chapman & Hall Ltd. (London).

The first type of clue is information about what field of knowledge is under discussion. The word "run" has special meanings in naval architecture, law, banking, or baseball and cricket, etc. In order to make field-of-knowledge clues useful in terminological work, two steps would have to be taken. The first is a careful inventory and classification of fields of knowledge for this special purpose. Many dictionaries already use field-of-knowledge markers to distinguish different senses of words. They frequently list several hundred field categories. Better for precise terminological work than such lists would be a classification that takes into account the inclusion of some fields of endeavor within others. For example, there are terms with special meanings in science, and terms with special meanings in physics, a part of science. Probably what is needed is not a tree type of classification like the Universal Decimal Classification, but a multidimensional classification of some sort. It should be developed by using the criteria of simplicity and the ability to separate the meanings of many words.

After we have provided a workable classification of fields of knowledge, we need to provide means for the reader to find out what field of knowledge is discussed in a given paper. An author, wishing to use the classification to achieve precision, could supply a simple statement of the field in which his terminology is to be construed. For mechanical translation of papers in which the field is not stated by the author, some other method of finding the field is needed. One method is to have the operator set the machine for the field corresponding to the source of the article. If the article is in a chemical journal, he might assume that the field is chemistry. Having knowledge of the article's source, he might determine the field from the title, from the abstract, or from a brief inspection of the article itself. There is an additional problem, however: The writer may suddenly switch fields when he reasons by analogy or takes an example from another field. There is also the problem of what to do with articles on the border line between two or more fields.

For mechanical translation a method is needed by which the machine can automatically determine which field is under discussion at any point in the text. This could possibly be done by having the machine look up the words in its dictionary and note the field of those that are listed as having only one field. These words would determine the field under discussion and might indicate the points in the text at which the field changes. Perhaps it would be better to have the machine accumulate information on all the possible fields for each word, then look for common fields. Words in a sentence might pertain to a number of fields, but there might be only one field common to all the words. The common field would be selected by the machine and the appropriate meaning of each word chosen on that basis. When one sentence uses terminology from two or more fields, the meaning of a noun is often expressed by its modifier or qualifier. Thus we have mechanical power and political power. It might be necessary therefore, to take into consideration the relation of the various words in the sentence.

As with all terminological work, there would be the problem of keeping the system up to date in the face of changing languages and changing technology. Not only would the meanings of the terms and their assignment to fields have to be checked periodically, but the field classification itself would have to be changed periodically as new fields open up and old ones die out. The careful and systematic application of a good field classification to scientific terminology, preferably on an international scale, would have the advantage that many technical terms could each be assigned several agreed-upon meanings (they tend to accumulate several meanings anyway) and they could still be used precisely and unambiguously.

The second type of clue that is used unconsciously by the reader in resolving the ambiguity of words of more than one meaning is the recognition of a word complex or idiom. Many languages like English form what may be called uncompounded compound words — here called word com-

plexes. An example is “red lead” and “white lead” as opposed to “red paint” and “white paint”. The difference between red paint and white paint is a matter of color and there can be paints of any color, but the difference between red lead and white lead involves chemical composition. “Red lead” consists of two words but in reality it constitutes one technical term. The meaning of a word and its qualifier or modifier is indicated by the separate words in a regular manner, whereas many semantic relations can be indicated by the complex. A pocket lighter is used for lighting cigarettes, not pockets; and a burning glass is not a glass that is burning.

Then there are borderline cases. Mechanical power and political power seem to be of the qualifier type and might be handled best by the use of field indicators in the dictionary. On the other hand, atomic power, private power, public power, and water power might be more advantageously considered as single complexes and listed in full in the dictionary, as many dictionaries already do. Most idioms can also be listed in full in the dictionary.

The third type of clue that the reader uses in understanding multiple-meaning words can be referred to as a syntactic category or syntactic function clue. One of the most important devices in language consists of the indication of a meaning or a distinction by an arrangement of forms — words, parts of words, or groups of words — into standard pattern or structures. We have the patterns involving an article and a noun that distinguish in German between “der Band” and “das Band”, and between “der Mann” and “der Frau”. We have cases determined by word order in “Die Männer sahen die Knaben”, and in English “the man saw the boy”. We have the difference between “to go” and “to him”. We have differences in meaning between a verb used with an object and without it: “When the cook boils the water, it is time to come”, and “When the water boils, it is time to come”. Then there is the difference between a noun and a verb, signalled not by the word itself, but by patterns of the surrounding words: “The cook has boils.”

The standard or traditional grammatical categories are approximately the right ones for the purpose in mind, but they lack good operational definitions. Definitions are needed that specify the exact circumstances under which a word will be a noun or a verb. Particularly in English, but in other languages too, there are many cases in which words of identical form can be used as totally different parts of speech. Eventually, the notion of grammatical or syntactic category could be extended to include several kinds of nouns, thus making a grammatical distinction between “When the water boils, it’s time to come”, and “When the cook boils, it’s time to run”. To achieve a sufficiently detailed syntactic understanding of at least two languages is perhaps the greatest hope of those working on mechanical translation. Syntactic category is the key to the solution of some of the most serious and most frequently occurring multiple-meaning problems. Prepositions, for example, have many uses under different circumstances of sentence pattern, exhibiting many different meanings and, consequently, many different translations.

For technical terminology, an extensive and carefully defined list of syntactic categories could be used to distinguish alternative meanings of technical terms in much the same way in which some dictionaries use the traditional categories. Under “sound”, one dictionary lists adjective, noun, transitive verb and intransitive verb, but there are over 30 meanings. A more detailed syntactic classification would help to separate these.

The fourth type of clue that readers use can be referred to as a meaning category clue. Some of the adjectival meanings of “sound” include flawless (a sound casting), healthy (sound teeth), strong or safe (a sound bridge), stable, secure or trustworthy (a sound corporation), true or logical (a sound argument), sensible or reasonable (sound advice), legal (a sound title), firm or solid (a sound foundation), thorough (a sound beating), and profound or deep (a sound sleep). These are examples of what

may be the most difficult kind of multiple-meaning problem faced by mechanical-translation or terminology workers. They are on the borderline between word complexes (sound sleep and sound beating should probably be handled as complexes), grammatical categories and field-of-knowledge categories. Probably the best way to handle problems of this type is to use meaning categories. Noun meanings could be classified into categories such as inanimate object, animate object, state of being, action, entity, organization, etc. The meaning of an adjective would then depend on the noun category with which it is used. Techniques similar to the field-of-knowledge indicators would be used. The difference here is that the words must be related syntactically in certain specified ways. Meaning categories would be helpful not only in the adjective-noun relation, but also in other relations like subject-verb, verb-object, etc. Meaning categories are similar to syntactic categories in that they would be set up in the same way — on the basis of the distribution of the words. They are tied to sentence structure much more closely than field-of-knowledge categories.

The fifth type of clue used by the reader involves the determination of the antecedent. In translating, it is frequently necessary to locate the antecedent of a pronoun in order to assign it to the proper gender in the second language. The grammatical notion of antecedent is one of the most difficult to define operationally. It is important to codify the rules for locating antecedents not only for the pronouns, but also for the technical terms. It often happens that in a paper, the same concept is called by many different names. In a recent article<sup>4)</sup> the same concept was referred to by the following expressions: *Groups with a future; small discussion groups, consisting largely of friends, and meeting in private homes, to talk over a series of radio programs developed by . . . ; home discussion groups; mutual attraction groups; natural friendship groups; groups; friendship groups; the former; these groups; they; them; groups that are significant both in the functional life of the community and personally significant to their members; and discussion groups.* If all of these references to the same concept by a different name could be recognized by the application of rules, many of the multiple-meaning problems of the several terminological variations could be dealt with.

An understanding of antecedents adequate for handling situations like the above will probably be the last requirement of practical mechanical translation to be met: A knowledge of the other four types of clues may be needed in order to find antecedents. In precise writing, the problem is handled very neatly in certain fields like physics. Concepts are defined in the beginning of a paper and assigned names or symbols which can then be used unambiguously throughout the paper. This technique deserves increased use in fields like law where precision is at a premium.

Some of the clues to the meaning of a passage containing multiple-meaning words have been grouped under five headings: field of knowledge to which the material refers; complexes and idioms; syntactic or functional categories; meaning categories; and determination of the antecedent. Having all of these possibilities, it is to be expected that we will run into multiple or contradictory clues to the meaning of a word. It may be found that the clues operate in some sort of hierarchy or in some order of priority. Syntactic category might take precedence over meaning category, and this over something else. The situation will become clear only when a deeper understanding of the devices of language is developed.

The sixth and last category of clue available to the reader might be called common sense, or perhaps better, specific knowledge of the subject under discussion. This is perhaps the last recourse that the reader has if he has trouble in understanding a passage. He can expand a bare statement, filling in the details from his own knowledge of situations similar to those being discussed, and then deduce

---

<sup>4)</sup> JOHNSON, Eugene J: Groups with a future. *The Journal of Communication* V. (1955); pp. 89 to 101.

the probable meaning of an ambiguous passage. A technical translation under such circumstances comes perilously close to an original composition on a theme suggested by the document. There is the risk of reading into the words of the author meanings that he never intended but that originated, instead, in the mind of the reader or translator. Perhaps the greatest skill that a translator needs is to be able to remain true to the author in such circumstances. Particularly in translating literary works of art, a great deal of study and care is required to make sure that the author's meaning is left intact. A translation of this kind requires much more than carefully defined mechanical routines. It requires that the translator be as competent an artist as the original writer. This type of translation must always be left to the human being. The most that the machine can be expected to do is dull and repetitious mechanical work.