[*Mechanical Translation*, vol.4, no.3, December 1957; pp. 52-53]

Some New Terminology

Erwin Reifler, University of Washington, Seattle, Washington

MT research requires cooperation between engineers and linguists. It is important, therefore, to develop a uniform linguistic terminology that can be understood and used by engineers. Furthermore, it is necessary that linguists develop an understanding of the engineering problems involved. The results of cooperation between linguists and engineers working with the MT Pilot Model at the University of Washington are presented here.

THE LINGUIST interested in pioneering in MT has to struggle with two difficult problems from the very outset: 1) the formulation of an adequate linguistic terminology that can be understood and used by the engineer, and 2) an understanding of the engineering problems involved. During our eight years of MT research at the University of Washington we have had the great advantage of close cooperation between linguists and engineers. I wish to submit for discussion under the heading of "Terminology" some of the results of this cooperation.

Recent developments in MT research at the University of Washington have necessitated the redefinition of some old linguistic terms and the formulation of some new ones. They concern the concepts of MT symbols, i.e., all graphic symbols used in the machine translation process. These MT symbols consist of the Control Symbols and Contextual Symbols.

1. Control Symbols — MT symbols which, coded into the machine memory, control certain steps in the translation process. Since they are not contextual symbols, they appear neither in the input nor in the output.

2. Contextual Symbols — the minimal contextual constituents used to produce a material stimulus for a machine-operational step relevant for MT, such as an alphabetic letter, a numerical figure, a dollar sign, a punctuation mark, a single space. Contextual symbols consist of Input Symbols and Output Symbols.

3. Input Symbols include all contextual symbols that may appear in a source text.

4. Output Symbols include:

a) Letter symbols of the target alphabet

b) Symbols for the numerals

c) Punctuation symbols

d) Editing symbols — target symbols intended to aid in the interpretation of the MT product. Examples are subscript numbers which are attached to some target equivalents to pinpoint the field or fields of science to which the scientific meanings of certain semantic units of the source language belong. (The term "semantic unit" will be explained below.)

5. Free Symbol — a contextual symbol preceded and followed by space. It is always meaningful and always used to symbolize both grammatical and non-grammatical meaning. An example is English 'I'.

6. Bound Symbol — a contextual symbol either not preceded or not followed, or neither preceded nor followed by space. We distinguish

a) Left-bound symbols

b) Right-bound symbols

c) Twice-bound symbols

7. Meaningful Bound Symbol — a contextual symbol used to symbolize:

a) Grammatical meaning, i.e., left-bound "s" in "father's, fathers", the right-bound " ' " in " 's" which indicates that the following "s" is a substantive ending, the twice-bound "o" in "arterio-sclerosis."

b) Non-grammatical meaning, i.e.., the left-bound "g" which distinguishes the meaning of "pang" from that of "pan", the right-bound "s" which distinguishes the meaning of "span" from that of "pan", the twice-bound "a" distinguishing the meaning of "seat" from that of "set."

c) Both grammatical and non-grammatical meaning, i.e., right-bound "o" distinguishing the grammatical and non-grammatical meaning of описать 'describe' (perfective aspect) from that of писать 'write' (imperfective aspect), left-bound "я" distinguishing the grammatical and non-grammatical meaning of ломя 'breaking' from that of лом 'crowbar', twice-bound "ж" distinguishing the grammatical and non-grammatical meaning of между 'between' from that of меду ' of the honey'.

8. Meaningless Bound Symbol — a bound symbol not intended by the author of a source text to symbolize anything, but treated as a separate entry by the MT planners in order to overcome engineering difficulties due to certain limitations of the MT equipment. An English example is the arbitrary left-bound final symbol "n" in "misinterpretation" which consists of 17 letters. If, for example, the input equipment cannot handle free symbol sequences longer than 16 letters, then "misinterpretation" may be split arbitarily into two constituents, the first of which contains the first 16 letters while the second consists of only one letter. These two constituents would then form two separate entries in the machine memory.

9). Symbol Sequence — a sequence of contextual symbols not interrupted by space.

10. Free Symbol Sequence — a symbol sequence preceded and followed by space. A free symbol sequence is always meaningful and is always used to symbolize both grammatical and non-grammatical meaning.

11. Bound Symbol Sequence — a symbol sequence either not preceded, or not followed, or neither preceded nor followed, by space. We distinguish:

a) Left-bound symbol sequence
b) Right-bound symbol sequence
c) Twice-bound symbol sequence

12. Meaningful Bound Symbol Sequence — a bound symbol sequence used to symbolize:

a) Grammatical meaning, i.e., left-bound "ren" in "children", and right-bound "be" in "befall" which changes the intransitive meaning of "to fall" into a transitive meaning, twice-bound ыв distinguishing the grammatical meaning of описывать 'to describe' (imperfective aspect) from that of описать 'to describe' (perfective aspect).

b) Non-grammatical meaning, i.e., left-bound "et" distinguishing the meaning of "ballet" from that of "ball", right-bound "bl" distinguish ing the meaning of "bleat" from that of "eat", twice-bound "ur" distinguishing the meaning of "gourd" from that of "god".

c) Both grammatical and non-grammatical meaning, i.e., left-bound "shore" in "seashore", right-bound "sea" in "seashore", and twice-bound "en" in "disentomb".

13. Meaningless Bound Symbol Sequence — a bound sequence not intended by the author of a source text to symbolize anything, but treated as an individual entry by the MT planners in order to overcome engineering difficulties due to certain limitations of the MT equipment. An English example is the meaningless left-bound symbol sequence "ss" in "irreconcilableness" which consists of 18 letters. The MT planners would have to split this free symbol sequence into two arbitrary constituents containing 16 and 2 letters respectively, and enter them as separate entries into the machine memory if the available input equipment cannot handle free symbol sequences longer than 16 letters.

14. Group of Free Symbol Sequences — a complete text or any part of a text, chapter, section, sentence or clause consisting of two or more free symbol sequences which symbolize a meaning intended by the author of the source text.

15. A Semantic Unit — a single free or bound meaningful symbol or symbol sequence, and any group of free symbol sequences which is idiomatic in terms of source-target semantics.

With the growth of MT development and the increase in the number of MT pioneers it is becoming more and more important to achieve some uniformity in linguistic terminology for MT. I submit the above definitions for criticism and suggestions.