[*Mechanical Translation*, vol.3, no.2, November 1956; pp. 36-37]

Report on Research

Cambridge Language Research Unit, Cambridge, England

THE CAMBRIDGE Language Research Unit is primarily concerned with analytic investigation of language, and in particular with a correlative study of the descriptive-linguistic, logical, algebraic and other notational characteristics of natural languages and of translation between natural languages. Much of this work is relevant to machine translation and the following four sections by members of the unit illustrate some of the applications that are being made. The first three are concerned with the possibilities of using a mechanical thesaurus; the fourth deals with mechanical translation via an interlingua.

Potentialities of a Mechanical Thesaurus

(M. Masterman)

The unit of a mechanical dictionary is the semantically significant "chunk", not the free word. From a logical point of view, uncoded MT dictionary entries form "trees", the paths of which can be determined, to a significant extent, by objective criteria. However, these, as they stand, are too complicated to be used directly for MT.

Attempts to construct multilingual MT dictionary entries show that the entries form, not trees, but algebraic lattices, with translation points at the meets of the sublattices. It also emerges that the complexity of the entries need not increase greatly with the number of languages, since translation points can, and do, fall on one another.

Such a multilingual MT dictionary is analogous, in various respects, to a thesaurus. A method of using such a thesaurus to refine the mechanical pidgin output of a bilingual mechanical dictionary has been devised.

Mechanical Translation Program Utilizing an Interlingual Thesaurus

(A.F. Parker-Rhodes)

The problem of setting the information contained in a fully general interlingual thesaurus into coded form for the use of an electronic computer would be formidable if not impossible of practicable solution, if it were necessary to include every entry as such. But if we could devise a system of coding such that each entry is represented by a numeral which would be calculated from information to hand respecting the current context-situation and the code-number given for the input word under consideration, then we could dispense with any actual thesaurus entries in the computer's storage, all the relevant information being contained in the input and output dictionaries which respectively provide the code-number of the input words and decode the numbers, calculated from these and the context, into the target language.

Mathematically, the problem is completely soluble, provided no limit is placed on the length of numerical symbols. If we limit ourselves to a practicable length of symbol, the question of adapting the general mathematical solution to actual use becomes one of ingenuity which can probably be solved, but which can only be assessed by practical effort. The mathematical procedure consists in finding a set of Boolean operations having certain prescribed properties which can be deduced from the conditions of the problem. These operations are few in number and could be built into a computer; being Boolean, they can be performed with very great speed.

A model solution, substantially simpler than that recommended for actual trial, will be described, and an example worked in it will be demonstrated. This will show up the sort of crossword-puzzle ingenuity required to devise a suitable context classification. The attraction of the method, despite this inevitable lack of elegance, is that it makes the computer actually calculate instead of merely looking things up in lists, and thus makes the whole procedure capable of sufficient speed to be feasible for a mechanical-translation program.

Linguistic Basis of the Thesaurus-Type Mechanical Dictionary and its Application to English-Preposition Classification.

(M.A.K. Halliday)

The thesaurus method of mechanical lexicography is an attempt to systematize the lexis in such a way that the "one-to-one word equivalence" principle can be maintained as the first stage in the dictionary, since the mechanical application of the concept of "primary

meaning" implicit in this principle requires the arrangement of secondary translation equivalents into contextually determined systems. Each entry, consisting of a "key word" and its associates, constitutes one such system.

Multiple translation equivalence requires the specification of the conditions under which one of the terms in the closed system of a thesaurus entry is to be selected, these conditions being contextual features of the target language. This is illustrated by a "context-continuum" showing some word equivalence in non-technical railway terminology in four languages.

The thesaurus exploits the redundancy of the target language by handling its word classes without comparative identification. The autonomous treatment of the target language reduces the loss of determination involved in the translation process.

Among the word classes established for English as a target language, prepositions are particularly suited by their relatively low entropy to non-comparative treatment. Prepositions are classified as "determined" and "commutative". The former are listed as sub-entries of the determining word, having a single or multiple sub-entry according as they are wholly or partially determined. The latter constitute separate headings and are placed in closed commutation systems which differ from those set up for e.g. nouns in that they are in the first instance grammatically, not contextually, restricted.

General Program for Mechanical Translation between Any Two Languages via an Algebraic Interlingua. (R.H. Richens)

It has become clear that the amount of lexical and syntactical analysis required to produce a smooth and idiomatic mechanical translation from any base language into any target language is very great. It is interesting, therefore, to examine the possibilities of mechanical translation via a notational interlingua. With this approach, only one program is envisaged for translation between any two languages, with the addition of specific mechanical dictionaries for each input and output language.

The notational interlingua being studied is ideographic and constructed so as to represent the ideas of any base passage divested of all lexical and syntactical peculiarities; for which reason it is called Nude. The words in Nude are constructed of some fifty elements (Roman letters, capitals and lower case letters being regarded as different symbols), each of which denotes some basic idea such as plurality. animal, or negation. A word in Nude may consist of one letter only; the more complex a notion, the more letters are required. Each word in Nude is regarded as a relation, either 0-ad, 1-ad, or 2-ad; 1-ads are preceded by a point, 2-ads by a colon. Punctuation in Nude is used to indicate the concatenation of the words. The words linked by a 2-ad relation precede it and are separated by a comma, e.g., A, B:C; coordinate conjunction is expressed by a hyphen, e.g., A-B.C.

The translation program involves the following operations:

(1) Matching semantically significant "chunks" of the base passage against the Base-Nude dictionary.
(2) Reorganization of the syntax into Nude syntax by the method of cyclical reduction described at the 1955 Symposium of the Cambridge Language Research Unit, utilizing the word-class sequence entries of the Base-Nude dictionary (cf. MT III, 1).
(3) Treatment of chunk-chunk, chunk conjugation and chunk-semantic interactions by comparison with the appropriate interaction entries in the Base-Nude dictionary.
(4) Repetition of the above stages, using the Nude-Target mechanical dictionary. The potentialities of this method are to be illustrated by translation from a Japanese passage into English, German, Latin and Welsh.

List of Publications of the C.L.R.U.

1. Progress Report I (January, 1953), obtainable cyclostyled from C. L. R. U., 20 Millington Road, Cambridge, England.
2. Progress Report II, MT Vol. 3, No.l.
3. Annexe V to same is obtainable cyclostyled from Editors (MT).
4. The Potentialities of a Mechanical Thesaurus by Margaret Masterman.
5. A General Program for Mechanical Translation between any Two Languages via an Algebraic Interlingua, by R.H. Richens.
6. The Linguistic Basis of the Thesaurus-Type Mechanical Dictionary and its Application to English Preposition Classification, by M. A. K. Halliday.
7. An Algebraic Thesaurus, by A.F. Parker-Rhodes.