

A PROJECTED STUDY OF SEMANTIC AMBIGUITY

David G. Hays

P-944A

September 24, 1956

The RAND *Corporation*
1700 MAIN ST. • SANTA MONICA • CALIFORNIA •

A PROJECTED STUDY OF SEMANTIC AMBIGUITY

The only approach yet conceived for reduction of semantic ambiguity is by examination of context. One of many possible translations for a word must be selected by consideration of adjacent words, or of the specialized nature of the document in which it appears, etc.

The original plan of Roget's Thesaurus seems adaptable to this problem. Roget devised a hierarchical classification of meanings--major classes are divided into sub-classes, which in turn are subdivided, and so on, until a large number of categories is obtained. Within each category are placed a set of words capable of expressing a common meaning.

If a suitable classification could be constructed--certainly Roget's is not suitable--it could serve as an aid to M.T. in the following way. A sentence for translation is presented. One word is ambiguous--that is, it has two or more locations in the classification. Each of these locations is compared with the locations of the other words in the given sentence. That location for the ambiguous word is chosen which is closest to the locations of the other words. In any fixed location, a word has only one translation, or several which are equivalent.

Many problems inhere in this approach. A classification scheme and a distance function must be defined so that "acceptable" translations are obtained. Only intuitive arguments have been found for the existence of such a scheme and function: the best argument would be an empirical demonstration.

If the desired classification is possible, it is probably not possible on the basis of pure synonymity. The intended use suggests that the "distance" function is inversely related to frequency of joint occurrence.

A program is now beginning at The RAND Corporation to examine the feasibility of statistical methods for design of the classification scheme. The plan is to devise a simple machine routine and test it with a small sample of language, continuing by successive elaborations of method until the ideas involved can be evaluated.