

BOOK REVIEWS

SYNONYMY AND SEMANTIC CLASSIFICATION

Karen Sparck Jones

(University of Cambridge, England)
(Edinburgh Information Technology Series 1)

Edinburgh University Press: Scotland, 1986,
viii+285 pp.
(Distributed in the U.S. by Columbia University
Press)

ISBN 0-85224-517-3, £25.00 (hb)

Reviewed by
Svatava Machová
Charles University

The focus of the work is on providing a principled account of lexical relations and classes, defining semantic primitives in terms of the textual behaviour of words and hence viewing them as embedded in their natural language and not as elements of another, independent language. (p. 1)

The book *Synonymy and Semantic Classification* by Karen Sparck Jones was published in 1986 as the first volume of Edinburgh Information Technology Series on computer science and artificial intelligence. The book is interesting both in its content and in some circumstances of its publication. The main part of the book comprises Sparck Jones's Ph.D. thesis, approved at the University of Cambridge as long ago as 1964. No changes in the text of the thesis have been made for this publication. All that has been added is a new chapter entitled "Twenty years later: A review", in which Sparck Jones acts as her own reviewer and considers her own work from the viewpoint of the results achieved in linguistics during the past 20 years.

The book consists of seven chapters and two appendices. It reflects the complex paths along which linguists have to proceed when searching for the semantic structure of a vocabulary, for semantic primitives, and for the establishment of the parts they play in the automatic process of natural language text understanding.

To realize the idea of machine translation of natural language texts, it is necessary, among other things, to have a dictionary that makes it evident what meanings each word-sign possesses, and to have a procedure for recognizing the meaning in which the word-sign is used in the context. For designing this kind of procedure, the author considers it important that a discourse is connected by one idea that recurs several times in the discourse. It should be discovered what kind of information is required to detect this idea. The author asks the question as to whether the existing thesauri can

fulfill these requirements, and arrives at the conclusion that they cannot. She suggests a different approach: group all the elements of a vocabulary on the basis of the interrelations the words have with each other and define the meanings of words with the aid of these interrelations.

The relations between words that a linguist stores in the dictionary are supposed to be semantic relations (i.e., relations between words), not factual relations (i.e., extra-linguistic references of words). After studying several types of relations (e.g., hyponymy, contrast, likeness, implication), the author arrives at the conclusion that synonymy as the exemplar of likeness of meaning is a semantic relation because it is dependent on the way in which the mechanism of using signs is functioning. It is therefore possible, on the basis of synonymy, to define the idea of the semantic structure of a vocabulary. The author gives analyses of several existing definitions of synonymy; however, she does not find them quite suitable for designing machine translation dictionaries and suggests a definition of her own, which is linked with the terms *sentence*, *length of the sentence*, *ploy* (a primitive notion, representing roughly the meaning, application, and form of a sentence), *row* (a set of elements that can replace one another with respect to a position n in a sentence s_n without changing its ploy), *word-sign*, *word-use*, and *word*: "Two or more word-uses¹ are synonymous, and therefore form a row, if their word-signs are mutually replaceable at some position in some sentence, without any change in the ploy of the sentence" (p. 71). To each word-use it is possible to assign one and only one row consisting of all the word-signs mutually replaceable in a context, hence synonymous. Let us take, for example, word-signs A, B, C, D, E, F , and G , and let us consider word uses represented by the word-sign A . We may find, for instance, that A has three word uses characterized by three rows:

A B C
A D
A E F G

This notation means that the first word-use of word-sign A can be replaced by the word-signs B and C in a certain context; the second word-use can be replaced by the word-sign D , and in the third word-use, it can be replaced by the word-signs E, F , and G . If the word-sign A acquires a new word-use, a new row is included in its set of rows. Thus, if we operate with the concept of row we can define precisely a total synonymy of two word-uses, a likeness of two word-uses, a similarity of two rows and a connectivity of two word-signs.

This definition of synonymy provides us with a tool to relate the words in the vocabulary of a language to one another, and, hence, a tool with the aid of which we can define all the word-uses of each word-sign simply and in a form allowing comparisons of definitions. It can be objected that if this approach is applied to special vocabulary (technical and scientific terminology) in which synonymy is rare, the definitions of word-uses will often remain empty. The author analyzes various potential classifications of sets of synonyms (sets of rows) enabled by the synonymy definition, as well as the possibilities of their groupings. She finds a few ways of row grouping resulting in conceptual groupings similar to those found in thesauri. She made a computer experiment whose aim was to find out whether the suggested procedure of grouping is practicable for the natural language vocabulary. She selected 21 word-signs and, on the basis of information about them contained in the *Oxford English Dictionary*, assigned 500 rows to them (given in her Appendix 2). With aid of the computer, she tried to establish groupings similar to those in thesauri. The results have been satisfactory in part only. After modifying classification criteria, the author intended to carry out a new experiment involving about 2,000 rows. The results of that experiment were not mentioned in the chapter "Twenty years later".

Sparck Jones is looking for an answer to the question as to why we expect to find synonyms in natural languages. She arrives at the conclusion that synonymy is not a mere redundancy and that it exists because, in the extra-linguistic world, we encounter situations that are unique, but, at the same time, similar to each other in certain aspects. Synonymy reflects this fact; otherwise language would be an inadequate representation of the extra-linguistic world. She gives four models of the way linguistic symbols are set up and she claims that her Model 4 is the one that represents natural language:

Model 4:

- a. A word-use may have more than one sign;
- b. Two or more word-uses may have the same sign, where these word-uses are similar. (p. 135)

It remains unclear why also homonymy, which is included in her Model 2 and which does constitute a language relation, is not dealt with in Model 4, and why the word *homonymy* is never used in the book.

The book exemplifies an excellent way the efforts made to solve practical problems in computational linguistics bring new and promising knowledge in a field of linguistic theory. The rendering of the subject has a solid, logical structure; it is clear and systematic. The text is not burdened with the artificial linguistic terminology that flooded the linguistic publications of the sixties. It can be regretted that the book was not published at the time of its origin. Even for the present time, however, it yields a number of suggestions for linguistic research.

NOTE

1. The author states that "word-meaning" and "word-use" are to be regarded as synonyms and makes use of "word-use" throughout her book. I will do the same in this review.

Svatava Machová received her Ph.D. in computational linguistics from Charles University, Prague, in 1970; she is presently a senior scientific worker in the university's Computer Center. Her research has included lexical components of question-answering systems and multilingual terminological databanks. Her address is: Matematicko-fyzikální fakultě, Univerzity Karlovy, Malostranské n. 25, 118 00 Praha, Czechoslovakia.

REASONING AND DISCOURSE PROCESSES

Terry Myers; Keith Brown; and Brendan McGonigle (eds.)

(Respectively: University of Edinburgh, University of Essex, and University of Edinburgh)
(Cognitive Science Series)

Academic Press: London, 1986, xii+295 pp.
Hardbound, ISBN 0-12-512320-5, \$69.00 (hb); ISBN 0-12-512321-3, \$29.95 (sb)

Reviewed by
Robin Cohen
University of Waterloo

Reasoning and Discourse Processes is a collection of papers in the Cognitive Science series of Academic Press. The authors are largely from psychology and linguistics departments in the U.K. and Europe.

The book begins with a brief preface, which clearly sets out the theme of the book: exploring the relation between verbal reasoning and discourse, with an aim of contributing to an adequate theory of natural language processing. The very first chapter of the book, written by the editors, serves to describe how the different authors address the theme mentioned in the preface. This chapter is an excellent summary of the upcoming chapters, providing the reader with an index into those parts of the book of most interest to his/her own research.

The editors divide the papers into two main topics: forms of representation and the role of inference for reasoning within discourse. The first six chapters discuss representation: "whether rules of inference formalized in a logical calculus adequately characterize the deductive component of the verbal reasoning capacity". The remaining chapters are primarily concerned with characterizations of coherence; these considerations may introduce a deductive component into discourse.

The book thus addresses issues of concern to computational linguists. Constructing models for the processing of natural language requires considerations both of the form of the representation and the inferencing