Feasible Fuzzy Semantics*

Burghard Rieger**

Abstract

Linguistic semanticists who think their discipline an empirical science, will not mainly be concerned with either language theory, formal logics or mathematics, but with the study of meaning as it is constituted in spoken or written texts used in the process of communication. Rather than focussing on the fiction of an 'ideal' speaker or the formal rules of an abstract and mere theoretical language usage, their linguistic point of view implies that they are much more interested in the *analysis* and *description* of natural language regularities that real speakers/hearers follow and/or establish when they interact verbally by means of texts in order to communicate.

For any *description* of natural language meaning, however, we are in need of a formally adequate representation to depict semantic phenomena, and for any anlysis of natural language meaning we need methods and procedures which are empirically adequate. Both, the postulates of *formal* and *empirical* adequacy will have to be met by a theory of communicative semantics that – other than word – or sentence-semantics – is comprehensive and satisfactory. Such a theory does not exist. But it seems that the concept of 'fuzzy' sets may prove to serve as a formally and numerically flexible link to connect the two main, seemingly divergent lines of research in modern semantics so far: namely; the more it theoretically oriented *algebraic models* of what logicians feel an 'ideal' speaker should, or would, do when he produces meaningful sentences and the more *empirically* oriented methods and *quantitative procedures* of experimental semanticists who try to find out what real speakers actually do when for communicative purposes they produce texts and/or try to understand them. As 'fuzzy' theory introduced by Zadeh (1965) has in the meantime been developed to an increasingly successful formal approach of even wider scope than semantics, it seems fit to bridge the gap between an abstract model of, and its application to, vagueness of natural language meaning.

It is assumed that the structural meaning of lexical item (word, lexeme, stem, etc.) – as distinct from its referential meaning – may be computed from sets of natural language texts according to the use the speakers/hearers make of an item when producing utterances in order to interact verbally. When spoken or written by real communicators in sufficiently similar pragmatical situations of actually performed or at least intended communication called *frames*, these natural language utterances, sentences, strings, texts, etc. will be considered 'pragmatically homogeneous'. The hypothesis now is that in a sufficiently large sample of pragmatically homogeneous texts, called *corpus*, only a restricted vocabulary, i.e. a limited number of lexical items will be used by the communicators however comprehensive their performative vocabularies in general might be. The

*in: Heggstad, K. (ed.): COLING 78 – Reprints 7th International Conference on Computational Linguistics, Bergen (ICCL) 1978, pp. 41-43

**Institut für mathematisch-empirische Systemforschung (MESY); Heerlener Straße 30, D-5100 Aachen (West Germany)

lexical items employed in those texts will be distributed over the corpus according to their communicative properties, constituting certain regularities which may be detected empirically by means of a modified correlation-measure, allowing to compute the relational interdependency of any two lexical items from their textual frequencies (Sparck Jones/Kay 1973). Those items which frequently occur (or do not occur) t o g e t h e r in a number of texts will positively be correlated and hence called 'affined', those of which only one (but not the other) frequently occurs in a number of texts will negatively be correlated and hence called 'repugnant'. For a limited number of lexical items in a corpus restricted pragmatically be the frame of communication, it has been shown elsewhere (Rieger 1976a; b; 1977a; b) that degrees of word-repugnancy and word-affinity – indicated by numeric values ranging from -1 to +1 and ascertained without recurring to an investigator's or his proband's knowledge of language (competence), but solely from the textual regularities (performance) observed – may be used to depict usage regularities within an continuum, called *corpus-space*, the elements of which are called *corpus-points*, each representing a lexical items's usage regularity. Differences in usage, resulting in different positions occupied by the corpus-points concerned within the corpus-space, may be calculated by a distance-measure, the values of which are real, non-negative numbers. These may be interpreted in two ways: (1) the distances between any one corpus-point and all the others are considered as coordinates which will define a so-called *meaning-point* as an element of a new space, called *semantic-space*; (2) the distances between any one corpus-point and all the others are interpreted as membership-grades which allow the differences of a lexical item's usage regularities to be employed in order to represent its meaning by a fuzzy subset of the vocabulary. Both these interpretations converge, however, on the notion of the meaning of meaning being a frame- or corpus-dependent *function* which maps a lexical item onto the vocabulary by means of all the differences of all the regularities observed according to the usage the lexical item is made of by the speakers/writers within the corpus of that particular frame. The lexical structure thus constituted and empirically detected my formally be modelled as a system, i.e. a set of 'fuzzy' sets each of which is a mapping of lexical item's meaning with the vocabulary serving as its (componentially extended) descriptor set.

The theory of fuzzy sets now provides some basic definitions (equality, similarity, containment) and operations (complementation, union, intersection) which in this model of the semantic-space are employed to specify some basic semantic relation like *synonymy, partial synonymy, hyponymy* etc. sentence-independently; operations like *negation, adjunction and conjunction*, allow new lexical meanings (fuzzy sets) to be generated from the finite set of those already determined empirically within the lexical structure.

From computation of a corpus of the 19th and 20th Century German students' poetry (comprising of some 3000 texts and a vocabulary of 300 lemmatized types/96000 tokens) some examples will be given to illustrate that fuzzy sets theory allows vague lexical meanings of words to be represented in precise terms. The vagaries of actual word usage by individuals, social groups, pragmatic frames, or else, do not even need to be reduced on a strict binary determinateness, but rather become the empirical basis for any structural meaning's representation in a system which depicts the (semantic) regularities followed and/or established by real communicators and consequently are computable from (individually, sociologically, pragmatically, or else determined) sets of

natural language texts.

Literature

Rieger, B. (1976a): "Fuzzy Structural Semantics. On a Generative Model of Vague Natural Language Meaning", Paper presented on the Third Meeting on Cybernetics and System Research (EMCSR/3) April 76, Vienna, Austria (to appear).

Rieger, B. (1976b): "Unscharfe Semantik natürlicher Sprache. Zum Problem der Repräsentation und Analyse vager Bedeutung", Paper presented on the Linguistics Symposion, Deutsche Akademie der Naturforscher Leopoldina, Halle: to appear in Nova Acta Leopoldina.

Rieger, B. (1977a): "Theorie der unscharfen Mengen und empirische Textanalyse", Paper presented on the 'Deutsche Germanistentag 1976', Düsseldorf; published in Klein, W. (Ed.): *Methoden der Textanalyse (Medium Literatur 3)*, Heidelberg: Quelle u. Meyer, p. 84-99.

Rieger, B. (1977b): "Vagheit als Problem der linguistischen Semantik", in Bald, W.D. (Ed.): *Semantik und Pragmatik.* Akten des 11. Linguist. Kolloqiums, Tübingen: Niemeyer, p. 91-101.

Sparck Jones, K./Kay, M. (1973): *Linguistics and Information Science*, New York/London: Academic Press, partic. 120-173.

Zadeh, L.A. (1965): "Fuzzy Sets", in: *Information and Control 8*, p. 338-353.