Linguistic Semantics and the Problem of Vagueness.

On analysing and representing word meaning.*

Burghard B. Rieger

One cannot ignore the impression, when going through the indexes of a sufficiently large number of relevant works in linguistics or language theory, that *vagueness* — other than, for instance, *ambiguity* — has in the majority of cases not yet gained the status of a keyword signalling a topic of current interest.

This might seem rather a paradox considering the fact that modern linguistics, and its development of formal theory in particular, has been deeply influenced by the ideas of a number of theoreticans, i.e. logicians, mathematicians, and philosophers, who, for their part, have been well aware of the vagueness of natural languages, and have reflected upon it, characterizing it as the problem of dealing precisely with the phenomena of imprecision (Rieger 1976). As their considerations have quite evidently fallen into oblivion, some of them ought to be mentioned in the following. *Firstly,* I shall be trying to illustrate very briefly the theoretical background of this paper, touching upon both referential and structural semantic theory in linguistics. *Secondly,* I shall sketch the course of my own, more empirical approach in analysing and describing natural language meaning within the frame of a pragmatically-based generative model of structural semantics. *Thirdly,* and finally, I shall give some examples from computational experiments on a corpus of nineteenth century German students' poetry to illustrate the feasibility of the approach (Rieger 1977a).

Theoretical Background

There will probably be no argument among semanticists that — whatever else has to be dealt with in detail — natural language meaning presents two major problems: firstly, what is known as the connotational aspect of how the signs, words, and sentences of a language are related to one another, constituting *structural meaning* as a system of intra-lingual relations; and secondly, what is known as the denotational aspect of how the signs, words, and sentences of a language are related to the objects and/or processes they refer to, constituting *referential meaning* as a system of extra-lingual relations.

---

*Published in: Ager, D.E./Knowles, F.E./Smith, J. (Eds.): Advances in Computer-aided Literary and Linguistic Research, Birmingham (AMLC University of Aston Pr.) 1979, pp. 271–288.

Structural and referential approaches differ in what they consider natural language meaning to be, yet, they converge on the notion of it being a relation between a representation (natural language) and that which it represents (meaning). Why, then, did they indeed notice this very relation's basic fallacy but still fail to topicalize it in a semantic theory of its own?

> Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. ... Vagueness, clearly, is a matter of degree, depending upon the extent of the possible differences between different systems represented by the same representation. Accuracy, on the contrary, is an ideal limit. Russell:1923:85–90

To start with the latter, *referential semantic theory* has developed along the line of Frege, Russell, early Wittgenstein, Carnap, and Quine. Their relevance to linguistics was not recognized until recent years. Linguists' increasing interest in the formal theory of semantics has produced quite a number of different approaches in the field since. These share the fiction, though, that natural language sentences ought to be either 'true' or 'false', or, at worst, have a third value like 'indeterminate'. Like the truth conditions for predicates or propositions, those for natural language sentences are analogously introduced in terms of classical set theory. Accordingly, the meaning of a word is basically identified either extensionally with a set of points of reference, or intensionally with a set of their properties in the universe of discourse, allowing a truth value to be determinably assigned to any (declarative) natural language sentence. These truth-value models of *sentence semantics* now tend to exhibit all the formalisms and idealizing abstractions the logical rigour of precise binary systems calls for. They do so, however, at the price of a rather limited coverage of basic and very obvious characteristics of natural language meaning, one of which, namely *vagueness,* is consequently encountered only in its degenerated form of ambiguity.

> Vague terms are only [!] dubiously applicable to marginal objects, but an ambiguous term such as 'light' may at once clearly be true of various objects (such as dark feathers) and clearly false of them. Quine:1960:192

And as vague words occurring in natural language sentences do not seem to perturb their truth values, the notion of vagueness necessarily never attracted any attention as a problem of sentence semantics unless this was founded on pragmatics.

> When sentences whose truth values hinge on the penumbra of a vague word gain [communicative] importance, they cause pressure for a new verbal convention or changed trend of usage that resolves the vagueness in its relevant portion. Quine:1960:128

Hence, the trouble with vagueness is that it may be permanently experienced in communication but vanishes when so resolved, whereas ambiguity needs logical disambiguation to be experienced at all.

Unlike referential semantics, *structural semantic theory* primarily has been a means of word semantics. As such, it has always considered vagueness somewhat fundamental to natural language meaning, but only so as to dispense with it all the more decisively as a problem which could not be solved in precise terms anyway. Structuralists have therefore been concerned with the question of how the lexical meanings of words — rather than being related to extra-lingual sets of objects — are intra-lingually related to one another, constituting a paradigmatic system of which people obviously make use when communicating. According to structural theory, the meaning ('sense') of each term depends on the position it occupies in that system. It is argued that, although the terms may referentially be vague, the position of each term in the system relative to another will nevertheless be defined with precision. This fiction of 'structural preciseness' as opposed to 'referential vagueness' has inspired linguists since Saussure, Trier, and Weisgerber up to Coseriu, Greimas, and Lyons, and even scholars from non-linguistic disciplines like Osgood, Goodenough, or Wallace — to mention only these few. Their models, methods, and metaphors have undoubtedly been fertile and influential for some time and/or discipline. But as they were based mainly upon intuitive introspection and the questioning of probands, and were abstracted from the pragmatic frame which any real communicant's language usage constitutes, they do not seem to have achieved either the theoretical consistency or the methodological objectivity that an empirical theory of language communication calls for. Thus, apart from the ethno-sciences or experimental psychology, structuralistic ideas have been of decreasing influence in modern linguistics and its recent semantic theories.

Although the problems of vagueness have attracted increasing attention from logicians during recent years (see, for instance, *Synthese,* **30,** 1975 and **33,** 1976), linguists seem to be even less interested. Might it not be true that what was said about logicans at the beginning of this century, now applies only to linguistic semanticists who are 'at fault in giving vagueness the go-by, so far as not even to analyse it'? (Peirce:1965:5,446)

> It should be noted that the semantic analysis of a given natural language poses enormous difficulties because of the great complexity and apparent vagueness of the relevant phenomena. Problems of this kind are relevant not only for the adequate description of particular languages, but also for the development of the general theory, since a general theory is valid only in so far as it is based on empirical facts. Bierwisch:1970:184

As a linguist, who thinks his discipline an empirical science, I shall in the sequel be concerned not so much with either language philosophy, formal logics, or mathematics, but mainly with the study of meaning as it is constituted in spoken or written *texts* used in the process of *communication.* Rather than focusing on the fiction of an 'ideal speaker' or the formal rules of an abstract and merely theoretical language usage, my linguistic point of view implies that I am much more interested in the *analysis* and *description* of natural language regularities which 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 *analysis* 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 is comprehensive and satisfactory. Such a theory does not exist. But I think that the concept of fuzzy sets may prove to serve as a formally and numerically flexible link to connect satisfactorily the two main, so far seemingly divergent lines of research in modern semantics: namely, the more *theoretically*-oriented *algebraic models* of what logicans 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 set theory introduced by Zadeh (1965) has in the meantime been developed into an increasingly successful formal approach of even wider scope than semantics (Zadeh 1975), it seems fit to bridge the gap between an abstract model of, and its application to, vague natural language and the processes it represents:

> Rather than regard human reasoning processes themselves 'approximating' to some more refined and exact logic process that could be carried out perfectly with mathematical precision, Zadeh has suggested that the essence and power of human reasoning is in its capability to grasp and use inexact concepts directly. He argues that the attempts to model, or emulate, it by formal systems of increasing precision will lead to decreasing validity and relevance. Gaines:1976:625

Basic to the notion of fuzzy sets is — other than in classical set theory — that the elements of fuzzy sets show gradual rather than abrupt transition from non- to full membership. Fuzzy sets are defined by characteristic or membership functions which associate with each element a real, non-negative number between 0 and 1, with 0 equalling 'non-membership' and 1 equalling 'full-membership' in the classical set-theoretical sense. Let A be a subset of X, then A can be defined by a membership function

$$\mu_A \colon X \to [0,1]$$

that will map X onto the interval $[0,1]$. Hence, the fuzzy set A is defined to be the set of ordered pairs

$$A := \big\{ \big(x, \mu_A(x)\big) \big\} \text{ for all } x \in X.$$

To give an example, let X be the set of existing (European) automobile types scaled according to increasing engine volume, then the *meaning* of a term like 'middle-class car' may referentially be represented as a fuzzy set, defined by a membership function μ_M that associates with each possible car model $x \in X$ a numeric value $\mu_M(x)$, stating the membership grade of x in the fuzzy sub-set M ('middle-class car') of X, illustrated in Figure 1.

Definitions of equality, containment, complement, union, and intersection may be given which reduce to those of classical set theory when 0 and 1 become the only membership grades admissible.

In his paper on 'Quantitative Fuzzy Semantics' Zadeh (1971) adopted a strictly reference-theoretical model, into which he successfully incorporated the notion of

	Fiat	VW Golf	Opel Rekord	Mercedes	Rolls Royce
X	f	g	o	m	r
ccm	500	1,100	1,700	3,200	6,300
$\mu_m(x)$	0.0	0.5	1.0	0.3	0.0

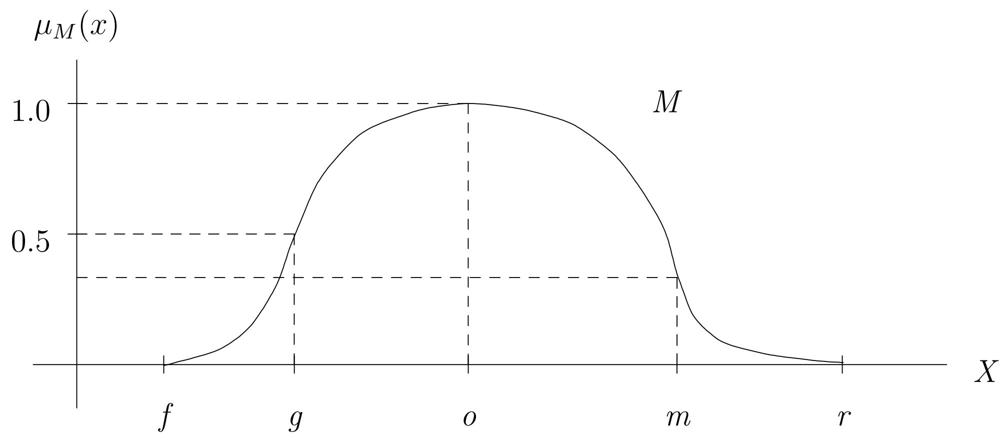

Figure 1: The (referential) *meaning* of 'middle class car', represented (subjectively) by the fuzzy set M, defined on the descriptor set X of existing (European) automobiles, scaled according to increasing engine capacity.

fuzziness. He was able to show that the meaning of a word or term may well be vague in the sense that it refers to a set of reference points whose boundary is not sharply defined, thus constituting a fuzzy set in the universe of discourse.

> In fact it may be argued that in the case of natural languages, most of the words occurring in a sentence are names of fuzzy rather than non-fuzzy sets, with the sentence as a whole constituting a composite name for a fuzzy subset of the universe of discourse.
> Zadeh:1971:160

The second aspect raised by Zadeh in the same paper, whether 'fuzziness of meaning can be treated quantitatively, at least in principle' has, however, been dealt with only formally. The empirical side of it concerning questions of how the meaning of a term described as a fuzzy set may be detected, how a descriptor set may be chosen, or how the membership grades may be ascertained and associated with the elements of the descriptor set in a particular case, has not even been touched upon. We are informed instead that membership functions

> can be defined in a variety of ways: in particular (a) by a formula, (b) by a table, (c) by an algorithm (recursively), and (d) in terms of other membership function (as in a dictionary). Zadeh:1971:161

From the empirical linguist's point of view this is rather unsatisfactory. In clearly recognizing the relevance of fuzzy set theory for the *description* of word meaning, he will also find that — in so far as its *analysis* is concerned — fuzzy set theory does not offer any new method. It seems that it merely allows a somehow quantified notation of more or less subjective, more or less acceptable, results, which traditional methods of linguistic introspection may yield anyway. I would therefore like to propose that fuzzy set *theory* be combined with *methods* of statistical text analysis in order to arrive at a generative model of structural semantics for which the notion of vagueness is constitutive.

Methodological Approach

It is assumed that the structural meaning of any lexical item (word, lexeme, stem, etc.) depends on its pragmatics and hence may be detected from sets of natural language texts according to the use the speakers/writers make of an item when they produce utterances in order to communicate. Such utterances are called 'pragmatically homogeneous' if they were written or spoken by real communicators in sufficiently similar situations of actually performed or at least intended verbal interaction, constituting its *frame.*

It has been shown elsewhere (Rieger 1971, 1972) that in a sufficiently large sample of pragmatically homogeneous texts, called a *corpus,* only a restricted vocabulary, i.e. a limited number of lexical items will be used by the communicators, however comprehensive their personal vocabularies in general might be. Consequently, the lexical items employed in these texts will be distributed according to their communicative properties, constituting *semantic regularities* which may be detected empirically (Sparck Jones and Kay 1973). For this purpose a modified correlation coefficient has been used by way of an experiment.

Given a lemmatized vocabulary

$$V := \{x_i\}, \quad i = 1, \ldots, n$$

and a pragmatically homogeneous corpus of texts

$$C := \{t\}, \quad t = 1, \ldots, T$$

where

$$U = \sum_{t=1}^{T} u_t, \quad 1_t \leq u_t \leq U \tag{1}$$

is the sum of all text lengths u_t measured by the number of words (tokens) in the corpus and

$$H = \sum_{t=1}^{T} h_t, \quad 1_t \leq h_t \leq H \tag{2}$$

is the total frequency of a word x (type) computed over all texts in the corpus, then the coefficient applied reads

$$\alpha(x, x') = \frac{\sum_{t=1}^{T} (h_t - {h_t}^*)(h'_t - {h'_t}^*)}{\left(\sum_{t=1}^{T} (h_t - {h_t}^*)^2 \sum_{t=1}^{T} (h'_t - {h'_t}^*)^2 \right)^{\frac{1}{2}}}; \quad -1 \leq \alpha(x, x') \leq +1 \tag{3}$$

$$\text{where} \quad {h_t}^* = \frac{H}{U} u_t \quad \text{and} \quad {h'_t}^* = \frac{H'}{U} u_t$$

This permits the computation of the relational interdependence of any two lexical items x and x' from their textual frequencies. Those items which co-occur frequently in a number of texts will be positively correlated and hence called ‘affined’, those of which only one (and not the other) frequently occurs in a number of texts will be negatively correlated and hence called ‘repugnant’. Different degrees of word repugnancy and word affinity — indicated by numeric values ranging from -1 to $+1$ — may thus be ascertained without recurring to an investigator’s or his probands’ knowledge of the language (competence), but solely from the regularities observed in a corpus of texts spoken or written by real speakers/writers within a particular frame of actual communication (performance).

Let C be such a corpus consisting of a number of texts t satisfying the conditions of pragmatic homogeneity. For illustrative purposes, we shall consider a simplified case where the vocabulary V employed in these texts is to be restricted to only three word types, say i, j, and k which have a certain overall token frequency. The correlation coefficient α will measure the regularities of usage by the ‘affinities’ and ‘repugnancies’ that may hold between any one lexical item and all the others used in the texts. That will yield for any item an n-tuple of correlation values, in this case for the lexical item i with $n = 3$ the triple of values ii, ij, and ik.

These correlation values are now interpreted as being co-ordinates that will define for each lexical item i, j, or k one point α_i, α_j or α_k in three-dimensional space. This is illustrated in Figure 2. There we have three axes representing the three word types i, j, and k which intersect in front of the three planes cutting the axes at their $+1$ values. The point α_i is defined by the correlation values $ii = +1$, $ij = -.25$, and $ik = -.75$; it is therefore situated in the i plane with the interrupted lines (parallel to the j and k axis) representing the ii and ik values. The other points α_j and α_k are defined analogously.

The position of α_i in this space now obviously depends on the regularities the lexical item i has been used with in the texts of the corpus. α_i therefore is called *corpus point* of i in the α or *corpus space.* Any two points in this space will consequently be the more adjacent to each other, the less their regularities of usage differ. This difference may be calculated now by a distance measure δ between any two α points which reads

$$\delta(y, y') = \left(\sum_{i=1}^{n} \left(\alpha(x, x_i) - \alpha(x', x_i) \right)^2 \right), \quad 0 \leq \delta(y, y') \leq 2\sqrt{n} \tag{4}$$

and is illustrated by dotted lines in Figure 2.

These distance values, which are real, non-negative numbers, do represent a new characteristic which may be interpreted in two ways. *Firstly,* the dotted distances between any one α point and all the others are interpreted as new co-ordinates, then these co-ordinates will again define a point in a new n-dimensional space, called *semantic space.* The position of such a *meaning point* in the semantic space will depend on all the differences (δ or distance values) in all the regularities of usage α or correlation values) any lexical item shows in the texts analysed. *Secondly,* the dotted distances between any one α point and all the others are interpreted as membership grades: then — after these δ values have been transformed appropriately into μ values, ranging from 0 to $+1$ — the differences of a lexical item's usage regularities may well be represented by a fuzzy set with the vocabulary serving as its descriptor set. Both these interpretations of δ values, as co-ordinates of points in the semantic space *or* as membership grades of fuzzy subsets in the vocabulary, are equivalent: they will equally map the vague meaning of a word as a function of *all* its differences in *all* its regularities onto the vocabulary, according to the usage a lexical item possesses, as exhibited by the speakers/writers in a corpus of pragmatically homogeneous texts.

Apart from this, the fuzzy set theoretic interpretation allows a considerable extension of this analytical model of structural meaning. Some basic definitions and formal operations may now be introduced which will allow an empirically based and formally satisfactory, though sentence-independent, explication of linguistic *sense relations* and — even more important than that — the formal *generation* of (at least in principle) infinitely many new meanings from the finite number of those lexical meanings, which prior to that were analysed empirically from the text corpus.

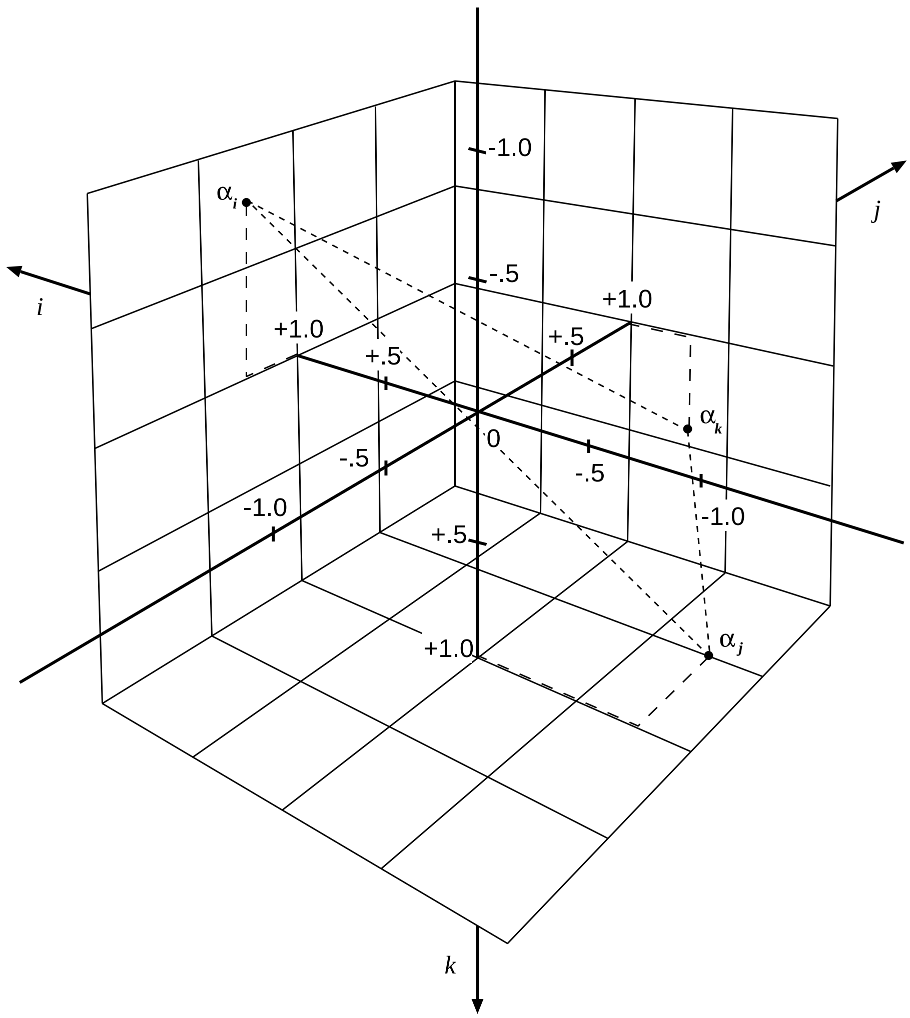

Figure 2: *Corpus* or α *space*, representing usage of terms i, j, and k by *corpus point* α_i, α_j and α_k, the δ distances (dotted lines) between which indicate usage differences of terms according to the texts analysed.

Synonymy of meanings may be explicated as equality of two fuzzy sets; *partial synonymy* of meanings may be defined by a similarity formula, introducing a threshold value s; *hyponymy* of a meaning relative to another may be explicated as containment of fuzzy sets. In so far as the operations of *negation, adjunction,* and *conjunction* are concerned, there has been quite a bit of critical discussion lately, particularly from the experimentalist's point of view. For the generation of new meaning points in the semantic space, I have thus far gone back to those definitions proposed by Zadeh (1965). The corresponding definitions in my linguistic model of structural lexical meaning are illustrated in Figures 3 and 4 and read as follows:

Synonymy (equality):

$$i = j \quad \text{iff} \quad \mu_i(k) = \mu_j(k) \quad \text{for all } k = 1, \ldots, n \tag{5}$$

Partial Synonymy (similarity):

$$i \approx j \quad \text{iff} \quad |\mu_i(k) - \mu_j(k)| \leq s \quad \text{for all } k = 1, \ldots, n \tag{6}$$

Hypnoymy (containment):

$$i \subset j \quad \text{iff} \quad \mu_i(k) \leq \mu_j(k) \quad \text{for all } k = 1, \ldots, n \tag{7}$$

Negation (complement):

$$\sim i := \mu_{i'}(k) = 1 - \mu_i(k) \quad \text{for all } k = 1, \ldots, n \tag{8}$$

Conjunction (intersection):

$$i \wedge j := \mu_{i \cap j}(k) = \min\big[\mu_i(k); \mu_j(k)\big] \quad \text{for all } k = 1, \ldots, n \tag{9}$$

Adjunction (union):

$$i \vee j := \mu_{i \cup j}(k) = \max\big[\mu_i(k); \mu_j(k)\big] \quad \text{for all } k = 1, \ldots, n \tag{10}$$

Empirical Analysis

By way of conclusion I would like to give some examples from the computer analysis of a corpus of nineteenth and twentieth century German students' poetry, the first part of which, covering the early nineteenth century, comprises of some 500 texts and a vocabulary of 315 lemmatized word types/21,000 tokens.

As there are serious difficulties in visualizing a 315-dimensional *semantic space* on the one hand, as there is, on the other, but little illustrative use in reproducing

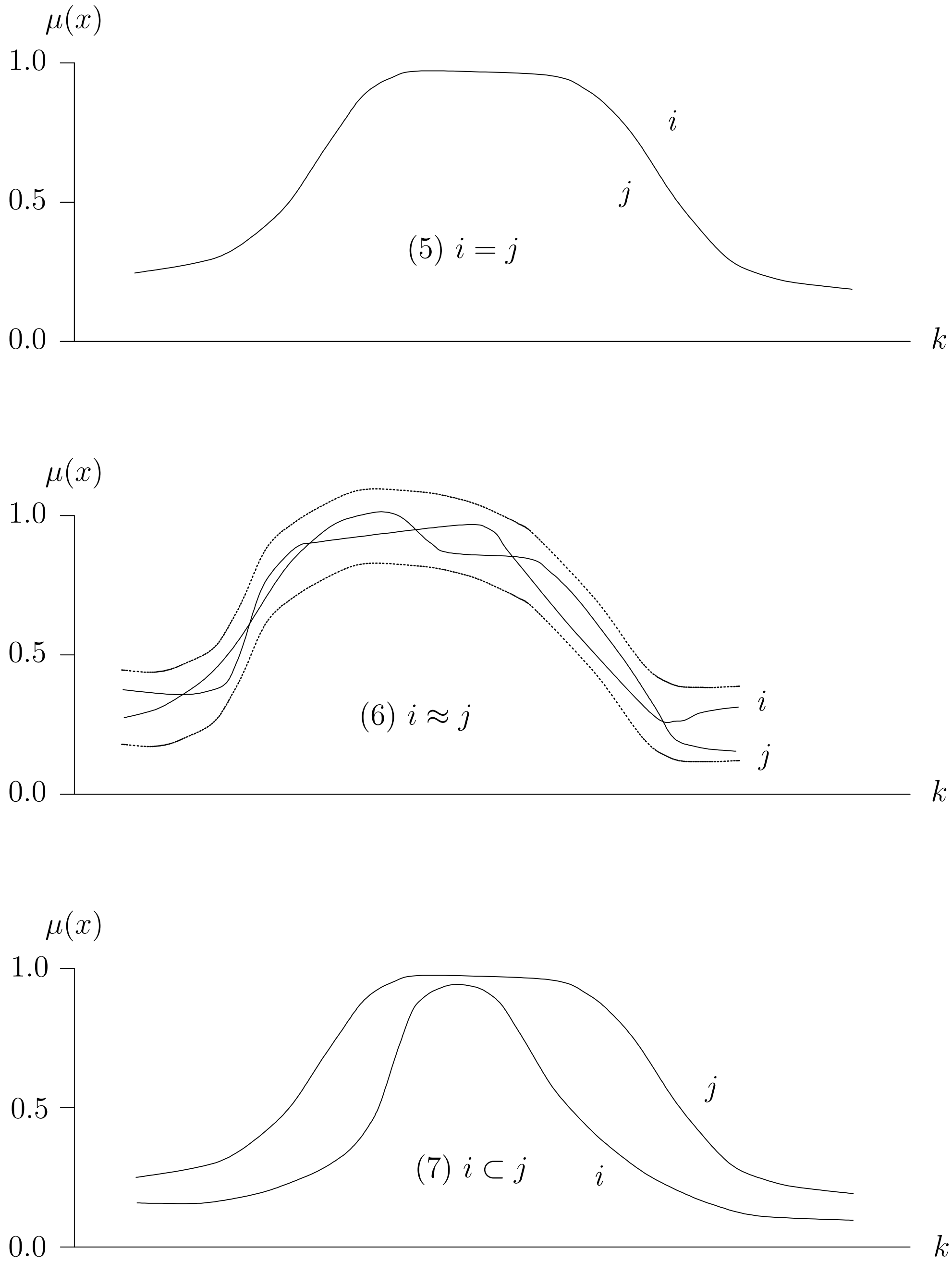

Figure 3: Fuzzy set-theoretical representation of *meaning points* (5) synonymy (equality), (6) partial synonymy (similarity), and (7) hyponymy (containment).

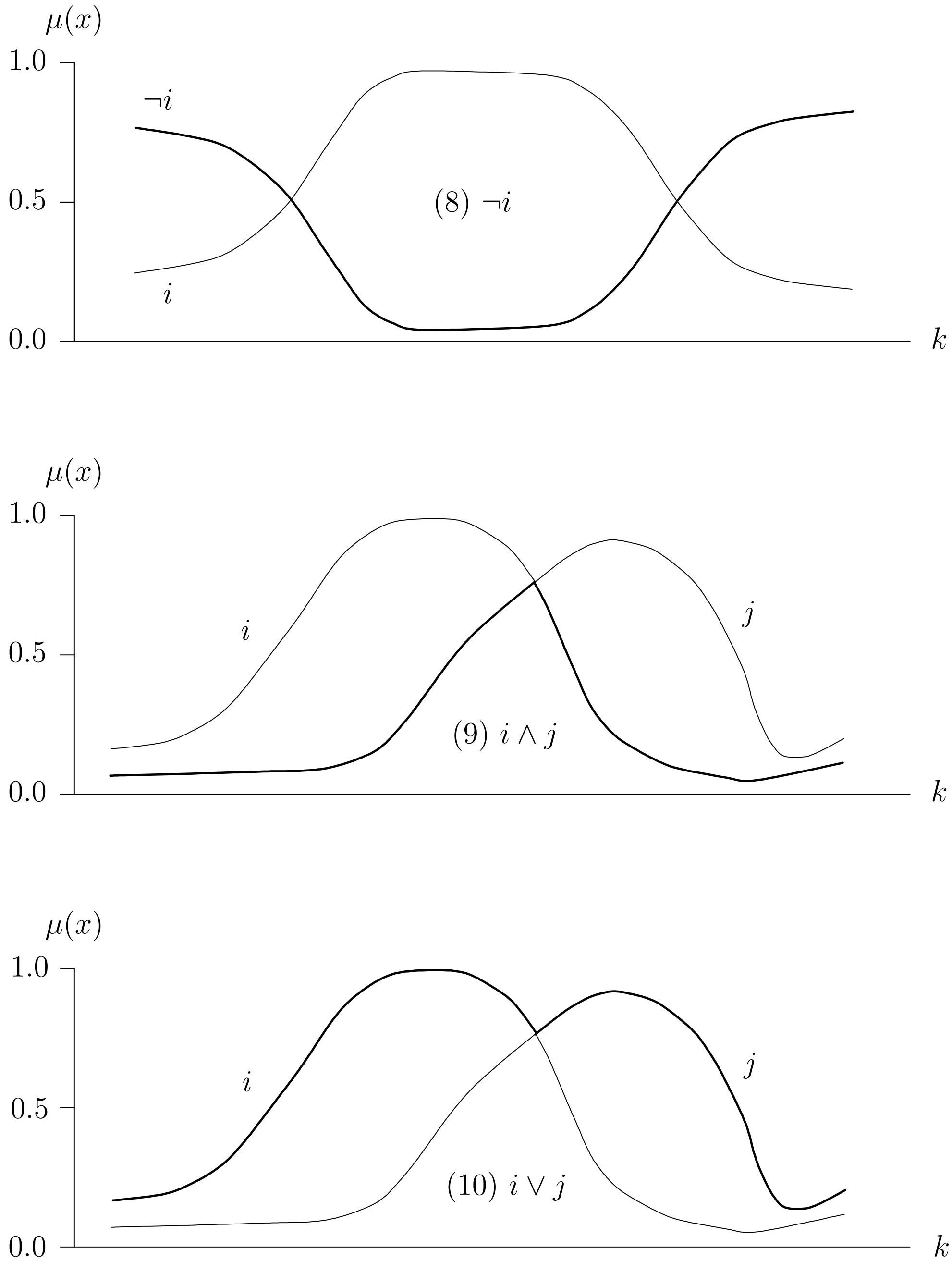

Figure 4: Fuzzy set-theoretical interpretation of *meaning points* (8) negation (complement), (9) conjunction (intersection), and (10) adjunction (union).

1.918	2.575	1.995	2.459	2.240	2.249	2.293	2.615
1.856	2.141	2.240	2.294	2.205	2.263	1.994	1.920
2.354	*0.000	2.084	2.165	1.652	1.850	2.381	2.008
2.101	2.062	2.495	1.933	2.179	2.494	2.404	0.000
1.?10	2.340	2.299	2.113	1.896	1.933	2.161	1.953
2.070	2.250	2.108	2.524	2.152	1.617	2.633	2.324
2.082	1.893	2.227	1.895	2.370	2.602	2.329	2.208
2.125	2.154	2.191	2.346	1.914	2.093	2.115	2.234
2.488	2.960	1.917	2.112	2.429	2.187	2.208	2.330
2.076	2.337	2.003	2.721	2.182	1.694	2.034	1.425
2.148	1.827	2.038	2.191	2.307	1.955	2.135	2.354
2.239	2.024	2.154	2.066	2.516	2.938	2.307	1.889
2.542	2.009	2.355	1.936	2.187	2.118	2.510	2.362
1.906	2.171	2.332	2.097	2.732	2.028	2.451	2.328
2.342	2.067	2.291	1.947	2.172	2.056	2.217	2.655
2.022	1.880	2.066	1.968	2.048	2.078	2.092	2.428
2.107	2.799	1.983	2.159	2.417	2.374	2.363	2.249
2.188	2.376	1.849	2.407	2.143	2.040	2.171	2.282
2.217	2.014	2.016	2.046	2.005	1.999	2.187	1.981
2.203	2.519	2.148	1.771	2.017	2.451	2.104	1.962
2.113	1.993	1.903	2.010	2.220	2.087	1.949	2.241
2.075	2.032	1.838	2.460	2.288	2.464	2.131	0.000
2.385	2.123	2.293	1.895	2.133	2.384	1.796	2.359
2.378	2.008	2.094	2.122	2.256	1.816	2.381	2.238
2.351	2.169	1.943	1.994	2.066	2.191	2.047	1.922
2.154	1.648	2.020	2.126	2.177	2.202	2.131	2.352
2.299	1.983	2.080	2.047	1.923	1.874	2.184	2.191
2.292	2.248	2.077	2.647	2.395	2.054	1.981	2.191
2.133	2.354	2.185	1.769	2.131	2.476	2.194	1.893
2.711	2.472	2.174	2.259	2.000	2.133	2.231	2.245
2.144	2.145	2.074	2.255	2.205	2.174	1.978	2.297
2.738	2.342	2.081	2.172	1.956	2.186	2.118	2.057
2.837	2.431	2.176	2.342	2.313	1.903	2.059	2.057
2.250	2.412	2.409	2.265	1.892	2.062	2.174	2.252
2.115	2.124	2.162	0.000	2.180	2.412	2.041	2.334
2.493	2.357	2.099	2.537	2.393	1.908	2.159	1.958
2.017	2.250	1.936	2.064	1.792	2.013	2.253	1.927
2.144	2.107	2.115	2.237	1.829	2.176	2.329	1.892
2.095	2.386	2.268	2.457	2.376	1.960	1.986	2.273
2.184	2.398	0.000					

Table 1: δ value n tuple ($n = 315$) of α or corpus point BLÜTE(BLOSSOM), defining its meaning point in semantic space/its fuzzy subset in the vocabulary, respectively.

Frühling spring	2.768	Baum tree	3.339	Duft fragrance	3.412
Rose rose	3.435	Lenz spring (poet.)	3.589	Schönheit beauty	3.641
Garten garden	3.788	Vogel bird	3.859	Wiese/Aue lea/meadow	3.971
hold gracious	3.983	Zweig/Ast branch	3.987	zart tender	3.995
Berg mountain	4.006	Traum dream	4.028	Gras grass	4.030
Nachtigall nightingale	4.031	Blume flower	4.042	Wunder miracle	4.050

Table 2: Blüte (blossom)

an n tuple of 315 δ values defining a meaning point or fuzzy set respectively of, say, the lexical item BLÜTE/blossom (Table 1), I have devised other ways of giving an impression of these lexical structures.

To illustrate the position of a meaning point, I have tabulated those points which are nearest to it in the semantic space, constituting something like a meaning point's topological environment. As I have shown elsewhere (Rieger 1974, 1975), these environments prove to be very similar to what linguists used to compose introspectively and what they have called *paradigmatic* or *semantic fields.*

When you let your eyes pass along the meanings points listed in Table 2 and Table 3, showing the environments of BLÜTE/blossom and FRIEDHOF/graveyard, you will get an idea of the semantic fields of these words as used in the German poems of the early nineteenth century. As far as paradigmatic relations are concerned, I think they are almost self-evident to a native speaker of German, or, to say the least, they are not contra-intuitive.

It should be noted, however, that paradigmatic relations vary considerably from one language to another; the word–word translation of meaning points from German into English in Tables 2 to 5 might be rather inadequate. It cannot be intended to depict comparable English paradigmatic relations, but has been given for illustrative purposes only. For comparable results, one would have to analyse a similar corpus of English texts.

My text-oriented approach to the empirical analysis and description of natural language meaning is obviously tentative and admittedly far away from a consistent theory of communicative semantics. I feel, however, that it might have a good chance of bringing statistical and theoretical linguistics together so as to arrive at a model of natural language word-meaning representation which, in its abstract (algebraic) parts, may be interpreted as a corpus-independent formal theory of semantic

Grab/Gruft grave/tomb	2.945	kalt cold	3.192	Stunde hour	3.494
Tod death	3.636	fahl/welk dim/faded	3.669	finster dark/sad	3.980
bleich pale/dim	4.141	hohl hollow	4.436	Schein shine	4.533
schwarz black	4.595	Abgrund gulf/depth	4.642	grau grey	4.718
Angst terror/fright	5.078	heilig holy	5.115	blaß colourless	5.487
Schweben hover	5.543	weiß white	5.977	gelb yellow	5.992

Table 3: Friedhof (graveyard)

From the material computed in order to test the above defined operations, generating *new* meanings, I have chosen two lexical items, namely BAUM/tree and BLÜTE/blossom which are closely related paradigmatically. The idea is, that the new meaning points 'BAUM/tree ∧ BLÜTE/blossom' (Table 4) and 'BAUM/tree ∨ BLÜTE/blossom' (Table 5) resulting from conjunction and adjunction of these two meanings, should be positioned somewhere in the same region of the semantic space, which in fact they are.

Zweig/Ast bough/branch	2.268	Blume flower	2.535	Blatt leaf	2.618
grün green	2.622	Garten garden	2.663	Frühling spring	2.673
Vogel bird	2.761	Duft fragrance	2.780	Wiese/Aue lea/meadow	2.865
Rose rose	2.885	leise low/faint	3.001	singen sing	3.121
Lenz spring (poet.)	3.127	Gold gold	3.158	Welle wave	3.221
Wunder miracle	3.224	Jubel joy	3.247	Luft air	3.302

Table 4: Baum ∧ Blüte (tree ∧ blossom)

Baum tree	2.489	Blüte blossom	2.489	Frühling spring	3.060
Garten garden	3.437	Duft fragrance	3.550	Zweig/Ast bough/branch	3.678
Gras grass	3.754	Lenz spring (poet.)	3.767	Traum dream	3.897
Laub leaves	3.937	Rose rose	3.941	Eiche oak tree	3.964
Blatt leaf	3.971	Vogel bird	3.972	Feld field	3.991
Pracht splendour	3.994	zart tender	4.017	Nachtigall nightingale	4.049

Table 5: Baum $\vee$ Blüte (tree $\vee$ blossom)

competence ('langue'), whereas its empirical (quantitative) parts will represent the performative lexical data ('parole') which are corpus-dependent and hence will vary according to the pragmatic frame of the texts analysed.

> The problem is that lexical relations are properly tested by being part of the information utilized by a grammar. But theoretical linguistics has had, for example, no place for a thesaurus except, possibly, as a source of semantic markers ... Until very recently, the structure of the lexicon has been a matter of relatively little interest to linguists, and they have as yet found little of substance to say about it. This situation can be expected to change ... Today, the interest is primarily in semantic problems and, more specifically, those associated with characterizing word meaning.
> Sparck Jones and Kay:1973:172

As I have been trying to show, fuzzy set theory allows vague lexical meanings of words to be represented in precise terms. The vagaries of actual word usage by individuals, social groups, in certain pragmatic situations, or otherwise, do not even need to be reduced to strict binary determinateness, but rather become the empirical basis for any structural meaning's representation which depicts the (semantic) regularities followed and/or established by real communicators, and consequently are computed from sets of natural language texts which might be determined individually, sociologically, pragmatically, or otherwise.

But fuzzy set theory may also provide the means of linking the numeric analysis of texts and model-theory semantics. This, of course, is a topic of its own which would have to be dealt with in a separate paper.

The only thing that can be indicated here is that (similar to the way the operations of conjunction, adjunction, negation, etc., of meanings were defined within lexical structure, called semantic space above) syntactical categories (which had been ignored during the empirical analysis of semantic regularities) might formally be introduced as functors in the sense of categorial grammar (Cresswell 1973). The meanings or intensions of these syntactical functors would be functions that operate on meaning points as their arguments and assign, when applied, a new meaning point to the complex called sentence. The notion of semantic space could possibly be interpreted as a model $\langle \pounds\langle i,j\rangle\rangle$ in Montague grammar (Montague 1970) where $\pounds$ is a Fregean interpretation of the language L and $\langle i,j\rangle$ is a point of reference of L with i being an actual world and j an actual context of language usage constituting the pragmatical basis specified by the formal model and its empirical data respectively.

Acknowledgements

This paper, an earlier version of which was presented at the Third International Conference on Computing in the Humanities (ICCH/3), University of Waterloo, Ontario, Canada, reports on some parts of a research project which was substantially supported by a DFG grant (German Research Foundation) during 1975 and 1976. A more elaborate discussion of the project's other aspects may be found in Rieger 1976, l977a, b.

I would like to thank my colleague Dr. H.M. Dannhauer who did the necessary programming for this project on the CDC/Cyber 175 at the Aachen Technical University Computing Centre.

References

Bierwisch, M. (1970) Semantics. *New Horizons in Linguistics,* pp. 116–84 (ed. Lyons, J.). Harmondsworth: Pelican.

Cresswell, M.J. (1973) *Logics and Languages.* London: Methuen.

Gaines, B.R. (1976) Foundations of fuzzy reasoning. *International Journal of Man-Machine Studies,* **8,** 623–68.

Montague, R. (1970) Universal grammar. *Theoria,* **36,** 373–98.

Peirce, C. S. (1965) *Collected Papers* (ed. Hartshorne & Weiss). Cambridge, Mass.: MIT Press.

Quine, W. van Orman (1960) *Word and Object.* Cambridge, Mass.: MIT Press.

Rieger, B. (1971) Wort- und Motivkreise als Konstituenten lyrischer Umgebungsfelder. Eine quantitative Analyse semantisch bestimmter Textelemente. *Zeitschrift für Literaturwissenschaft und Linguistik (LiLi),* **4,** 23–41.

Rieger, B. (1972) Warum mengenorientierte Textwissenschaft? Zur Begründung der Statistik als Methode. *LiLi,* **8,** 11–28.

Rieger, B. (1974) Eine tolerante Lexikonstruktur. Zur Abbildung natürlich-sprachlicher Bedeutung auf unscharfe Mengen in Toleranzräumen. *LiLi,* **16,** 31–47.

Rieger, B. (1975) On a tolerance topology model of natural language meaning. Second International Conference on Computer and the Humanities ICCH/2. Los Angeles: University of Southern California. (Mimeograph).

Rieger, B. (1976) Unscharfe Semantik natürlicher Sprache. Zum Problem der Repräsentation und Analyse vager Bedeutung. Linguistics Symposion, Deutsche Akademie der Naturforscher Leopoldina, Halle. To appear 1979 in *Nova Acta Leopoldina.*

Rieger, B. (1977a) Theorie der unscharfen Mengen und empirische Textanalyse. *Methoden der Textanalyse (medium literatur 3),* pp. 84–99 (ed. Klein, W.) Heidelberg: Quelle u. Meyer.

Rieger, B. (1977b) Vagheit als Problem der linguistischen Semantik. *Semantik und Pragmatik,* pp. 91–101 (ed. Sprengel, K, Bald, & Viethen, H.W.). Tübingen: Niemeyer.

Russell, B. (1923) Vagueness. *Australian Journal of Psychology and Philosophy,* **1,** 84–92.

Sparck Jones, K. & Kay, M. (1973) *Linguistics and Information Science.* New York & London: Academic Press.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control,* **8,** 338–53.

Zadeh, L.A. (1971) Quantitative fuzzy semantics. *Information Science,* **3,** 159–76.

Zadeh, L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning. *Information Science,* **8,** 199–249 and 301–57, **9,** 43–80.