Feasible Fuzzy Semantics
On Some Problems of How to Handle Word Meaning Empirically[*]

Burghard B. Rieger

> "There is no need for us to emphasize either the fact that meaning is the crucial feature of language or the fact that it is the most intractable problem of linguistics. Semantics, the study of meaning, has a long and eminently respectable history as an activity for philosophers, logicians, grammarians, philologists and linguists, but unfortunately the obviousness of meaning of words and discourse is matched by its eel-like slipperyness when the philosopher or linguist tries to catch it." (Sparck-Jones / Kay 1973, 120)

This introductory quotation strikingly characterizes the situation and illustrates metaphorically the sort of difficulties we are going to encounter when engaging in the investigation of natural language word meaning. Besides, it hopefully will stimulate the reader's expectations (where necessary), or (where appropriate) will let her / him be prepared to be left empty-handed at the end.

0. Introduction

In discussing word-semantics and its empirical, possibly automatic, procedures at that, one has to face at least two problem areas

a) the specification of the language data to be analysed empirically, preferably by aid of computer, and

b) the sort of operational procedures, preferably algorithmic, to be employed in view of both, word meaning analysis and representation.

My current research in the field of quantitative and computational linguistics, being concerned with statistical methods of text-analysis and formal notations of fuzzy sets theory, suggests these to be applied to problem-area b). But before going about some of the procedures developed, and results tested in the project so far, there will have some remarks to be made on the frame-conditions the basic language material has to satisfy in order to make the approach work. And this concernes problem-area a).

As these issues have evolved through various stages which consequently have been discussed at some length elsewhere (Rieger 1977a, c; 1978; 1979 a, d), I shall be rather brief on them here, referring to focal issues only. However, as problems of word-semantics should be discussed where and when they come up, I will try to give an account of the philosophy — so to speak — behind my approach. I therefore shall have to point out

[*] Published in: Eikmeyer, H.J. / Rieser, H. (Eds.): Words, Worlds, and Contexts. New Approaches in Word Semantics (Research in Text Theory 6), Berlin / New York (W. de Gruyter) 1981, pp. 193–209.

some aspects of formal and descriptive theory construction, the empirical complications to be expected in view of a semiotic domain like word semantics, and how some of its problems can be tackled and even be solved.

1. Aims of Word Semantics

It is a truism by now that there is no linguistic theory of semantics that could explain why automatic retrieval procedures do in fact work — and that there is quite a number of indexing and retrieval systems' designers who can do very well without any specific linguistic analysis of their material. And yet, when we look up linguistic theories of sentence- or even text-semantics on the one hand, and procedures of intellectual or statistical indexing systems on the other, and see what both of them can offer in respect of word-meaning, we will in either case be confronted with special word-lists. The purpose of these lists, which may be relationally structured or just sequential, is to specify more or less comprehensively the conditions under which a term listed in a dictionary or thesaurus may be related to, or even identified with, certain meanings, represented by meaning-components, semantic-markers or -descriptors.

A *dictionary* in generative grammars may be considered as a sequential word-list that specifies syntactical, semantical and perhaps pragmatical restrictions of each of its entries. These have to be observed for the proper insertion of elements, or groups of elements into sentential or textual structures to generate or parse grammatically correct and meaningful surfaces.

And a *thesaurus* in indexing systems can be regarded as a structured word-list that specifies the lexicological or conceptual relations of each of its entries. These will serve in turn as meaning descriptors which are assigned to elements, or groups of elements in sentences or texts to constitute relevant meaning characterizations.

On the basis of such listings which provide different kinds of semantic information under each word-heading, sentence-semantics as well as indexing systems *are making use* of word-meaning instead of *analysing* it. Apart from tentative departures within generative semantics or statistical indexing, there have no operational procedures yet been devised for the semantic analysis and description of natural language terms, as a result of which, when applied to language data, a lexical structure may eventually be obtained.

Now, this is what word-semantics should and could do, and where exactly the problems begin.

2. Theory and Model Construction

If we agree that linguistics is, or at least ought to be an empirical discipline, then the paradigm of empirical sciences should be followed, although it needs modification in view of the scope of natural language semantics.

To adopt the paradigm of empirical sciences for linguistic research is tantamount to at least two postulates,

first: not to rely on ready-made theories or models taken from another domain, because these may be grossly inadequate in respect of the phenomena to be investigated;

second: not to rely on the introspective exploration of one's own knowledge and competence as the allegedly inexhaustible data-source, although it very often serves as a first guide and may produce valuable initial ideas.

Instead, the investigation of linguistic problems in general, and that of word-semantics in particular, should start with more or less pre-theoretical working hypotheses, formulated and re-formulated for continuous estimation and / or testing against observable data, then proceed to incorporate its findings tentatively in some preliminary theoretical set up which finally may perhaps get formalized to become part of an encompassing abstract theory. Our objective being natural language meaning, this operational approach would have to be what I would like to call *semiotic*.

This term is meant to refer to certain new proceedings which have in common that they do not insist to make imprecise phenomena precise in order to render them accessible to exact analysis. Their descriptive and / or formal framework is designed to fit the phenomena, not to straiten the phenomena to fit a model or theory (Gaines 1977). Thus, they are well suited to cover processes of semantisation, both (a) of how in acts of cognition (via processes of continuous social and / or physical experience, conditioning, learning or whatever) entities first undistinguished become more and more differentiated until finally more or less discrete units may be isolated from the unstructured mass of encountered phenomena by establishing (fuzzy) memberships in, relations among, and whole structures of categories of social and / or physical environments which constitute what has come to be known as *world-knowledge systems* in general, and in particular (b) of how in acts of communication (via environmentally similar processes of verbal interaction, exercise, repetition, or whatever) this system of word-knowledge, however vague or fragmentary yet, is simultaneously associated with conventionalized verbal behaviour the regularities of which will not only determine (more or less strictly) the structure of discourse on any of its semiotic (morpho-phonetic, syntactic, semantic, pragmatic) levels but as such will also be utilized to represent (denote, enlarge, modify, generate, or even create new) world-knowledge, constituting what is commonly called the *natural language system* (Rieger 1977b).

Following the line of Labov (1973) and others, prevailing linguistic theory and linguistic semantics in particular is dominated by what has been called the 'categorial view'. According to it, linguistic entities should either be discrete, invariant, qualitatively distinct, and composed of atomic primes, or else be of no use in linguistic theory at all. This view has led to the exclusion of very obvious object-level features of language usage, which only recently have begun to be recovered by linguistics proper, in some cases reluctantly but nevertheless continuously. Most prominent among these features is that of *word-meaning* itself, which — although recognized — is not an integral part of linguistic sentence- or text-semantics yet. Features of language *variation* on the morpho-phonetic level and those of *vagueness* on the lexico-semantic level are other well-known instances. They too gain increasing importance since language usage regularities are investigated empirically.

These aspects of the object-level semiotic phenomena however, are to be complemented by aspects of their formal notation. Hence, even theories of language performance, designed to account for phenomena like word-meanings' vagueness or variation, have to meet basic conditions of theory construction. Consequently these entities should again be well-defined on the meta-theoretic levels of representation where the dominance and validity of the 'categorial view' has to be maintained for formal, simulative, or descriptive models of controlled reconstruction even of semiotic phenomena.

This admittedly rough-and-ready distinction of object- and meta-theory, corresponding to different notational levels, requires some mediation. This can be provided, as I see it, *formally* by means of fuzzy set theoretical notations, and *operationally* by means of empirical procedures assigned to them. Applied to natural language data, they will interrelate observable but essentially *fuzzy* language phenomena on the one hand, and formal but finally *categorial* notations of their linguistic descriptions on the other.

Thus, findings and / or hypotheses on either side may become testable against each other, allowing for mutual modifications in the course of gradual improvement and increasing adequacy of the model and what it represents.

3. Structure of Meaning

Up to this point we have been reflecting upon only one part of the problem, or if you like to keep the picture, we have seen only one side of the slippery eel, namely, *how* semiotic phenomena (which are permanently experienced and observed in language use) should be accounted for by different notational levels of formal representation. What makes the study of natural language meaning an even more intricate problem, depends on the other part of the picture and that concerns the particular nature of *what* has to be represented, namely, a representational structure in its own.

It is this representational aspect of language which traditional theories of semantics have particularly been convinced of. According to the most influential theories, natural language meaning can be characterized by its denotative and its connotative aspects leading to different models:

Referential approach	*Model construction*
Denotation $:= L \to W$	Den: $L \times W \to M_D$
Structural approach	*Model construction*
Connotation $:= L \to L$	Con: $L \times L \to M_C$

Denotation is understood to constitute *referential* meaning which may be modelled as a system M_D of relations between words or sentences of a language L and the objects or processes they refer to in W.

Connotion is defined to constitute *structural* meaning which may be modelled as a system M_C relating words or sentences of a language L conceptually to one another.

Referential semantic theory is truth-functional and formally elaborated but as such not prepared to account satisfactorily for the vagueness inherent in natural language word meaning; whereas structural semantics has considered vagueness of meaning somewhat fundamental of natural language but, being based mainly upon intuitive introspection, it has not achieved the theoretical or methodological consistency of formal theories.

Although both approaches differ in what they consider natural language meaning to be, they nonetheless converge on the central notion of it, being a relation between a representation (i.e. natural language discourse) and that which it represents (i.e. referential or structural meaning) without bothering too much about what this relation stands for, how it is established and, hence, may be reconstructed operationally.

Let us assume that the communicator's ability to intend and comprehend meanings in verbal (natural language) interaction can be considered a phenomenological undoubtable, empirically well established, and theoretically at least defensible common basis of any study of natural language meaning. Then our approach should not end up with but rather start from the communicative property of discourse.

If it is true that verbal communication is a highly coded form of social interaction in which — as Halliday (1977, p. 207) put it — ”the interactants are continuously supplying the information that is missing in the text” from their situational and environmental knowledge of the world that particular piece of discourse is possibly referring to, the empirical task of reconstructing this complex system of (common) world-knowledge or fragments of it and its (possible) linguistic form of representation and / or perception gains prior importance over formal theory construction.

> By ‘text’, then, we understand a continuous process of semantic choice. Text is meaning and meaning [implies, rather than] is choice, an ongoing current of selections each in its paradigmatic environment of what might have been meant (but was not). It is the paradigmatic environment — the innumerable subsystems that make up the semantic system — that must provide the basis of the description, if the text is to be related to higher orders of meaning, whether social, literary or of some other semiotic universe. (Halliday 1977, p. 195).

As we accept a *sentence* to be a lexico-grammatical unit which is not composed of phonemes but is perceived in phonemes that constitute a morpho-phonological system, or as we may interpret a *text* to be a semantic unit which is not composed of sentences but is realized in sentences that constitute a lexico-grammatical system, we may quite as well conceive of a *frame* as a pragmatic unit which is not composed of texts but may be given access to by texts that constitute a semantic system environmentally conditioned as its register (Rieger 1979d).

Instead of focussing on the singular text and its comprehension, however, it is the lexico-semantic description of the paradigmatic environment of possible discourses we are interested in now, with the texts serving as the only accessible data source for the system's description.

> The system is a meaning potential which is actualized in the form of text; a text is an instance of social meaning in a particular context of situation. We shall therefore expect to find the situation embodied or enshrined in the text not piecemeal, but in a way which reflects the systematic relation between the semantic structure and the social environment. (Halliday 1977, p. 199).

Thus, it must not be the singular realization the analysis and description of word meanings has to focus on but the great number of textual instantiations (tokens) of a particular register (type). Again, other than Halliday's taking up the systematic relation between the semantic structure and the social environment, we will focus here on the semantic system's relational structure by trying to keep constant the type variable of social and general communicative environment for the selection of possible discourse tokens.

Assembled in a corpus, these discourses may serve as a sample of all possible texts which in fact have been (or could have been) produced under the particular frame conditions concerned, providing the database for an empirical analysis and description of the structural properties of the paradigmatic environment in terms of lexical units employed (Rieger 1979c).

4. A New Formal Approach

It is the throughout relational structure of meaning that obviously allows the concept of fuzzy sets and relations to be employed to incorporate vagueness into both referential and structural theories of semantics.

The most recent, and at that, most comprehensive formal approach I know of to tackle the problem of natural language meaning, is that of L. A. Zadeh (1978).

Under the acronym PRUF for 'Possibilistic, Relational, Universal, Fuzzy' he has devised a meaning representation language for natural languages which is possibilistic instead of truth-functional, and whose dictionary provides linguistically labeled fuzzy subsets of the universe, instead of sets of semantic markers under word-headings.

The conceptual structure of PRUF is based on the premise that, in contrast to formal languages and notational systems, natural languages are intrinsically incapable of precise characterization on either the syntactic or semantic, to say nothing of the pragmatic level.

> In the *first* place, the pressure for brevity of discourse tends to make natural languages maximally ambiguous in the sense that the level of ambiguity in human communication is usually near the limit of what is disambiguable through the use of an external body of knowledge which is shared by the parties in discourse.
> *Second,* a significant fraction of sentences in a natural language cannot be characterized as strictly grammatical or ungrammatical. [...]
> *Third,* [...] a word in a natural language is usually a summary of a complex, multifaceted concept which is incapable of precise characterization. For this reason, the denotation of a word is generally a fuzzy — rather than non-fuzzy — subset of a universe fo discourse. (Zadeh 1978, p. 397).

The basic idea, upon which this approach hinges, is that a referential meaning may be explicated as a fuzzy correspondence between language terms and a universe of discourse.

This correspondence, *L*, is formally defined to be a fuzzy binary relation from a set of language terms, *T*, to a universe of discourse, *U*. As a fuzzy relation, *L*, is characterized by a membership-function

$$\mu_L : T \times U \to [0,1]; \quad x \in T,\ z \in U; \quad 0 \leq \mu_L(x,z) \leq 1 \tag{1}$$

which associates with each ordered pair (x,z) its grade of membership $\mu_L(x,z)$, being a numeric value between 0 and 1, in *L*, so that

$$L := \Big\{ \big((x,z), \mu_L(x,z)\big) \Big\} \tag{2}$$

The fuzzy relation L now induces a bilateral correspondence according to which

first: the *referential meaning* of an element x' in T may be explicated as the fuzzy subset $M(x')$ in U, assigned to it by the membership function μ_L conditioned on x',

$$M(x') := \mu_L(z,x') := \Big\langle \big((x',z_1), \mu_L(x',z_1)\big), \ldots \big((x',z_n), \mu_L(x',z_n)\big) \Big\rangle \tag{3}$$

second: the *linguistic description* of an element z' in U may be given as a fuzzy subset $D(z')$ in T assigned to it by the membership function μ_L conditioned on z'

$$D(z') := \mu_L(x,z') := \Big\langle \big((x_1,z'), \mu_L(x_1,z')\big), \ldots, \big((x_n,z'), \mu_L(x_n,z')\big) \Big\rangle \tag{4}$$

Although formally satisfactory — as outlined and illustrated by PRUF — the basic assumption of the approach concerning the referential nature of natural language meaning proves to be crucial for its empirical applicability: in order to determine the membership-grades of a fuzzy set, or fuzzy relation respectively, one has to have access to relevant empirical data defined to constitute the sets, and some operational means to calculate the numerical values from these data.

As the domain of the fuzzy relation μ_L contains not only the set of terms of a language, T, but also the set of objects and / or processes these terms are believed to denote in the universe, U, both these sets should be accessible in order to let an empirical procedure be devised that could be assigned to μ_L. All that Zadeh (1978) is offering in that respect, stays empirically rather vague. He assumes that "each of the symbols or names in T may be defined ostensively or by exemplification. That is by pointing or otherwise focussing on a real or abstract object in U and indicating the degree — on the scale from 0 to 1 — to which it is compatible with the symbol in question" (p. 418).

This cannot be considered a solution which may be called both *semiotic* and *operational* in the above given sense. Taken to be executable, Zadeh's suggestion necessarily involves probands' questioning about what they think or believe a term denotes. Thus, the procedure would again have to rely on the individual introspection of a multitude of competent speakers, instead of making these speakers employ the term's denotational and / or connotational function in the course of communicative verbal interaction. However, experimental psychology has taught us to expect considerable differences between what people think they *would do* under certain presupposed conditions, and what in fact they *will do* when these conditions are real. And there is every reason to assume that this difference is found in cases of language performance, too.

So, it would seem to be more realistic to make natural language usage the basis for identifying those language regularities, which real speakers / hearers follow and / or establish in discourse as a consequence of which natural language meaning (whatever that may be) can obviously not only be intended and understood, but may also be analysed and represented.

As this seems to be the only certainty about meaning, anyway, namely that it can only be constituted by means of natural language texts, these should also be able to provide the necessary data with the advantage of being empirically accessible. Assembled in a corpus, the usage regularities which the lexical items employed produce, may thus be analysed statistically with the numerical values obtained to define fuzzy vocabulary mappings (Rieger 1979b).

5. *An Empirical Reconstruction*

Following this line of argument is to ask for a connotational supplement to the denotational approach Zadeh forwarded so far. This goes along with a necessary re-interpretation of what the sets T and U (1) in the referential meaning relation possibly stand for.

From a structural point-of-view, and in accordance with what has been said above about the semiotic nature of processes constituting both, natural language systems and world-knowledge systems, T is to be regarded not just as a set of terms of a language any more, but as a system of lexical units the usage regularities of which induce a relational structure. This structure does not just allow for a set of objects and / or processes in U

to be denoted, but it constitutes a (fuzzy) relational frame for the notation of a system of concept-points which is dependent on, but not identical with the one induced by the usage regularities of terms as employed and identified in natural language discourse. Hence, some operational means to analyse (preferably automatically) and to represent (preferably numerically) these regularities have to be devised (preferably as a best fit) for the process of semantisation (Rieger 1977b).

Thus, being a non-symmetric, fuzzy relation, μ_L can empirically be reconstructed only on the basis of natural language discourse data. So far, statistical procedures have been used for the reconstruction by a consecutive mapping in three stages from T to constitute U, providing the membership-grades for μ_L.

On the *first* stage co-occurrences of terms are not just counted but the intensities of co-occurring terms in the texts of the database are calculated. This is done by a modified correlation-coefficient α that measures mutual (positive) affinity or (negative) repugnancy of pairs of terms $x,\ x' \in T$ by real numbers from the interval $[-1,+1]$. α can therefore be considered a fuzzy relation in the cross-product of the set of terms T used in the texts analysed

$$\alpha : T \times T \to I,\ I = [-1,+1]; \quad T := \{x_i\},\ i = 1,\dots,n \tag{5}$$

By conditioning this fuzzy relation α on the $x_i \in T$, we get a non-fuzzy mapping

$$\alpha|x_i : T \to C,\ C := I^n \tag{6}$$

This mapping assigns to each $x \in T$ one and only one so-called *corpus-point* y in the corpus space C, defined by the n-tupel of membership-grades $\alpha(x_i, x)$

$$\alpha(x_i, x) := y \in C \tag{7}$$

Each corpus-point $y' \in C$ may thus be considered a formal notation of the usage regularities, measured by grades of intensity, any one term x' shows against all the other terms $x_i \in T$.

On the *second* stage the differences of usage is calculated. This is done by a distance measure δ_1, which yields real, non-negative, numerical values from an interval standardized to $[0,1]$ to denote the distances between any two corpus-points $y,\ y' \in C$. δ_1 can also be considered a fuzzy relation in the cross-product of C, namely the set of all corpus-points y_i defined to constitute the corpus space

$$\delta_1 : C \times C \to J; \quad J := [0,1]; \quad C := \{y_i\},\ i = 1,\dots,n \tag{8}$$

By conditioning this fuzzy relation δ_1 on the y_i (or — following (7) — the x_i respectively) we get a non-fuzzy mapping

$$\delta_1|x_i : C \to U; \quad U := J^n \tag{9}$$

This mapping assigns to each $y \in C$ (or $x \in T$ respectively) one and only one so-called meaning- or *concept-point* z in the semantic space U, defined by the n-tupel of distance-values $\delta_1(y_i, y)$

$$\delta_1(y_i, x) = \delta_1(y_i, y) := z \in U \tag{10}$$

Each concept-point $z' \in U$ may thus be considered a formal notation of all the differences of all usage regularities, as a function of which the meaning of a term $x' \in T$ can be characterized.

Therefore it can be identified — according to (7) — with (4) the *linguistic description*, $D(z')$, of a concept-point z' which is a fuzzy subset in T

$$\delta_1(x_i, z') := D(z') \subseteq T \tag{11}$$

On the *third* stage of the consecutive mapping, there will topological environments of concept-points be calculated — in analogy to (8) — by a distance measure δ_2 which specifies the distances between any two $z,\ z' \in U$. Thus again, δ_2 may also be interpreted as a fuzzy, binary relation in the cross-product of U, i.e. the set of all concept-points z_i defined to constitute the semantic space

$$\delta_2 : U \times U \to J; \quad J := [0,1]; \quad U := \{z_i\}; \quad i = 1, \ldots, n \tag{12}$$

The conditioning of δ_2 on the z_i results in a non-fuzzy mapping

$$\delta_2 | z_i : U \to J^n \tag{13}$$

which assigns to each $z \in U$ (and — following (10) — $x \in T$ respectively) one and only one n-tupel of distances that — scaled according to decreasing values — will constitute the environment $E(z)$

$$\delta_2(z_i, x) = \delta_2(z_i, z) := E(z) \tag{14}$$

Any such environment $E(z')$ can be considered a formal means to describe the position of a concept point z' by its adjacent neighbours in the semantic space which is constituted by functions of differences of language usage regularities. $E(z')$ can therefore be identified — following (10) and (14) — with (3) the *conceptual meaning, $M(x')$*, of a term x' which is a fuzzy subset in U

$$\delta_2(z_i, x') := M(x') \subseteq U \tag{15}$$

We are now in the position to assign to the fuzzy relation

$$\mu_L : T \times U \to [0,1] \tag{16}$$

and the two-sided correspondence (3) and (4) induced by it, the following operations:

The two distance measures δ_1 (8) and δ_2 (12), operating on numerical data obtained from the correlational analysis (5) of lexical items employed in a corpus of natural language texts, will determine the membership-grades to be associated with (16), namely for the correspondence (4) induced by μ_L according to (9) inserting

$$\delta_1 | x_i = \mu_L(x_i, z_i) = \{D(z)\} \subseteq T \tag{17}$$

and for its inversion the correspondence (3) according to (13) inserting

$$\delta_2 | z_i := \mu_{L^{-1}}(x_i, z_i) = \{M(x)\} \subseteq U \tag{18}$$

To conclude with, we have to give the coefficients alluded to above which have experimentally been used by insertion into (5) to calculate the actual α-values, and inserted into (8) and (12) respectively, to calculate the δ_1- and δ_2-values from the data.

Given the lemmatized vocabulary V as a proper subset of T of lexical units

$$V := \{x_i\};\ i = 1, \ldots, n \tag{19}$$

employed in a corpus K of natural language texts as specified above

$$K := \{t\};\ t = 1, \ldots, m \tag{20}$$

where

$$S = \sum_{t=1}^{m} s_t;\ 1_t \leq s_t \leq S \tag{21}$$

is the sum S of all text-lengths s_t measured by the number of lexical units (tokens) in the corpus, and

$$H = \sum_{t=1}^{m} h_t;\ 1_t \leq h_t \leq H \tag{22}$$

is the total frequency H of a lexical unit x (type) computed over all texts in the corpus, then the modified correlation-coefficient α (5) reads

$$\alpha(x, x') = \frac{\sum\limits_{t=1}^{m} (h_t - h_t^*)(h'_t - h'^*_t)}{\left(\sum\limits_{t=1}^{m} (h_t - h_t^*)^2 \sum\limits_{t=1}^{m} (h'_t - h'^*_t)^2 \right)^{\frac{1}{2}}};\quad -1 \leq \alpha(x, x') \leq +1 \tag{23}$$

$$\text{where}\quad h_t^* = \frac{H}{S} s_t \quad \text{and} \quad h'^*_t = \frac{H'}{S} s_t.$$

The distances δ_1 (8) and δ_2 (12) have been calculated according to the Euclidean measure which reads

$$\delta_1(y, y') = \left(\sum_{i=1}^{n} \big(\alpha(x, x_i) - \alpha(x', x_i)\big)^2 \right)^{\frac{1}{2}};\quad 0 \leq \delta_1(y, y') \leq 2\sqrt{n} \tag{24}$$

and

$$\delta_2(z, z') = \left(\sum_{i=1}^{n} \big(\delta_1(y, y_i) - \delta_1(y', y_i)\big)^2 \right)^{\frac{1}{2}};\quad 0 \leq \delta_2(z, z') \leq 2n \tag{25}$$

As these distance measures are to be considered the metrics of the corpus space C and the semantic space U respectively, it should be noted here that so far the assumption of it being Euclidean is nothing but a first (although operational) guess. Experiments with different and more sophisticated distance measures developed are currently undertaken which eventually might prove to be more adequate in modelling linguistical systems' structures.

STRUKTURELLE BEDEUTUNG B(X) DIE WELT
X = INDUSTRIE / IEREN

ELEKTRO / NISCH	.704	COMPUTER	.715
DIPLOM	.765	ERFAHREN / UNG	.838
SUCHE / N	.859	SCHREIBEN	.926
SYSTEM / ATISCH	.951	FAEHIG / KEIT	.973
ZONE	1.000	SCHULE / R	1.046
BERUF / LICH	1.072	KENNEN / TNIS	1.141
WUNSCH / EN	1.318	TECHNIK / ISCH	1.339
ORGANISATION	1.411	VERBAND	1.515
UNTERRICHT / EN	1.611	STADT	1.787
GEBIET	1.935	STELLE	2.027
VERANTWORTEN / TUNG	2.085	UNTERNEHMEN / R	2.095
BITTE / N	2.114	AUSGABE / GEBEN	2.210
ALLGEMEIN	2.287	VERWALTEN / UNG	2.420
PERSON / LICH / KEIT	2.545	VERKEHR / EN	2.686
EINSATZ / EN	2.699	ANBIETEN / GEBOT	2.744

STRUKTURELLE BEDEUTUNG B(X) NEUES DEUTSCHLAND
X = INDUSTRIE / IEREN

WISSENSCHAFT	1.407	BAUER	1.866
LANDWIRT / SCHAFT	1.869	STUNDE	1.956
AUSSTELLUNG	2.030	BERATEN / UNG	2.218
ARBEITER / IN	2.263	BAUER / IN	2.406
KULTUR / EN	2.627	ERFOLG / REICH	2.632
PROMINENT	2.750	QUALITAET	3.035
CHEMIE / SCH	3.239	BESUCH / EN	3.273
GENOSSE / IN	3.297	SPEZIAL / IST	3.320
PLAN	3.322	LPG	3.373
BERICHT / EN	3.400	KONGRESZ	3.409
PRODUKTION	3.472	JAHR / IG / LICH	3.512
SEKRETAER	3.521	LOHN	3.545
LEITEN / UNG	3.563	SOWJET / ISCH / UNION	3.571
KONZERT	3.616	BAU / EN	3.624
REGIEREN / UNG	3.630	TOD	3.646

Table 1. Linguistic description $D(z) = B(x)$ of the lexical entry INDUSTRIE / IALISIEREN (industry / ialise) from DW and ND

SEMANTISCHE UMGEBUNG E(X) DIE WELT
X = INDUSTRIE / IEREN

ELEKTRO / NISCH	2.106	LEITEN / R / UNG	2.369
BERUF / LICH	2.507	SCHULE / R	3.229
SCHREIBEN	3.328	COMPUTER	3.667
FAEHIG / KEIT	3.959	SYSTEM / ATISCH	4.040
ERFAHREN / UNG	4.294	KENNEN / TNIS	5.285
DIPLOM	5.504	TECHNIK / ISCH	5.882
UNTERRICHT / EN	7.041	ORGANISATION	8.355
WUNSCH / EN	8.380	ZONE	8.546
BITTE / N	9.429	STELLE	11.708
UNTERNEHMEN / R	14.430	STADT	16.330
GEBIET	17.389	VERBAND	17.569
PERSON / LICH / KEIT	18.983	AUSGABE / GEBEN	19.302
ANBIETEN / GEBOT	20.335	ALLGEMEIN	21.685
ARBEIT / EN	22.182	VERANTWORTEN / UNG	24.320
WERBEN / UNG	25.119	VERKEHR / EN	26.932

SEMANTISCHE UMGEBUNG E(X) NEUES DEUTSCHLAND
X = INDUSTRIE / IEREN

BAUER	5.264	ZIEL / EN	5.539
LANDWIRT / SCHAFT	6.071	AUSSTELLUNG	6.781
BAUER / IN	6.937	BERATEN / UNG	7.072
STUNDE	7.227	QUALITAET	9.619
PROMINENT	10.262	ERFOLG / REICH	10.288
KULTUR / EN	10.457	SPEZIAL / IST	10.638
KONGRESZ	11.024	PLAN	11.033
LEITEN / UNG	11.268	GENOSSE / IN	11.291
BESUCH / EN	11.522	LOHN	11.602
SOWJET / ISCH / UNION	11.628	CHEMIE / SCH	11.923
ARBEIT / EN	12.023	HEBUNG / EN	12.048
BESCHLUSZ / EN	12.056	PRODUKTION	12.159
OEKONOMISCH	12.264	VORSCHLAG	12.509
ARBEITER / IN	12.585	BRIGADE	12.745
SEHEN / SICHT	12.746	WEIHNACHT	12.777

Table 2. Conceptual meaning $M(x) = E(x)$ of the lexical entry INDUSTRIE / IALISIEREN (industry / ialise) from DW and ND

STRUKTUELLE BEDEUTUNG B(X) DIE WELT
X = KOENNEN

NUTZEN / NUETZEN	2.697	GABE / GEBEN	2.699
BEGINN / EN	2.738	HOEREN	2.809
LEBEN	2.813	SEKRETAER	2.831
GEHEN / GANG	2.878	FREUND / SCHAFT	2.904
SAGEN	2.939	EHRE / N	3.061
ABEND	3.166	ZIEL / EN	3.184
VERBINDEN / UNG	3.203	FEST / IGEN / UNG	3.209
GANZ	3.273	ALLE	3.304
LANG / E	3.310	BLEIBEN	3.345
GRUSZ / EN	3.352	GRENZE / N	3.361
ERNTE / N	3.389	GETREIDE	3.444
VERSUCH / EN	3.444	OEFFNEN / UNG	3.449
LIEBE / N	3.462	STEHEN	3.463
OSTEN	3.470	HAUPT	3.487
NEHMEN	3.488	NENNEN / UNG	3.528

STRUKTURELLE BEDEUTUNG B(X) NEUES DEUTSCHLAND
X = KOENNEN

LIEBE / N	2.508	FUEHRUNG / EN	2.592
NUTZEN / NUETZEN	2.680	SCHWER	2.685
MACHEN	2.762	BEWEGEN / UNG	2.801
EUROPA	2.863	GEHEN	2.863
HERR / EN / SCHAFT	2.867	GABE / GEBEN	2.899
HAUPT	2.954	SAGEN	3.002
STELLE	3.023	PUNKT	3.062
FRANKREICH / ZOESISCH	3.084	VERSUCH / EN	3.084
LANG	3.124	BLEIBEN	3.176
PREIS	3.207	ELEKTRO / NISCH	3.272
MUESZEN	3.299	KONTROLLE / IEREN	3.313
OSTEN	3.362	MINISTER	3.368
ATOM / AR	3.432	ARBEITER / EN	3.473
ENDE / N / LICH	3.490	ABKOMMEN	3.523
POLITIK / ER / ISCH	3.540	FRAGE / N	3.540

Table 3. Linguistic description $D(z) = B(x)$ of the lexical entry KOENNEN (able / ability) from DW and ND

SEMANTISCHE UMGEBUNG E(X) DIE WELT
X = KOENNEN

BEGINN / EN	7.809	FREUND / SCHAFT	8.513
GABE / GEBEN	8.855	GANZ	9.116
HOEREN	9.253	LEBEN	9.488
SEKRETAER	10.275	NUTZEN / NUETZEN	10.488
SAGEN	10.525	ALLE	12.020
EHRE / N	12.142	LIEBE / N	12.196
LANG / E	12.561	ZIEL / EN	13.602
GEHEN / GANG	13.618	GRUSZ / EN	14.281
FEST / IGEN / UNG	14.939	HERZ	15.275
MACHEN	15.510	ALT / ER	15.677
DANK / EN	15.731	HAUPT	15.744
DIENST / EN	15.864	TOD	16.102
STEHEN	16.183	ENDE / N / LICH	16.192
PREIS	16.283	RECHT	16.299
NAH / E / NAEHERN	16.304	FRAGE / N	16.347

SEMANTISCHE UMGEBUNG E(X) NEUES DEUTSCHLAND
x = KOENNEN

LIEBE / N	6.667	FUEHRUNG / EN	6.944
BEWEGEN / UNG	7.053	MACHEN	7.110
SCHWER	7.529	HERR / EN / SCHAFT	7.529
GABE / GEBEN	7.813	HAUPT	8.312
FRANKREICH / ZOESISCH	8.454	VERSUCH / EN	8.454
NUTZEN / NUETZEN	8.674	GEHEN	9.118
SAGEN	9.165	MINISTER	9.253
LANG	9.298	MUESZEN	9.395
ALLE	9.569	FEST / IGEN / UNG	9.869
OSTEN	10.401	ATOM / AR	10.472
RECHT	10.711	WELT	10.927
KONTROLLE / lEREN	10.932	PUNKT	11.030
POLITIK / ER / ISCH	11.064	LAND	11.375
GUT / GUETE	11.435	EUROPA	11.615
ENDE / N / LICH	11.645	FRAGE / N	11.704

Table 4. Conceptual meaning $M(x) = E(x)$ of the lexical entry KOENNEN (able / ability) from DW and ND

6. *Examples*

To show the feasibility of the empirical approach and to at least partly invalidate the apprehension of being left empty-handed at the end, the following examples of linguistic descriptions $D(z)$ and of conceptual meanings $M(x)$ may serve as an illustration. They are taken from the data of a pilot-study on semantic differences in lexical structure (Rieger 1980) that has been done within a major project on East-West-German language comparison.

So far, two samples from corpora consisting of texts from the East-German newspaper 'Neues Deutschland' and the West-German newspaper 'Die Welt', both of 1964, have been analysed according to the procedures outlined. The above (11) *linguistic description* $D(z)$ of a concept point z is equivalent to "Strukturelle Bedeutung $B(x)$", and the above (15) *conceptual meaning* $M(x)$ of a vocabulary term x is equivalent to "Semantische Umgebung $E(x)$" in the following print-outs reproduced (Tables 1 to 4). In addition, the distance values given here have not been standardized to the intervall $[0, 1]$ as stated above in (8) and (12); details may be found in Rieger (1980).

Although the samples analysed are rather small — approximately 3000 running words (tokens) of roughly 300 lemmatized words (types) — the results look quite promising to the native speaker of German. A word-word translation into English can hardly be given and were bound to miss the point, because the linguistic descriptions of concept points and the conceptual meanings of vocabulary terms are based upon the very connotational relations of a lexical structure which is idiomatic and as such varies considerably from one language, socio- and / or idiolect to the other. In mapping the connotational differences, however, that some morphologically identical German lexical entries have developed almost simultaneously after twenty years of usage in a devided country's rather strictly separated population, the pilot-study's results seem to indicate that — linguistically — an additional analysis of comparable text-corpora of earlier and / or later years could provide the *diachronic* complement to the so far *synchronic* investigation into the lexical structures concerned, allowing for the empirical reconstruction not only of their instantaneous word-meanings, but of their time-dependent procedural changes (Nowakowska 1980). Being induced by varying language usages, these can operationally be analysed as regularities followed and / or established to differing degrees, which hence may formally be represented as functions that constitute dynamic systems to model semiotic structures.

7. *Acknowledgement*

This paper, an earlier version of which under the title "Probleme der automatischen Wortsemantik" was presented at the joint Annual Meeting of the Association for Literary and Linguistic Computing (ALLC) and the LDV Fittings e.V. at the University of Bonn, Germany, in December 1979, reports on some of the model constructions resulting from a project in Empirical Semantics supported by the Northrhine-Westphalia Ministry of Science and Research (II B 6 – FA 7519), and their application to the language data, provided by the German Research Foundation project on East-West-German language comparison (DFG He 983 / 1 / 2). I would like to thank my colleague Dr. H. M. Dannhauer who kindly provided his programming abilities to make the CDC-Cyber 175 at the Aachen Technical University Computing Center collaborate efficiently.

Gaines, B.R., 1977: “System Identification, Approximation and Complexity”, Intern. Journ. General Systems 3, 145–174

Halliday, M.A.K., 1977: “Text as Semantic Choice in Social Context” in: van Dijk, T.A. / Petöfi, J.S. (Eds.): Grammars and Descriptions, Berlin / New York 1977, 176–225

Labov, W., 1973: “The boundaries of words and their meaning” in: Bailey / Shuy (Eds.): New Ways of Analyzing Variation in English, Washington, 340–373

Nowakowska, Maria, 1980: “Semiotic systems knowledge representation and memory” in: Rieger, B. (Ed.): Empirical Semantics. A Reader of New Approaches in the Field, Bochum (forthcoming)

Rieger, B., 1977a: “Theorie der unscharfen Mengen und empirische Textanalyse” in: Klein, W. (Ed.): Methoden der Textanalyse, Heidelberg 84–99

Rieger, B., 1977b: “Bedeutungskonstitution. Einige Bemerkungen zur semiotischen Problematik eines linguistischen Problems”, Zeitschrift für Literaturwissenschaft und Linguistik, LiLi 27 / 28, 55–68

Rieger, B., 1977c: “Vagheit als Problem der Linguistischen Semantik” in: Sprengel / Bald / Viethen (Eds.): Semantik und Pragmatik, Tübingen, 91–101

Rieger, B., 1978: “Unscharfe Semantik natürlicher Sprache. Zum Problem der Repräsentation und Analyse vager Bedeutungen”, Nova Acta Leopoldina (forthcoming)

Rieger, B., 1979a: “Fuzzy Structural Semantics. On a generative model of vague natural language meaning”, Trappl / Hanika / Pichler (Eds.): Progress in Cybernetics and Systems Research, Vol. V, New York / London / Sydney, 495–503

Rieger, B., 1979b: “Linguistic Semantics and the Problem of Vagueness: on analysing and representing word meaning”, in: Ager / Knowles / Smith (Eds.): Advances in Computer-Aided Literary and Linguistic Research, Birmingham, 271–288

Rieger, B., 1979c: “Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung”, in: Bergenholtz / Schaeder (Eds.): Textcorpora. Materialien für eine empirische Textwissenschaft, Kronberg / Ts., 52–70

Rieger, B., 1979d: “Revolution, Counterrevolution or a New Empirical Approach to Frame Reconstruction instead?” in: Petöfi, J.S. (Ed.): Text vs. Sentence. Basic Questions of Textlinguistics, Vol. II, Hamburg, 555–571

Rieger, B., 1980: “Ein statistisches Verfahren zur lexikalisch-semantischen Beschreibung des in Texten verwendeten Vokabulars im Rahmen eines Strukturmodells unscharfer (fuzzy) Wortbedeutungen”, in: Hellmann, M.W. (Ed.): Ost-West-Wortschatzvergleich. Sprache der Gegenwart, Schriften des Instituts für deutsche Sprache 48, Düsseldorf (forthcoming)

Sparck-Jones, Karen / Kay, M., 1973: Linguistics and Information Science, New York

Zadeh, L.A., 1978: “PRUF — a meaning representation language for natural languages”, Intern. Journ. Man-Machine Studies 10, 395–460