Fuzzy Representation Systems in Linguistic Semantics

An Empirical Approach to the Reconstruction of Lexical Meanings from East- and West-German Newspapers*

Burghard B. Rieger

Arbeitsgruppe f. mathem.-empirische Systemforschung (MESY) German Department

Technical University of Aachen, Germany

Word Semantics is gaining increasing importance within linguistics. Due to the fact that both, formal and operational means have been devised to analyse and represent word connotation and/or denotation adequately, this paper discusses some of the empirical problems connected with natural languages' essentially varying and vague meanings, how these can be analysed statistically from discourse data, and represented formally as fuzzy system of vocabulary mappings.

Some examples computed from East- and West-German newspaper texts will be to illustrate the approach's feasibility.

1 INTRODUCTION

In talking about fuzzy representation systems in linguistic semantics I will confine myself on discussing the question of how lexical meanings may possibly be reconstructed empirically, i.e. analysed and represented. Tackling this problem of word semantics is to be concerned with at least two central aspects of it

a) the specification of the data base to be analysed automatically, and

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

Those of you who happen to remember my paper [8] presented here on EMCSR/3 will probably expect to find some application of fuzzy sets theory to problem-area b) — and they are of course perfectly right in doing so.

But before I go about some of the procedures we have in the meantime developed and results tested in the Technical University of Aachen MESY-group so far, I will have to make some points on the frame-conditions, the basic language material has to satisfy in order to make our procedures work. And this, of course, concerns problem-area a).

*Published in: Trappl, R./Findler, N.V./Horn, W. (eds.): Progress in Cybernetics and Systems Research, Vol. XI, Washington/New York/London (McGraw-Hill Intern.) 1982, pp. 249–256.

As these issues have been discussed at some length elsewhere [5], [6], [7], [9], I shall only be refering to them here. However, as problems of word-semantics should be discussed where and when they come up, I would like to give an account of the philosophy (so to speak) behind my approach. I therefore will have to spend some time on aspects of formal and descriptive theory construction and the empirical complications to be expected in view of a semiotic domain like word-semantics.

Let me start with an introductory quotation which strikingly characterizes the situation: "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." This quotation from Sparck-Jones/Kay [14] will hopefully stimulate your expectations (if necessary), or (if appropriate) will let you be prepared to be left empty-handed at the end.

2 AIMS OF WORD SEMANTICS

According to Moskovich [3] 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 semantic-descriptors. Thus, *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 descriptions.

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 be obtained.

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

3 STATUS OF WORD SEMANTICS

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

a) not to rely on ready-made theories or models taken from another domain, because these may be grossly inadequate, in respect to the phenomena, and

b) not to rely on the introspective exploration of one's own knowledge and competence as the allegedly inexhaustible data-source, although it may produce valuable initial ideas.

Instead, the investigation of linguistic problems in general, and that of word-semantics in particular, should start with hypotheses, formulated and reformulated for continuous estimation and/or testing against observable data, then proceed to incorporate the 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 [5]. According to Gaines [1] their descriptive and/or formal framework is designed to fit the phenomena, not to straiten the phenomena to fit a model or theory.

Following the line of Labov [2] 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 welldefined 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 reconstruction even of semiotic phenomena.

My 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

Referential Approach	Model construction
Denotation $:= L \longrightarrow W$	Den: $L \times W \longrightarrow M_D$
Structural Approach	Model construction
Connotation $:= L \longrightarrow L$	Con: $L \times L \longrightarrow M_C$

Figure 1

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.

4 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 focussed on.

According to the most influential of them, natural language meaning can be characterized by its denotative and its connotative aspects (Fig. 1). *Denotation* is understood to constitute *referential* meaning as a system M_D of relations between words or sentences of a language L and the object or processes they refer to in W. *Connotation* is defined to constitute *structural* meaning as a system M_C by which words or sentences of a language L are conceptually related to one another. Referential semantic theory is truth-functionally and formally elaborated but as such not prepared to account satisfactorily for the vagueness of natural language meaning; whereas structural semantics has considered vagueness somewhat fundamental of 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. the body of natural language discourse) and that which it represents (i.e. referential or structural meaning constituted by this body).

5 ZADEH'S APPROACH

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

The most recent, and at that most comprehensive formal approach (at least I know of) to tackle the problem of natural language meaning, is that of L.A. Zadeh [15]. 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 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

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

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

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

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

a) 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 Φ_L conditioned on x',

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

b) 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 Φ_L conditioned on z',

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

Although formally satisfactory — as outlined and illustrated by PRUF — the basic assumption of the approach concerning the referential nature of natural 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 Φ_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 Φ_L. All that Zadeh [15] 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 appear more appropriate 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 produce, may thus be analysed statistically with the numerical values obtained to define fuzzy vocabulary mappings [10].

6 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 reinterpretation of what the sets T and U (1) in the referential meaning relation possibly stand for.

From a structural point-of-view, T is not just a set of terms of a language any more, but 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 them as 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 [9].

Thus, being a non-symmetric, fuzzy, binary relation, Φ_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 U, providing the membership-grades for Φ_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

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

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

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

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

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

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 are 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 corpuspoints $y,\ y' \in C$. δ_1 can also be considered a fuzzy, binary relation in the set of all corpus-points y_i defined to constitute the corpus space

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

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

$$\delta_1 \parallel x_i : C \to \boldsymbol{I}^n; \quad U := \boldsymbol{I}^n \tag{9}$$

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

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

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), i.e. 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 set of all concept-points z_i defined to constitute the semantic space U

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

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

$$\delta_2 \parallel z_i : U \rightarrow \mathrm{I}^n \tag{13}$$

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

$$\Phi_L : T \times U \rightarrow [0, 1] \tag{16}$$

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 Φ_L according to (9) inserting

$$\delta_1 \parallel x_i = \Phi_L(x_i, z_i) = \big\{D(z)\big\} \subseteq T \tag{17}$$

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

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

This concludes the empirical reconstruction, leaving open only the coefficients alluded to above.

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

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

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

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

where

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

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

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

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

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

and

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

As these distance measures are to be considered the metric of the corpus 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 word-semantic systems' structures.

7 EXAMPLES

To show the feasibility of the empirical approach and to leave you not completely empty-handed at the end, the following examples of linguistic description $D(z)$ 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 [11] 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' have been analysed according to the procedures outlined. 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. In mapping the connotational difference which 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 that Nowakowska [4] aims at. Being induced by

Table 1

Linguistic description $D(z)$, z =ELEKTRO/NISCH — DIE WELT

FAEHIG/KEIT	.701	INDUSTRIE/IEREN	.704
TECHNIK/ISCH	1.083	ORGANISATION	1.257
STADT	1.837	GEBIET	1.851
STELLE	1.899	VERANTWORTEN/NG	2.113
WIRTSCHAFT	2.970	ARBEIT/EN	3.011
HERRSCHAFT/EN	3.097	ENTWICKELN/UNG	3.371
FOLGE/N/UNG	3.409	SPRACHE/N	3.877
MOEGLICH/KEIT	3.987	JUNG/END	4.200
NEU	4.209	ZENTRAL/UM	4.513
GUT	4.585	FUEHREN/ER/UNG	4.650
HOCH/HOEHE	4.661	FRANKREICH/ISCH	4.843
HAUS	5.172	FORDERN/UNG	5.230
BAU/EN	5.233	PLATZ	5.373
HANDEL/N	5.545	KRAFT	5.613
STRASZE	5.852	WISSENSCHAFT	5.882

Conceptual meaning $M(x)$, x =ELEKTRO/NISCH — DIE WELT

LEITEN/R/UNG	3.001	FAEHIG/KEIT	4.730
TECHNIK/ISCH	6.220	ORGANISATION	7.327
STELLE	10.891	STADT	15.541
GEBIET	16.578	ARBEIT/EN	21.428
VERANTWORTEN/NG	23.494	HERRSCHAFT/EN	30.038
WIRTSCHAFT	35.258	FOLGE/N/UNG	35.427
ENTWICKELN/UNG	39.070	GUT	43.283
JUNG/END	44.809	NEU	44.822
ZENTRAL/UM	45.256	SPRACHE/N	46.399
MOEGLICH/KEIT	46.626	PLAN/EN	50.211
HOCH/HOEHE	51.107	MUESZEN	51.651
FRANKREICH/ISCH	52.979	BRAUCHEN	53.932
FUND/EN	54.943	BEREIT/EN	55.255
KUNST/LERISCH	55.276	PUNKT	55.434
FUEHREN/ER/UNG	55.945	STRASZE	57.334

Table 2

Linguistic description $D(z)$, z =ELEKTRO/NISCH — NEUES DEUTSCHLAND

SCHWER	2.249	BAU/EN	2.341
PREIS	2.555	GEHEN	2.670
FUEHRUNG/EN	2.710	FORDERN/UNG	2.726
GUT	2.791	ARBEITER/EN	2.848
FRANKREICH/ISCH	2.876	VERSUCH/EN	2.876
TOD	2.900	ALLE	2.915
NAH/E/RN	2.990	FAEHIG/KEIT	2.996
HERR/EN/SCHAFT	3.017	AFRIKA/NER/ISCH	3.027
BLEIBEN	3.067	SOLDAT	3.075
OSTEN	3.080	ARBEITER/IN	3.089
BEWEGEN/UNG	3.119	LIEBE/N	3.138
ABKOMMEN	3.148	DIENST/EN	3.154
STELLE	3,160	HAUPT	3.162
ENDE/N/LICH	3.174	HOCH/HOEHE	3.185
REGIEREN/UNG	3.186	MUESZEN	3.191

Conceptual meaning $M(x)$, x =ELEKTRO/NISCH — NEUES DEUTSCHLAND

SCHWER	7.802	FORDERN/UNG	7.971
PREIS	8.266	BAU/EN	8.399
TOD	8.501	ARBEITER/EN	8.818
ABKOMMEN	8.924	AFRIKA/NER/ISCH	9.040
REGIEREN/UNG	9.149	ALLE	9.255
HANDEL/N	9.491	DIENST/EN	9.631
SOLDAT	9.658	GEHEN	9.778
VERSUCH/EN	9.779	FRANKREICH/ISCH	9.779
STELLE	9.862	MUESZEN	9.881
PUNKT	9.980	ENDE/N/LICH	10.039
BLEIBEN	10.104	GUT	10.143
WASSER	10.200	JAHR/IG/LICH	10.335
HAUPT	10.359	PRAESIDENT	10.431
ATOM/AR	10.477	ARBEITER/IN	10.555
PLATZ	10.649	FAHRT/EN	10.719

Linguistic Description $D(z)$ and Conceptual Meaning $M(x)$ of the lemmatized lexical entry ELEKTRO/NISCH (electro/nic) as employed in texts of German from the newspapers DIE WELT (West) and NEUES DEUTSCHLAND (East) calculated according to (17) and (18); the values listed behind the descriptors, however, have not been standardized to the unit interval.

varying language usages, these can operationally be analysed as regularities followed and/or established by language users to differing degrees, which hence may formally be represented as functions that constitute dynamic systems to model semiotic structures.

In the above Tables 1 and 2 the *linguistic description* $D(z)$ of a concept point z is given as well as the *conceptual meaning* $M(x)$ of a vocabulary term x from both of the newspaper corpora further details of which may be found in Rieger [11].

ACKNOWLEDGEMENT

This paper parts of which were presented in German under the title "Probleme der automatischen Wortsemantik" 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, is an abbreviated version of Rieger [12]. It draws from a pilot-study [11] which took up some model construction resulting from a project in Empirical Semantics supported by the Northrhine-Westphalia Ministry of Science and Research (II B 6 – FA 7519) applied to the language data provided by the German Research Foundation's project on East-West-German language comparison (DFG He 983/1/2).

I would like to thank Dr. H.M. Dannhauer for providing the necessary programming to process these language data at the Technical University of Aachen Computing Center.

REFERENCES

1. GAINES, B.R.: "System Identification, Approximation, and Complexity", *Intern. Journ. General Systems* 3 (1977) 145–174
2. LABOV, W.: "The boundaries of words and their meaning" in: Bailey/Shuy (Eds.): New Ways of Analyzing Variation in English, Washington 1973, 340–373
3. MOSKOVICH, W.: "Perspective Paper: Quantitative Linguistics" in: Walker/Karlgren/Kay (Eds.): Natural Language in Information Science, Stockholm 1977, 57–74
4. NOWAKOWSKA, Maria: "Semiotic Systems, knowledge representation, and memory" in: Rieger, B. (Ed.): Empirical Semantics. A Collection of New Approaches in the Field, Bochum 1980 (forthcoming)
5. RIEGER, B.: "Bedeutungskonstitution. Einige Bemerkungen zur semiotischen Problematik eines linguistischen Problems", *Zeitschrift für Literaturwissenschaft und Linguistik, LiLi* 27/28 (1977) 55–68
6. RIEGER, B.: "Vagheit als Problem der Linguistischen Semantik" in: Sprengel/Bald/Viethen (Eds.): Semantik und Pragmatik, Tübingen 1977, 91–101
7. RIEGER, B.: "Unscharfe Semantik natürlicher Sprache. Zum Problem der Repräsentation und Analyse vager Bedeutungen", *Nova Acta Leopoldina* 1978 (in print)

8. RIEGER, B.: “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 1979, 495–503
9. RIEGER, B.: “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 1979, 555–571
10. RIEGER, B.: “Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung”, in: Bergenholtz/Schaeder (Hrsg.): Textcorpora. Materialien für eine empirische Textwissenschaft, Kronberg/Ts 1979, 52–70
11. RIEGER, B.: “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 1980 (forthcoming)
12. RIEGER, B.: “Feasible Fuzzy Semantics. On some problems of how to handle word meaning empirically” in: Eikmeyer/Rieser (Eds.): New Approaches in Word Semantics, Berlin/New York 1980 (forthcoming)
13. RIEGER, B.: (Ed.): Empirical Semantics. A Collection of New Approaches in the Field, Bochum 1980 (forthcoming)
14. SPARCK-JONES, Karen/Kay, M.: Linguistics and Information Science, New York 1973, p. 120
15. ZADEH, L.A.: “PRUF — a meaning representation language for natural languages”, *Intern. Journ. Man-Machine Studies* 10 (1978) 395–460