

Fuzzy Word Meaning Analysis and Representation in Linguistic Semantics

An Empirical Approach to the Reconstruction of Lexical Meanings in East- and West-German Newspaper Texts*

Burghard B. Rieger

Arbeitsgruppe f. mathem.-empirische Systemforschung (MESY)

German Department, Technical University of Aachen, Germany

Summary

Word semantics is gaining increasing interest within linguistics in view of both, more adequate representational structures of the semantic system and methods and procedures to analyse it empirically. Due to the fact that formal and operational means have been devised to describe and represent word connotation and/or denotation, this paper discusses some of the empirical problems connected with natural languages, 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 given at the end to illustrate the approach's feasibility.

1 Introduction

When we look up linguistic theories of sentence- or even of text-semantics to see what they can offer in respect to word-meaning, we will be confronted with basically two types FILLMORE [3] has referred to as *checklist-semantics* and *prototype-semantics*. According to this distinction, *checklist-semantics* provides listings of meaning components, semantic markers, or semantic descriptors which must be satisfied for a term to be (grammatically, truth-functionally, or else) interpretable within a linguistic expression; whereas *prototype-semantics* allows for the (paradigmatical, syntagmatical, or else) identification of a term as part of a linguistic expression within a network structure of labeled nodes and relations. Examining how these listings and networks are assembled, i.e. questioning from which sources and by what procedures the data necessary for their composition were acquired, we will invariably

*Published in: COLING 80. Proceedings of the 8th International Conference on Computational Linguistics, Tokyo (ICCL) 1980, pp. 76–84.

come across the individual analysts', or group of analysts' own assumedly comprehensive and reliable knowledge of the world and/or the natural language system concerned. In the majority of cases, these will not have been made accessible by intersubjectively defined operations but rather by way of intuitive introspection. In doing so, linguists tend to *make use* of word-meaning instead of *analysing* it when they set up matrices for componential analysis or define semantic networks. 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 natural language discourse — a lexical structure may be obtained.

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

2 Epistemology

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 inadequate in respect to the phenomena under investigation; and
- b) not to rely on the introspective exploration of one's own knowledge and competence as the allegedly inexhaustible datasource although valuable initial ideas might be produced that way.

Instead, the investigation of linguistic problems in general, and that of word-semantics in particular, should start with hypotheses formulated 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 theory.

Within such a set-up, the formal expressions which give an abstract representation of the domain, and the numerical expressions which give a quantitative account of the observable data, are normally to be complemented by correspondence rules. These allow for the operational interpretation of formal notations and theoretical constructs in terms of empirical methods of counting and measuring observable data. Linguistic theory has not been interested too much in developing correspondence rules of that kind so far [15].

Following the line of LABOV [9] and LEECH [10], prevailing linguistic theory and linguistic semantics in particular is dominated by what has been called the “categorical view”. According to it, linguistic entities are at least implicitly asserted to be discrete, invariant, qualitatively distinct, conjunctively definable, and composed of atomic primes. Membership in categories, and relations of inclusion and exclusion

among units and categories, are established by a deterministic type of rule that allows only for binary (positive or negative) or triple (positive, negative, or optional) assignment, but has no means to represent probable and/or possible degrees of transition. This type of rule — particularly when employed for meaning representation purposes — has come under severe criticism from as seemingly disparate disciplines like cognitive theory (e.g. [12]) and experimental psychology (e.g. [20]), information and computer science (e.g. [23], [4]), psycholinguistics (e.g. [16], [11]) sociolinguistics (e.g. [8], [7]), computational semantics (e.g. [14]) and artificial intelligence (e.g. [22], [5]).

From the increasing amount of strong empirical evidence piling up in favour of some re-adjustment, a (meta-theoretical) modification appears to be overdue. Accordingly, it may be argued that — contrary to the experimentally and simulatively well established (object-theoretical) fuzziness of cognitive categorizing and its linguistic correspondences — any formal representation of it using only binary systems' notations will inevitably result in inadequately sharp-edged lattices. When imposed upon the varying and vague structures constituted and modified continuously during the process of verbal communication observed to be modelled, this will render formal representations of discrete entities with clear-cut boundaries where blurred margins and continuous transitions would be adequate.

The modifications suggested so far may be summarized to concern both, the observable manifestation and/or formal representation of discourse, allowing gradual rather than abrupt transitions to account for imprecise phenomena in a precise way. This can be achieved, as I see it, *formally* by means of fuzzy set theoretical notations [24], and *operationally* by means of empirical procedures assigned to them [19]. Applied to natural language data, they will interrelate observable but essentially *fuzzy* language phenomena on the one hand, and formal but finally *categorical* 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

What makes the analysis of natural language meaning so intricate a problem depends on the particular nature of what has to be represented as its results, namely, a representational structure in its own. It is this representational aspect of language which theories of semantics and cognition have been, and still are focussed on in particular.

According to the more traditional theories, natural language meaning can be characterized by its denotative and connotative aspects. *Denotation* is understood to constitute referential meaning as a system of relations between words or sentences of a language and the objects or processes they refer to. *Connotation* is defined to constitute structural meanings as a system by which words or sentences of a language are conceptually related to one another. Referential semantic theory

is truth-functional 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.

In the course of recent, more procedural approaches to cognition and language comprehension, the former distinction of referential and structural meaning was embedded in what became to be known as frame semantics [17]. The central notion of it is that of memory which serves as a paradigm for the operational aspects of both, world system structures and language system structures. The basic distinction of what may propositionally be formulated as opposed to what may only prototypically be realized in some system structure of stored experiences, is reflected in the great variety of notional pairings which different disciplines have produced facing a similar, if not identical research problem. Thus, their notions of *formal* vs. *experiential knowledge* [2], *semantic* vs. *episodic memory* [21], *frame* vs. *scene* [3], *description* vs. *schema* [1], etc. show a striking resemblance: although their 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 discourse) and that which it represents (i.e. a referentially and/or prototypically defined system structure).

4 A formal 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 formal theories of semantics.

The most recent, and at that most comprehensive approach (at least I know of) to tackle the problem of natural language meaning, is that of L.A. Zadeh [24]. 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 labelled 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

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

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

$$L := \left\{ ((x, z), \mu_L(x, z)) \right\} \quad (2)$$

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 μ_L conditioned on x' ,

$$M(x') := \mu_L(z, x') := \left\langle ((x', z_1), \mu_L(x', z_1)), \dots, ((x', z_n), \mu_L(x', z_n)) \right\rangle \quad (3)$$

- 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 μ_L conditioned on z'

$$D(z') := \mu_L(x, z') := \left\langle ((x_1, z'), \mu_L(x_1, z')), \dots, ((x_n, z'), \mu_L(x_n, z')) \right\rangle \quad (4)$$

The definitions given in fuzzy sets theory for equality, containment, complement, intersection, and union allow for an application both, to *referential meanings* $M(x)$ as subsets of elements in U and to *linguistic descriptions* $D(z)$ as subsets of units in T . This corresponds to the distinction between scenic, or conceptual relations on the one hand, and frame, or semantic relations on the other — the latter of which only will be introduced here.

Thus, *synonymy* of two terms $x, x' \in T$ may be given as the equality of the two fuzzy subsets $M(x)$ and $M(x')$ representing the referential meaning in U

$$x = x' \quad \text{iff} \quad \mu_L(z, x) = \mu_L(z, x') \quad \text{for all } z \quad (5)$$

Partial synonymy may be defined by a similarity formula introducing some threshold-value s

$$x \approx x' \quad \text{iff} \quad |\mu_L(z, x) - \mu_L(z, x')| \leq s \quad \text{for all } z \quad (6)$$

Hyponymy of a term x relative to x' may be explicated as containment of the meaning representing fuzzy sets concerned

$$x \supset x' \quad \text{iff} \quad \mu_L(z, x) \geq \mu_L(z, x') \quad \text{for all } z \quad (7)$$

In so far as the operations of complement, intersection and union are concerned which correspond to negation, conjunction and adjunction respectively, there has been some critical discussion lately, particularly on the grounds of experimental results. These suggest that different definitions of operations should be maintained according to and comparable with the scene-frame-distinction aluded to above. For the generation of new meanings which denote possible but not yet labeled elements (or sets of elements) in U , it can well be argued that the following definitions should operate on both, referential meanings $M(x)$ and linguistic descriptions $D(z)$ the former of which only are given here.

Negation (complement):

$$\neg x := M(x) = 1 - \mu_L(z, x) \quad \text{for all } z \quad (8)$$

Conjunction (intersection):

$$x \wedge x' := M(x \cap x') = \min [\mu_L(z, x), \mu_L(z, x')] \quad \text{for all } z \quad (9)$$

Adjunction (union):

$$x \vee x' := M(x \cup x') = \max [\mu_L(z, x), \mu_L(z, x')] \quad \text{for all } z \quad (10)$$

Although formally satisfactory — as outlined and illustrated by PRUF — the approach’s basic assumption 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 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 [24] 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”.

This cannot be considered a solution which may be called both adequate and operational in the above 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 use the basis for identifying those language regularities, which under certain communication frame conditions real speakers/hearers follow and/or establish in discourse. These will consequently allow natural language meaning (whatever that may be) not only to be intended and understood, but also to be analysed and represented. As this apparently is 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 pragmatically homogeneous corpus, the usage regularities which the lexical items produce, may thus be analysed statistically with the numerical values obtained to define fuzzy vocabulary mappings [16].

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, 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 of its own. So, 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 [17].

Thus, being a non-symmetric, fuzzy, binary 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 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 Cartesian-product of the set of terms T used in the texts analysed

$$\alpha : T \times T \rightarrow \mathbf{I}, \quad \mathbf{I} = [-1, +1]; \quad T := \{x_i\}, \quad i = 1, \dots, n \quad (11)$$

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

$$\alpha|x_i : T \rightarrow \mathbf{I}^n, \quad C := \mathbf{I}^n \quad (12)$$

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

$$\alpha(x_i, x) := y \in C \quad (13)$$

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 corpus-points $y, y' \in C$. 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 C

$$\delta_1 : C \times C \rightarrow I; \quad I := [0, 1]; \quad C := \{y_i\}, \quad i = 1, \dots, n \quad (14)$$

By conditioning this fuzzy relation δ_1 on the y_i (or — following (13) — the x_i respectively) we get a non-fuzzy mapping

$$\delta_1|x_i : C \rightarrow I^n; \quad U := I^n \quad (15)$$

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 ,

$$\delta_1(y_i, x) = \delta_1(y_i, y) := z \in U \quad (16)$$

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 (13) — with (4), i.e. the *linguistic description*, $D(z')$, of a concept-point z' which is a fuzzy subset in T

$$\delta_1(x_i, z') := D(z') \subset T \quad (17)$$

On the *third* stage of the consecutive mapping, there will topological environments of concept-points be calculated — in analogy to (14) — 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

$$\delta_2 : U \times U \rightarrow I; \quad I := [0, 1]; \quad U := \{z_i\}, \quad i = 1, \dots, n \quad (18)$$

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

$$\delta_2|_{z_i} : U \rightarrow I^n \quad (19)$$

which assigns to each $z \in U$ (and — following (16) — $x \in T$ respectively) one and only one n -tuple of distances that — scaled according to decreasing values — will constitute the environment $E(z)$

$$\delta_2(z_i, x) = \delta_2(z_i, z) := E(z) \quad (20)$$

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 (16) and (20) — with (3) the *conceptual meaning*, $M(x')$, of a term x' which is a fuzzy subset in U

$$\delta_2(z_i, x') := M(x') \subset U \quad (21)$$

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

$$\mu_L : T \times U \rightarrow [0, 1] \quad (22)$$

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

The two distance measures δ_1 (14) and δ_2 (18), operating on numerical data obtained from the correlational analysis (11) of lexical items employed in a corpus of natural language texts, will determine the membership-grades to be associated with (22), namely for the correspondence (4) induced by μ_L according to (15) inserting

$$\delta_1|x_i := \mu_L(x_i, z_i) = \{D(z)\} \subseteq T \quad (23)$$

and for its inversion the correspondence (3) according to (19) inserting

$$\delta_2|z_i := \mu_{L^{-1}}(x_i, z_i) = \{M(x)\} \subseteq U \quad (24)$$

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

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

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

$$K := \{t\}; \quad t = 1, \dots, m$$

where

$$S = \sum_{t=1}^m s_t; \quad 1_t \leq s_t \leq S \quad (25)$$

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

$$H = \sum_{t=1}^m h_t; \quad 1_t \leq h_t \leq H \quad (26)$$

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 (11) reads

$$\alpha(x, x') = \frac{\sum_{t=1}^m (h_t - h_t^*)(h'_t - h'_t^*)}{\left(\sum_{t=1}^m (h_t - h_t^*)^2 \sum_{t=1}^m (h'_t - h'_t^*)^2 \right)^{\frac{1}{2}}}; \quad -1 \leq \alpha(x, x') \leq +1 \quad (27)$$

$$\text{with } h_t^* = \frac{H}{S} s_t \quad \text{and} \quad h'_t{}^* = \frac{H'}{S} s_t. \quad (28)$$

The distances have been calculated according to the following measures which for δ_1 (14) reads

$$\delta_1(y, y') = \frac{1}{n} \sum_{i=1}^n \frac{(\alpha(x, x_i) - \alpha(x', x_i))^2}{\alpha(x, x_i)^2 - \alpha(x', x_i)^2}; \quad 0 \leq \delta_1(y, y') \leq 2 \quad (29)$$

and δ_2 (18) reads

$$\delta_2(z, z') = \left(\sum_{i=1}^n (\delta_1(y, y_i) - \delta_1(y', y_i))^2 \right)^{\frac{1}{2}}; \quad 0 \leq \delta_2(z, z') \leq 2\sqrt{n} \quad (30)$$

As these distance measures satisfying the conditions are to be considered the metric 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 (30) is nothing but a first (although operational) guess. Experiments with different distance measures one of which is (29) are currently undertaken. Eventually, these might prove to be more adequate one day in modelling word-semantic systems' structures.

CONCEPTUAL MEANING $M(x)$ $x = \text{EUROPA/ISCH}$	DIE WELT	LINGUISTIC DESCRIPTION $D(z)$ $z = \text{EUROPA/ISCH}$	DIE WELT
VERTRAG	.469	GEMEINSCHAFT/CH	.507
VOLK	.540	GEWOHNHEIT/LICH	.541
MINISTER/IUM	.546	DEUTSCH/LAND	.548
TEILNAHME/N	.554	STAAT	.558
KRIEG/ERISCH	.578	BLUT/EN	.585
AMERIKA/ER/ISCH	.587	BESUCH	.588
ERKLAEREN/UNG	.589	HOLLAND/ER/ISCH	.594
FRAKTION	.594	HOFFEN/UNG	.595
REGIEREN/UNG	.598	SOWJET/ISCH/ION	.598
PRAESIDENT	.615	WELT	.625
FRIEDE/LICH	.629	DELEGATION/ER	.629
ANTWORT/EN	.631	BUND	.633
KOMMUNIST/MUS	.634	ATLANTIK/SCH	.635
NATION/AL	.636	CHINA/SISCH	.639
MACHT/IG	.639	VERSUCH/EN	.641
CONCEPTUAL MEANING $M(x)$ $x = \text{EUROPA/ISCH}$		NEUES DEUTSCHLAND	
SPALT/EN	.136	ANTI	.165
MILITARIST/MUS	.197	VERSUCH/EN	.204
FRANKREICH/ISCH	.204	HERR/EN/SCHAFT	.254
ABSICHT	.261	REVANCHE/IST	.279
BEWEGEN/UNG	.309	LIEBE/N	.326
SCHWER	.341	STREBEN/UNG	.354
OSTEN	.373	DOKTRIN	.381
SAGEN	.392	SPORT/LER	.392
WERK/ER	.402	WATT	.402
DYNAMO	.402	WALZE	.402
BLECH	.402	HOEREN	.408
FUEHRUNG/EN	.416	MINISTER	.427
ABGEORDNETER	.432	RAD	.434
NATION/AL	.438	FAHREN/LR/T	.442
HALBJAHR	.447	GANZ	.452
CONCEPTUAL MEANING $M(x)$ $x = \text{EUROPA/ISCH}$		NEUES DEUTSCHLAND	
SPALT/EN	.955	ANTI	1.049
MILITARIST/MUS	1.510	REVANCHE/IST	2.026
FUEHRUNG/EN	2.049	SAGEN	2.148
NATION/AL	2.175	STREBEN/UNG	2.494
GANZ	2.629	SCHWER	2.646
HERR/EN/SCHAFT	2.718	BEWEGEN/UNG	2.753
OSTEN	2.770	VERSUCH/EN	2.818
FRANKREICH/ISCH	2.818	LIEBE/N	2.842
KOENNEN	2.863	MINISTER	2.893
HAND	2.953	GEHEN	3.018
HOEREN	3.077	GABE/EN	3.204
ELEKTRO/NISCH	3.266	ABSICHT	3.358
HAUPT	3.464	TRITT/EN	3.441
LANG	3.464	POLITIK/ER/ISCH	3.530
MACHEN	3.541	OEKONOMISCH	3.592

Table 1

Conceptual Meaning $M(x)$ and Linguistic Description $D(z)$ of EUROPA/ISCH as employed in the newspapers DIE WELT and NEUES DEUTSCHLAND, calculated according to (29) and (30).

CONCEPTUAL MEANING $M(x \wedge x')$		LINGUISTIC DESCRIPTION $D(z)$		DIE WELT
$x = \text{SKI}, x' = \text{ABFAHRT/EN}$	$x = \text{SKI}, x' = \text{ABFAHRT/EN}$	$x = \text{SKI}, x' = \text{ABFAHRT/EN}$	$x = \text{SKI}, x' = \text{ABFAHRT/EN}$	DIE WELT
SKI	.248 ABFAHRT/EN	SKI	.598 ABFAHRT/EN	.598
LIFT	.305 PISTE	BERG	1.189 URLAUB	1.259
TOUR	.371 ALPEN	LIFT	1.303 TOUR	1.358
AUTO	.431 FAHREN/T/ER	SPORT/LER	1.427 PISTE	1.435
BERG	.441 URLAUB	RETTEN/UNG	1.497 ALPEN	1.606
SPORT/LER	.456 RETTEN/UNG	AUTO	1.891 BAHN	2.011
BAHN	.553 LAUF/EN	FAHREN/T/ER	2.089 ORT	2.103
DOPPEL	.573 LUFT	STUECK	2.104 LAUT/EN	2.131
STRASZE	.639 GAST	LUFT	2.321 DOPPEL	2.415
STUECK	.648 LAUT/EN	MARKT	2.526 STAMN/EN	2.562
ORT	.672 MUSIK/ALISCH/R	TOD	2.571 HERZ	2.603
JAZZ	.686 TOD	HOLLAND/ER/ISCH	2.608 MUSIK/ALISCH/R	2.610
KUR	.720 MARKT	PREIS	2.641 UEBERZEUGEN/UNG	2.642
STUNDE	.727 BAD/N	JAZZ	2.653 ALT/ER	2.655
ARZT/LICH	.731 HOLLAND/ER/ISCH	ALLE	2.660 STRASZE	2.663

Table 2

Conjunction of the Conceptual Meanings of SKI and of ABFAHRT/EN, $M(x \wedge x')$, and the resulting concept point's Linguistic Description $D(z|z = x \wedge x')$.

6 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 [18] 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 divided 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 [13] aims at. Being induced by 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 [18].

Acknowledgement

This paper, an earlier version of which was presented under the title “Fuzzy Representation Systems in Linguistic Semantics” at the 8th European Meeting on Cybernetics and Systems Research (EMCSR/8) in Vienna, Austria, in April 1980, is in some parts identical with [19]. It takes up the model construction resulting from a project in Empirical Semantics supported by the Northrhine-Westphalia Ministry of Science and Research, applied to the language data provided by the German Research Foundation’s project on East-West-German language comparison. I would like to thank Dr. H.M. Dannhauer for providing his programming abilities to process these language data so efficiently at the Technical University of Aachen Computing Centre.

References

- [1] BOBROW, D.G./NORMAN, D.A.: “Some Principles of Memory Schemata” in: Bobrow/Collins (Eds): Representation and Understanding, New York 1975, 131–149
- [2] CHAFE, W.L.: “Language and Memory”, *Language* 49 (1973), 261–281
- [3] FILLMORE, C.J.: “Scenes-and-frames semantics” in: Zampolli, A. (Ed): Linguistic Structures Processing, Amsterdam 1977, 55–81
- [4] GAINES, B.R.: “System Identification, Approximation, and Complexity”, *Intern. Journ. General Systems* 3 (1977), 145–174
- [5] GAINES, B.R./SHAW, M.L.G.: “Exploring Personal Semantic Space” in: Rieger, B.(Ed): Empirical Semantics. A Collection of New Approaches in the Field, Bochum 1981 (forthcoming)
- [6] KINTSCH, W: The Representation of Meaning in Memory, Hillsdale, N.J., 1974
- [7] KLEIN, W: “Einige wesentliche Eigenschaften natürlicher Sprache und ihre Bedeutung für die linguistische Theorie”, *Zeitschr. f. Literaturwissenschaft und Linguistik (LiLi)*, 23/24 (1976), 11–31
- [8] LABOV, W.: “The Study of Language in its Social Context”, *Studium Generale* 23 (1970), 30–87
- [9] LABOV, W.: “The Boundaries of Words and their Meanings” in: Bailey/Shuy (Eds): New Ways of Analysing Variation in English, Washington, D.C., 1973, 340–373
- [10] LEECH, G.N.: “Being precise about lexical vagueness”, *York Papers in Linguistics* 6 (1976), 149–165
- [11] LEVELT, W.J.M./et al.: “Struktur und Gebrauch von Bewegungsverben”, *Zeitschrift f. Literaturwiss. u. Linguistik (LiLi)*, 23/24 (1976), 131–174
- [12] MILLER, G.A./JOHNSON-LAIRD, P.N.: Language and Perception, Cambridge, U.K. 1976
- [13] NOWAKOWSKA, M.: “Semiotic Systems, Knowledge Representation, and Memory” in: Rieger, B. (Ed) : Empirical Semantics, Bochum 1981 (forthcoming)
- [14] RIEGER, B.: “Eine tolerante Lexikonstruktur. Zur Abbildung natürlichsprachlicher Bedeutung auf unscharfe Mengen in Toleranzräumen”, *Zeitschr. f. Literaturwiss. u. Linguistik (LiLi)*, 16 (1974) , 31–47

- [15] RIEGER, B.: “Bedeutungskonstitution: Bemerkungen zur semiotischen Problematik eines linguistischen Problems”, *Zeitschr. f. Literaturwiss. u. Linguistik* (LiLi), 27/28 (1977), 55–68
- [16] RIEGER, B.: “Repräsentativität: von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung” in: Bergenholtz/Schaeder (Eds): *Textcorpora. Materialien für eine empirische Textwissenschaft*, Kronberg, Ts., 1979, 52–70
- [17] RIEGER, B.: “Revolution, counter-revolution, or a new empirical approach to frame reconstruction instead?” in: Petöfi, J.S. (Ed): *Text vs. Sentence. Basic Questions of Text Linguistics*, Hamburg 1979, 555–571
- [18] 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*, Düsseldorf 1980 (forthcoming)
- [19] 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)
- [20] ROSCH, E./MERVIS, C.: “Family resemblances: Studies in the internal structure of categories”, *Cognitive Psychology* 7 (1975), 573–605
- [21] TULVING, E.: “Semantic and episodic memory” in: Tulving/Donaldson (Eds): *Organisation of Memory*, New York, '72
- [22] WAHLSTER, W.: Die Repräsentation von vagem Wissen in natürlichsprachlichen Systemen der künstlichen Intelligenz, University of Hamburg 1977, Iff-Report HH-B-38/77
- [23] ZADEH, L.A.: “A fuzzy-algorithmic approach to the definition of complex or imprecise concepts”, *Intern. Journ. Man-Machine Studies* 8 (1976), 249–291
- [24] ZADEH, L.A.: “PRUF — a meaning representation language for natural languages. An up-dated version in: Rieger, B. (Ed): *Empirical Semantics*, Bochum 1981 (forthcoming)