

Inducing a Relevance Relation in a Distance-like Data Structure of Fuzzy Word Meaning Representation*

Burghard B. Rieger

Modelling representational systems for word meanings and/or world knowledge is a problem of mutual and complex relatedness. Different formats have been used with differing success among which that of stereotypical and/or prototypical meaning and knowledge representation appeared to be most adequate in view of how conceptual knowledge is made use of and/or new concepts are being conveyed. Under the notion of lexical *relevance* and semantic *disposition* this interdependency may operationally be clarified and empirically be reconstructed from natural language discourse — although most approaches to word semantics and conceptual modelling do not address these issues. Instead, linguists and psychologists, as well as artificial intelligence experts engaged in word meaning and/or world knowledge representation still provide the necessary semantic and external world information introspectively, i.e. they are exploring (or make testpersons explore) their own competence and memory capacities to depict their findings in some semantic or conceptual structures (lists, arrays, networks, etc.).

Other than these introspective explorations, the present approach strives to derive directly via automatic analysis of natural language discourse (input) some basic data (output) whose relational structure need not be defined *statically* in declarative terms of *logical-deductive* hierarchies but will instead be imposed procedurally by algorithms which allow for the *dynamic* induction of relevant *analogical-associative* dependencies to form semantic dispositions¹.

By way of a sketchy overview rather than a qualifying introduction, it will (*first*) be outlined according to what principles the natural language discourse is analysed statistically and how the data obtained is represented formally. Constituting the semantic space model (*second*), its structure is examined for specific meaning representations, their positions, environments, and clustering properties. Starting from the notion of priming and spreading activation in memory as a cognitive model for comprehension processes, we will

*This paper reports on some of the objectives of a project in Computational Semantics, worked on by the MESY-Group at the German Department of the Technical University of Aachen, West Germany, under support of the Northrhine-Westphalia Ministry of Science and Research under grant IV A 2-FA 8600. The project is concerned with the development of automated means for the construction of lexical and/or semantic systems of stereotype/prototype knowledge representation from natural language discourse. Published in: Allen, R.F. (Ed.): Data Bases in the Humanities and Social Sciences (Proceedings of the 4th International Conference ICDBHSS/83), Osprey, FL (Paradigm Press) 1985, pp. 374–386.

¹Instead of formally introducing any of the algorithms developed and tested so far for the purposes at hand, some ideas of their performance and application shall in the sequel be tried to be given by way of some — hopefully illustrative — transparencies and examples. For more detailed introductions the reader is referred to the bibliography at the end of this paper where additional informations on the MESY-project in general and its procedural approach in particular may be found in a number of recent publications.

(*third*) deal with our procedural method of representing semantic dispositions by way of inducing lexical relevance relations within semantic space. Concluding (*fourth*) we shall point to two or three problem areas connected with word meaning and concept processing which may be tackled anew and perhaps brought to a more adequate though still tentative solution under an empirically founded approach to procedural semantics.

1 Statistical Text Analysis and Data Representation

It has been shown elsewhere² that in a sufficiently large sample of pragmatically homogeneous texts, called *corpus*, only a restricted vocabulary, i.e. a limited number of lexical items will be used by the interlocutors however comprehensive their personal vocabularies in general might be. Consequently, the lexical items employed to convey information on a certain subject domain under consideration in the discourse concerned will be distributed according to their conventionalized communicative properties, constituting *semantic regularities* which may be detected empirically from the texts.

The empirical analysis of discourse and the formal representation of vague word meanings in natural language texts as a system of interrelated concepts is based on the WITTGENSTEINIAN notion of *language games* and his assumption that a great number of texts analysed for the terms' usage *regularities* will reveal essential parts of the concepts and hence the meanings conveyed.

A meaning of a word is a kind of employment of it. For it is what we learn when the word is incorporated into our language. That is why there exists a correspondence between the concept *rule* and *meaning*. [...] Compare the meaning of a word with the function of an official. And *different meanings* with *different functions*. When language games change, then there is a change in concepts, and with the concepts the meanings of word change. [No. 61–65], WITTGENSTEIN (1969), p. 10e

The statistics which have been used so far for the systematic analysis not of propositional strings but of their elements, namely words in natural language texts, is basically descriptive. Developed from and centered around a correlational measure to specify intensities of co-occurring lexical items used in natural language discourse, these analysing algorithms allow for the systematic modelling of a fragment of the lexical structure constituted by the vocabulary employed in the texts as part of the concomitantly conveyed world knowledge.

A correlation coefficient appropriately modified for the purpose has been used as a mapping function. It allows to compute the relational interdependence of any two lexical items from their textual frequencies. Those items which co-occur frequently in a number of texts will positively be correlated and hence called *affined*, those which only one (and not the other) frequently occurs in a number of texts will negatively be correlated and hence called *repugnant*. Different degrees of *word-repugnancy* and *word-affinity* — indicated by

²See e.g. RIEGER (1977) where the principle of *semantization* is introduced as a procedural means to constitute *meanings* by restricting choices from the level of pragmatics, via semantics and syntactics down to morpho-phonetics. The ranges of possible choice on each of these *semiotic* levels are established by an equally generative, however inverted, corresponding restrictions of formal *combinatorial* limitations of the numbers of possible string combinations of any set's elements and/or symbols to a limited number of recurring realizations which — on one semiotic level — allow for redundancies that will serve as interpreted string elements of new sets to be combined — on the next semiotic level — again without exhausting all their combinatorial possibilities, and so forth, from phonemes to syllables, syllables to words, words to phrases, phrases to discourses, etc.

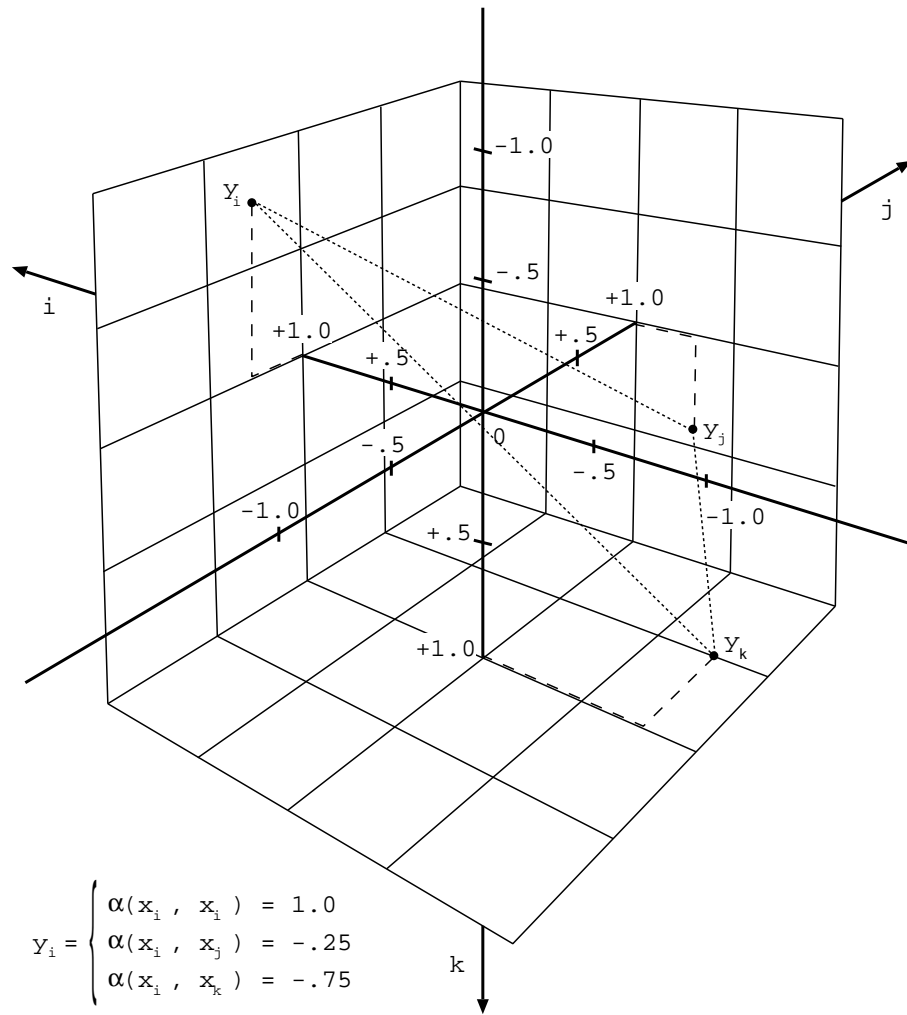


Fig. 1

numerical values ranging from -1 to $+1$ — may thus be ascertained without recurring to an investigator's or his test-persons' word and/or world knowledge (semantic competence), but can instead solely be based upon the usage regularities of lexical items observed in a corpus of pragmatically homogeneous texts, spoken or written by real speakers/hearers in actual or intended acts of communication (communicative performance).

Let T be such a corpus that consists of t texts belonging to a specific *language-game*, i.e. satisfying the condition of pragmatic homogeneity. For the sake of illustrating the analysing algorithm's performance, we will consider a simplified case where the vocabulary V employed in the texts shall be limited to only three word-types, namely x_i, x_j and x_k which have a certain overall token-frequency. Then the modified correlation coefficient A will measure the regularities of usage by the *affinities* and *repugnancies* that may hold between anyone lexical item and all the others employed in the discourse analysed. That will yield for any item an n -tuple of correlation-values α , in this case for the lexical item x_i with $n = 3$ the triple of values $\alpha_{ii}, \alpha_{ij}, \alpha_{ik}$. These correlation-values are now interpreted

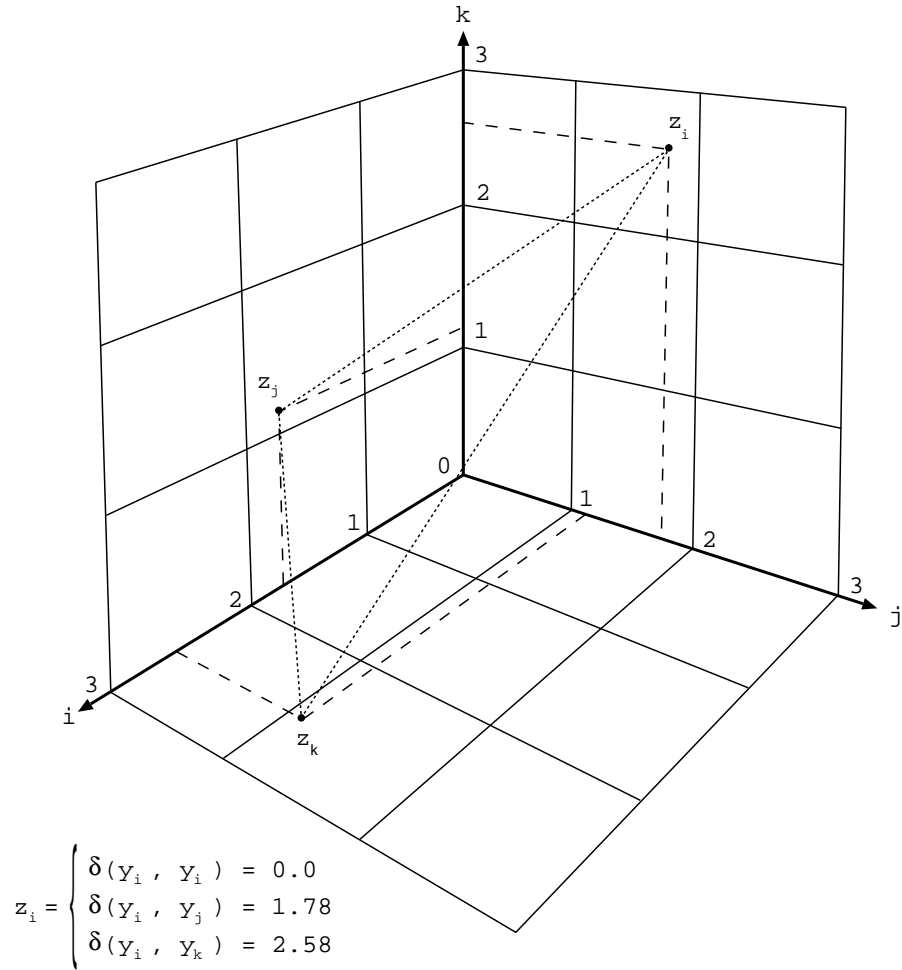


Fig. 2

as being coordinates that will define for each lexical item x_i , x_j and x_k one point $y(\alpha_i)$, $y(\alpha_j)$, and $y(\alpha_k)$ respectively in a three-dimensional space structure spanned by the three axis i , j and k as illustrated in Fig. 1. As the positions of these points now obviously depend on the regularities the lexical items concerned have been used with in the texts of the corpus, the y -points are called *corpus-points* of i , j and k in the α - or *corpus-space*.

Two y -points in this space will consequently be the more adjacent to each other, the less their usages differ. These differences may be calculated by a distance measure δ between any two y -points, as illustrated in Fig. 1 by dotted lines. The distance-values are real, non-negative numbers which represent a new characteristic. For any item y_i , y_j , and y_k an n -tuple of δ -values, i.e. for y_i the triple δ_{ii} , δ_{ij} , δ_{ik} is obtained which may be interpreted as new coordinates. These will again for each item x_i , x_j , and x_k define new points $z(\delta_i)$, $z(\delta_j)$, and $z(\delta_k)$ in a new n -dimensional space, called *semantic space*, as illustrated in Fig. 2.

The positions of such points in the semantic space will clearly depend on all the differences (δ - or distance-values) in all the regularities of usage (α - or correlation-values)

any lexical item shows in the texts analysed. Thus, each lexical item is mapped onto a fuzzy subset of the vocabulary according to the numerically specified regularities these items have been used with in the discourse analysed. Measuring the differences of any one's lexical item's usage regularities against those of all others allows for the above interpretation and consecutive mappings of items onto theoretical constructs. These new entities — called *meanings* — are operationally defined, and may verbally be characterized as a function of all the differences of all regularities any one item is used with compared to any other item in the same corpus of discourse.

2 Cluster Analysis and Structure of Semantic Space

The resulting system of sets of fuzzy subsets is a relational datastructure which may be interpreted topologically as a hyperspace with a natural metric. Its linguistically labelled elements represent *meaning-points*, and their mutual distances represent *meaning-differences*. The position of a meaning point may be described by its semantic environment. This is determined by those other points in the semantic hyperspace which — within a given diameter — are most adjacent to the first one.

Fig. 3 shows the topological environments, i.e. those points being situated within the hypersphere of a certain diameter of three meaning points, namely ATOM (atom), INDUSTRIE (industry) and COMPUTER (which needn't be translated) as computed from a corpus of newspaper texts comprising some 8000 tokens of 360 types in 175 texts from the 1964 editions of the German daily DIE WELT.

Having seen that the environments do in fact assemble meaning points of a certain semantic affinity, a couple of questions came up which I will only touch upon not, however, discuss in detail here:

- are there regions of *point density* in the semantic space, forming clouds and clusters which might indicate a semantic (syntagmatic and/or paradigmatic) structuredness?
- can such regions be detected and described automatically by statistical methods of multi-varied and cluster analysis, and how would they look like?
- could the *internal relation* according to which certain meaning points cluster be specified in terms of the logical-declarative vs. analogical-associative opposition of *semantic relatedness*?

The investigation of these questions (RIEGER 1981, 1982, 1983) have produced results according to which regions of point density could be ascertained by cluster analysing methods, assembling lexical items, however, which seemed to be both, paradigmatically and syntagmatically relatable, forming more of a *connotative cloud* than a *semantic field*. Its internal relations appeared to be *declaratively* unspecifiable beyond their contents-driven *associative* connectedness of "having something to do it" that any distance-related representational format might be translated to.

3 Spreading Activation and Connotative Dependencies

One of the problems of distance-like data structures in semantic processing is that — *distance* being a symmetric relation — well-known search strategies for retrieval, matching,

Semantische Umgebung $E(x)$		DIE WELT	
$x = \text{ATOM/AR}$			
AMERIKA/ER/ISCH	5.106	WAFFE	5.337
MACHT/IG	5.447	ENTSCHEIDEN/UNG	5.891
RÜSTEN/UNG	6.144	ENG/E/N	6.149
VERSUCH/EN	6.222	TEILNAHME/N	6.442
KAMPF/EN	6.494	FRIEDE/LICH	6.583
SOWJET/ISCH/ION	6.862	ABKOMMEN	6.926
ZONE	7.340	ATLANTIK/SCH	7.403
VERTRAG	7.571	KONTROLLE/IEREN	7.619
STARK/E	7.651	KRIEG/ERISCH	7.727
OSTEN	7.772	FEIER/N	7.777
DIENST/EN	7.849	GEWOHNHEIT/LICH	7.885
REPUBLIK/ANISCH	7.986	KOMMUNIST/MUS	8.064
GRENZEN/N	8.093	GEMEINSCHAFT/CH	8.194
DEUTSCH/LAND	8.208	SONNTAG	8.220
EUROPA/ISCH	8.224	FRAKTION	8.250
Semantische Umgebung $E(x)$		DIE WELT	
$x = \text{INDUSTRIE/IEREN}$			
ELEKTRO/NISCH	2.106	LEITEN/R/UNG	2.369
BERUF/LICH	2.507	SCHULE/R	3.229
SCHREIBEN	3.328	COMPUTER	3.667
FÄHIG/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/EIT	18.983	AUSGABE/GEBEN	19.302
ANBIETEN/GEBOT	20.335	ALLGEMEIN	21.685
ARBEIT/EN	22.182	VERANTWORTEN/NG	24.320
WERBEN/UNG	25.119	VERKEHR/EN	26.932
Semantische Umgebung $E(X)$		DIE WELT	
$x = \text{COMPUTER}$			
ERFAHREN/UNG	1.294	LEITEN/R/UNG	1.529
FÄHIG/KEIT	1.722	SYSTEM/ATISCH	2.065
DIPLOM	2.067	KENNEN/TNIS	2.737
SUCHE/N	2.864	INDUSTRIE/REN	3.667
ELEKTRO/NISCH	4.339	TECHNIK/ISCH	4.344
BERUF/LICH	4.777	SCHULE/R	5.905
SCHREIBEN	6.371	UNTERRICHT/EN	8.839
BITTE/N	10.340	ORGANISATION	11.076
WUNSCH/EN	11.659	STELLE	14.238
UNTERNEHMEN/R	17.635	STADT	19.592
GEBIET	20.654	VERBAND	20.819
PERSON/LICH/KT	21.591	AUSGABE/GEBEN	22.232
ANBIETEN/GEBOT	22.910	ALLGEMEIN	24.816
ARBEITEN/EN	24.849	WERBEN/UNG	26.969
VERANTWORT/UNG	27.642	VERKEHR/EN	30.073

Fig. 3

and inferencing purposes cannot be applied. In order to make such procedures operate on the semantic space data, its distance-like structure has to be transformed into some hierarchical organisation of its elements. How can that be done?

Taking up the heuristics as provided by *Spreading Activation Theory* in memory structures initially presented by QUILLIAN (1968) and COLLINS/LOFTUS (1975) and studied under the notion of *priming* in subsequent publications (e.g. SWINNEY 1979; FLORES D'ARCAIS/JARVELLA 1983), the semantic space may be interpreted as a means of empirically sampled, discourse-based, raw material which — other than material gathered from isolated word association task experiments — provides the necessary data for the dynamic structuring of meanings as contextual processes of choice restrictions. Represented as meaning points in a relational data structure, selecting from it the most relevant, i.e. contextually motivated relations between them thus allows for the generation of *semantic dispositions* as possible paths along which in case of *priming* activation might spread when one meaning point is stimulated.

Originally developed as a model to cope with observed latencies in processes of concept identification and recognition tasks, the notion of *priming* and *spreading activation* explaining those observations is based on network-type models of word-meaning or world-knowledge structures. Essentially, these are defined by labeled *nodes*, representing concepts, meanings or objects, and labeled *links* which relate them conceptually, semantically, or logically to one another.

Unlike these ready-set and fixed relations among nodes, we have devised an algorithm which operates on the semantic space data structure as its base to induce dependencies between its elements, i.e. among subsets of the meaning points. The recursively defined procedure detects fragments of the semantic space according to the meaning point it is started with and according to the semantic adjacencies, i.e. the distance relations it encounters during operation, constituting what we termed semantic *relevance*. Stop-conditions may deliberately be formulated either qualitatively (naming a target point) or quantitatively (number of points to be processed).

Given one meaning point's position being primed, the algorithm will first start to list all neighbouring points by their increasing distances. Then, the algorithm's generic procedure will open a tree with the initially primed point as its root before taking the first on the list, determining its most adjacent point among those already primed to identify it as its mother node, and then deleting the new daughter-node's label from the list.

Repeated successively for each of the meaning points listed and in turn primed in accordance with this procedure, the *algorithm of least distance* will select a particular fragment of the relational structure latently inherent in the semantic space, depending on the aspect, i.e. the primed meaning point the algorithm is initially started with. Working its way through and consuming all labeled points in the space system, the algorithmic procedure transforms prevailing similarities of stereotype meanings as represented by adjacent points to establish a binary, non-symmetric, and transitive relation between them. This relation which — according to the representational format it is derived from — we call *relevance relation* allows for the hierarchical re-organisation of meaning points as nodes under a primed head, i.e. the root in a general or n-ary dependency tree of *semantic dispositions*. This verbal description of the algorithm's operative characteristics may be exemplified by some hopefully instructive illustrations given in Fig. 4 to Fig. 7.

Starting from a distance-like data structure as shown in the two-dimensional configuration of 11 points and labeled *a* to *k* in Fig. 4, we observe the stimulation of e.g. points *a*

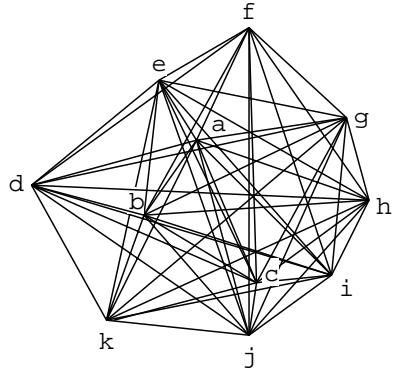


Fig. 4

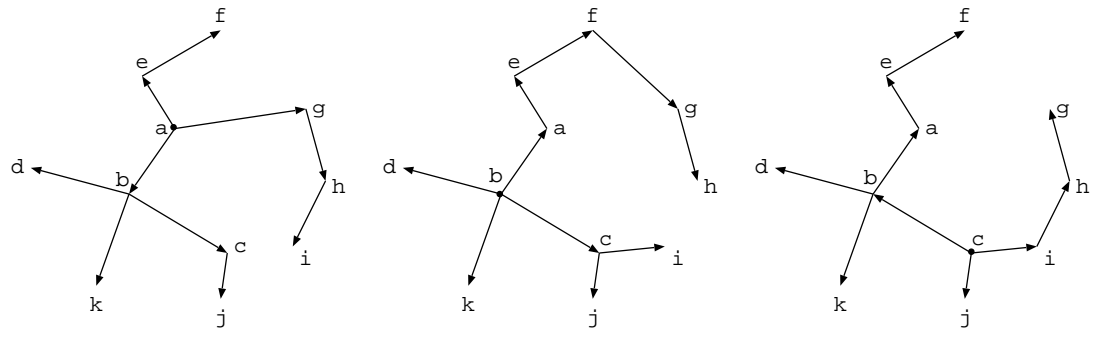


Fig. 5

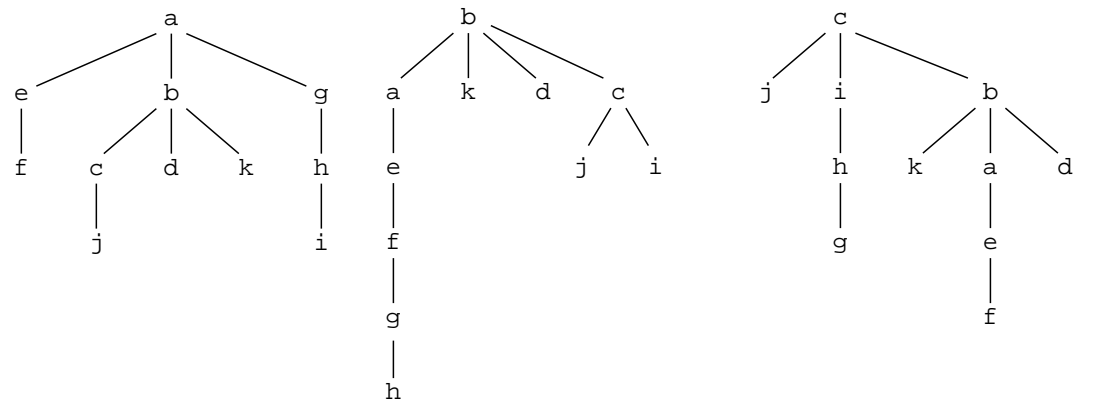


Fig. 6

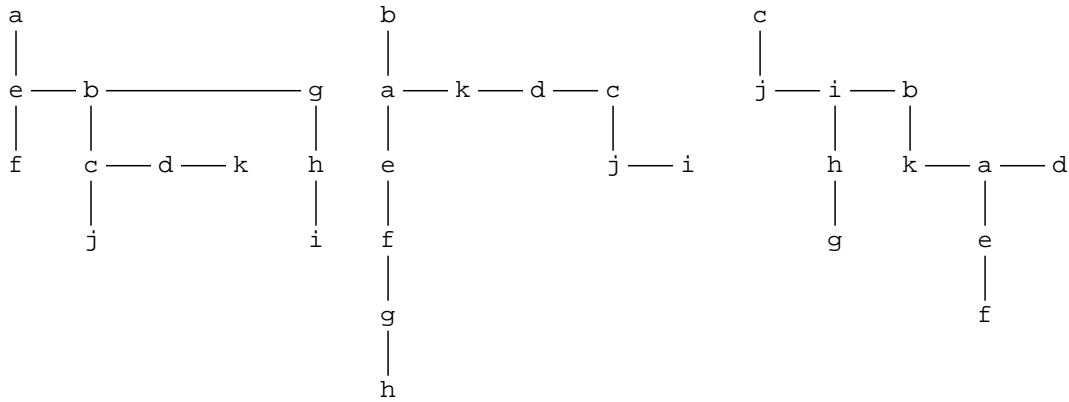


Fig. 7

whose neighbours' distances are detected and the least one's selected to form its characteristic configurations of related points in Fig. 5, which then is represented as an n -ary or *general tree* in Fig. 6 and transformed to a *binary tree* in Fig. 7 respectively to represent this meaning point's *dispositional dependency structure (DDS)*.

Stimulating other points within the same point configuration (as for example b and c as illustrated in Figs. 5 to 7), results in similar but nevertheless differing trees, depending on the aspect under which the structure is accessed, i.e. the point initially stimulated to start the algorithm with.

Applied to the semantic space data of 360 defined meaning points of the newspaper DIE WELT, Figs. 8 and 9 show what the *DDS-trees* of ERFAHR/experience and GESCHÄFT/business look like as generated by the above procedure described. In Fig. 8 we have on the tree's first level the three associative (or connotative) alternates, namely TECHNIK/technique, ORGANISATION/organisation and BERUF/profession, dependent from the head ERFAHR/experience, and so forth on the next level of the *DDS-tree*.

Attention is drawn to the marked path in this tree, signifying a dependency of SUCH/search via COMPUTER/computer, ELEKTRON/electronic and LEIT/guidance. This dependency is found in exactly the same order in the *DDS-tree* of GESCHÄFT/business, but here it is situated farther from the root, starting on the tree's sixth level only, instead of its third.

To calculate such differences, we have devised a numerical measure of *criteriality* of a node with respect to its root or aspect. This measure will not be introduced here, but can be characterized to be defined as a function of both, the distance values and the tree's levels concerned. Thus, for the simulation of analogical inferencing processes in natural language understanding systems based upon the flexible contents-structured format of *dispositional dependency trees* in procedural semantics, the different criterialities of nodes will be used to estimate which paths are more likely being taken against others which will be followed less likely under priming of certain meaning points³.

It goes without saying that generating *DDS-trees* is a prerequisite to source-oriented,

³It appears that on the foundation of *DDS-criterialities* there is a good chance to develop a numerical expression to measure the amount of meaning conveyed, based upon structural properties of open sets and systems of symbols, instead of probabilities as calculated from finite symbol sets in classical information theory.

DEPENDENCY PATHS

(primed node x, targetted node z)



INFERENCE PATHS

(premise nodes x_i , concludes node z)

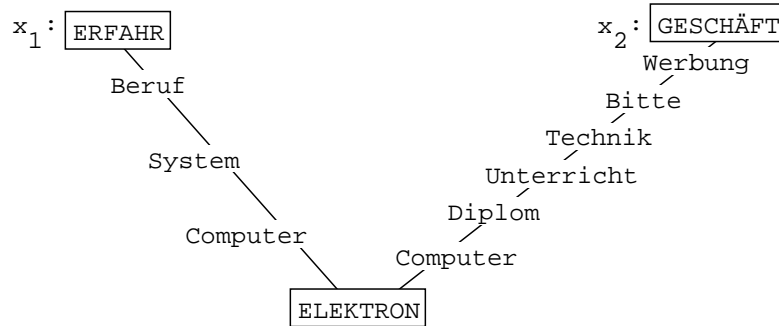


Fig. 10

contents-driven *search* and *retrieval* procedures which may thus be performed effectively on the semantic space structure. Given, say, the meaning point ERFAHR/experience to be stimulated, and, say, GESCHÄFT/business as the target point to be searched for, then, the *DDS* of ERFAHR/experience will be generated first. The nodes primed accordingly will with decreasing criterialities provide the *semantic dispositions* inherent in the semantic space data and triggered under the aspect of ERFAHR/experience. Then, the tree structure generated will be searched (breadth first) for the target node which — when hit — will stop the search procedure. Its *dispositional dependency* path will then be activated to trace those intermediate nodes which determine the connotative transitions of any target node under the selected aspect concerned. When we look up GESCHÄFT/business as a target node, we get its dependency path under the aspect of ERFAHR/experience to consist of WERBUNG/advertise, BITTE/request and TECHNIK/technique, which — not surprisingly though — proves to be the *dispositional dependency* of ERFAHR/experience under the aspect of GESCHÄFT/business but in inverted order (Figs. 8 and 9).

Using these source-oriented search and retrieval processes, an *analogical*, contents-driven form of inference — as opposed to *logical deduction* — may operationally be devised by way of parallel processing of two (or more) *DDS*-trees. For this purpose an algorithm is started by the two (or more) meaning points considered to represent the semantic *premises*, of say, ERFAHR/experience and GESCHÄFT/business. Their *DDS*-trees will be generated before the inferencing procedure begins to work its way through both trees, taking highest criterialities first in tagging each encountered node. When the first node in either tree is

met that has previously been tagged already, the search procedure stops to activate the dependency paths from this *concluding* common node — in our case ELEKTRON/electronic — in the *DDS*-tree concerned (marked by dotted lines in Fig. 8 and 9, and separately presented as Fig. 10).

4 Conclusions and Possible New Vistas?

4.1 Among others, the *DDS*-procedure provides a flexible, source-oriented, contents-driven method for the induction of a relevance relation among stereotypically represented concepts linguistically conveyed by natural language discourse on specified subject domains.

4.2 Applied to distance-like data structures, the *DDS*-procedure allows for the generation of possible paths of spreading activation which branch across semantic space, submitting relevant portions of it to associatively guided search strategies and retrieval operations.

4.3 The problem of identifying stored meaning constructions with distorted instantiations of them, can be circumvented. The procedural approach replaces the storage of fixed and ready-set networks by a contents-driven induction of relevance relations between nodes. Triggered by any identifiable label, the *DDS* will be generated according to the database provided and the resultant tree-structure will therefore vary according to the possibly varying status of the data in space structure.

4.4 In view of tacit knowledge and implied information the *DDS*-procedure offers an empirically based approach and a dynamic representation of *semantic dispositions* which — in language understanding systems — might serve as connotative default values in identifying and/or interpreting input labels and solving ambiguity and/or vagueness problems of input strings.

References

- COLLINS, A.M./LOFTUS, E.F. (1975): A spreading activation theory of semantic processing, *Psychological Review* 6 (1975) 407–428
- FLORES D'ARCAIS, G.B./JARVELLA, C. (Eds)(1983): The Progress of Language Understanding. New York/Sydney/Toronto (Wiley Sons) in press
- QUILLIAN, M.R. (1968) Semantic Memory. In: Minsky, M. (Ed.): Semantic Information Processing, Cambridge, Mass. (MIT Press) 70–106
- RIEGER, B. (1977): Bedeutungskonstitution. Bemerkungen zur semiotischen Problematik eines linguistischen Problems, *Zeitschrift f. Linguistik u. Literaturwissenschaft* 27/28 (1977) 55–68
- RIEGER, B.B. (1981): Connotative Dependency Structures in Semantic Space. in: Rieger, B.B. (Ed.): Empirical Semantics I & II. A Collection of New Approaches in the Field (Quantitative Linguistics No. 12 & 13), Bochum (Brockmeyer) 622–710

- RIEGER, B.B. (1982): Procedural Meaning Representation. An empirical approach to word semantics and analogical inferencing. In: Horecky, J. (Ed.): COLING 82. Proceedings of the 9th Intern. Conf. on Computational Linguistics (Linguistic Series 47), Amsterdam/New York (North Holland) 319–324
- RIEGER, B.B. (1983): Clusters in Semantic Space. In: Delatte, L. (Ed.): Actes du Congrès International Informatique et Sciences Humaines, Liège (LASLA) in press
- SWINNEY, D.A. (1979): Lexical processing during sentence comprehension, *Journal of Verbal Learning and Verbal Behaviour* 18 (1979) 733–743
- WITTGENSTEIN, L. (1969): Über Gewißheit — On Certainty. New York/San Francisco/London (Harper & Row)