Reconstructing Meaning from Texts
A Computational View on Natural Language Understanding*

Burghard B. Rieger
Department of Computational Linguistics
FB II: LDV/CL – University of Trier

1 Introduction

1.1 Recent achievements being made within the intersection of cognitive psychology, artificial intelligence and quantitative linguistics appear to combine promising components for wordsemantics, conceptual structures, and knowledge representation. These are likely to become seminal in the future for a wide range of disciplines and applications concerned with *natural language understanding* by machine. With regard both to the prospects of new technologies and to the potential benefits or detriments that these could (though not necessarily will) imply, cognitive theory and applied cognitive science will consequently play an increasingly important role in the information society of the future [1]. Significant effects have been witnessed already as produced by advances in some related areas as well as in rather remote branches of science and society. However, our understanding of the bunch of complex intellectual activities subsumed under the notion of *cognition* is still very limited, particularly in how knowledge is acquired from texts and how this process can be modeled.

1.2 From the linguistic viewpoint natural language texts, whether stored electronically or written conventionally, will in the foreseeable future provide the major source of scientifically, historically, and socially relevant information. Due to the new technologies, the amount of such textual information continues to grow beyond manageable quantities. Availability of *data*, therefore, no longer serves to solve an assumed problem of lack of *information* to fill a *knowledge* gap in a given instance, but will instead create a new problem which arises from the abundance of information that confronts the potential user.

1.3 There is an increasing need to employ computers more effectively than hitherto for the analysis of natural language material. Although the demand is high for intelligent machinery to assist in or even provide speedy and reliable selection of relevant information under individual aspects of interest from any subject domain, such systems are not yet available. Development of earlier proposals [2], have resulted in some advances [3] towards an artificial *meaning learning and understanding* system (MLU) as core of a cognitive information processing system (CIPS) which will be capable of learning to understand (i.e. identify and interpret) the meanings implied in natural language texts by generating perspectival and dynamic conceptual dependencies (i.e. semantic inferencing) [4]. In view of a text skimming system under development [5], a basic *cognitive* algorithm has been

*Paper presented at the 2nd German-Chinese-Electronic-Week (DCEW 91), October 25–29, 1991, Jiaotong University, Shanghai, Peoples Republic of China.

Published in: Proceedings of 2nd German-Chinese Electronics Week. Shanghai, October 7th - 11th.

designed which detects from the textual environment the system is exposed to those structural information which the system is able to collect due to its own two-level knowledge structuredness. It allows for the automatic generation of a pre-predicative and formal representation of conceptual knowledge which the system will both, gather from and modify according to the input texts processed. The system's internal knowledge representation is planned to be made accessible in a dialog interface. This will allow users to make the system *skim* masses of texts for them, display its acquired knowledge in dynamic structures of conceptual dependencies, provide valuable clues for relevant connections, and help to avoid unnecessary reading of irrelevant texts.

2 Word Meaning and World Knowledge

2.1 The *representation* of knowledge, the *understanding* of meanings, and the *analysis* of texts, have become focal areas of mutual interest of various disciplines in cognitive science whose (preferably dynamic) *computational* modelling obviously serves to unify descriptive, explicative, procedural, and simulative purpose at stake [6]. Although current semantic theories of word meanings and world knowledge generally refer to memory in human or artificial systems of cognition and understanding as a complex structure of interrelated concepts, rather different approaches and models have been proposed.

2.2 In linguistic semantics, cognitive psychology, and knowledge representation most of the necessary data concerning lexical, semantic and external world information is still provided introspectively. Researchers are exploring (or make test-persons explore) their own linguistic or cognitive capacities and memory structures to depict their findings (or to let hypotheses about them be tested) in various representational formats. It is widely accepted that model structures resulting from these analysis do have a more or less *ad hoc* character, and tend to be confined to their limited theoretical or operational performances within a specified knowledge domain or implemented system. By definition, these approaches can map only what is already known to the analysts, not, however, what of the world's fragments under investigation might be conveyed in texts unknown to them.

2.3 Being *interpretative* and unable of auto-modification, such knowledge representations will not only be restricted to rule-based predicative and propositional structures which can be mapped in well established (concept-hierarchical, logically deductive) formats, but they will also lack the flexibility and dynamics of *associative* model structures more adapted to re-constructive meaning analysis and automatic representation from input texts. These have been recognized to be essential [7] for any *learning* device capable to set up and modify a system's own knowledge structure, however shallow and vague such knowledge may appear compared to human understanding. New connectionistic models of neural networking and learning algorithms appear to be promising though not yet available on the semantic level.

3 Empirical Approach

3.1 Other than introspective data acquisition, our present approach has been based on the algorithmic analysis of discourse that real speakers/writers produce in actual situations of performed or intended communication in a certain subject domain. Under the notion of *lexical relevance* and *semantic disposition* [8], a conceptual meaning representation system has operationally been defined which may empirically be reconstructed from natural language texts. Based upon the WITTGENSTEINian concept of *language games* it is assumed that a great number of texts analysed for the terms' usage *regularities* will reveal the central concepts employed and hence their meanings conveyed [9].

3.2 It has been shown elsewhere [10] that in a sufficiently large sample of *pragmatically homogeneous* texts 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 words 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 constraints*. These may be detected empirically from masses of texts which are considered systems or structured sets of strings of linguistic elements.

3.3 The statistics used so far for the analysis of *syntagmatic* and *paradigmatic* relations on the level of *words* in discourse, is basically descriptive. Developed from and centred around a correlational measure to specify intensities of co-occurring lexical items, these analyzing 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. Thus, a modified correlation coefficient has been used as a *first* mapping function α. 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 of 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* will be indicated by numerical values ranging from -1 to +1. The regularities of usage of any lexical item will be determined by the tuple of its *affinity/repugnancy*-values towards each other item of the vocabulary which — interpreted as coordinates — can be represented as points in a vector space, spanned by the number of axes each of which corresponds to an entry of the vocabulary. Any two of such points will be the more adjacent to each other, the less the usages of their corresponding lexical items differ. These differences may be calculated by a distance measure δ of, say, EUCLIDian metric. It serves as a *second* mapping function to represent any items differences of usage regularities measured against those of all other items. The resulting sets of distance values may again be interpreted as coordinates to define a new entity, called *meaning point*, in another space structure, called *semantic hyperspace* (*SHS*).

4 Semantic Hyperspace

4.1 As a result of these *consecutive* mappings, any meaning point's position in *SHS* is determined by all the differences (δ- or distance-values) of all regularities of usage (α- or correlation-values) a lexical item shows against all others in the text-corpus analysed. Thus, it is the basic analyzing algorithm which by processing NL texts provides the *MLU*--system with the information necessary to represent the system's status of knowledge. This is achieved without recurring to any investigator's or his test-persons' word or world knowledge (*semantic competence*), but solely on the basis of usage regularities of lexical item in discourse which is produced by real speakers/hearers in actual or intended acts of communication (*communicative performance*).

4.2 The systematic constraints represented by the system of meaning points may be formalized as a set of *fuzzy subsets* [11] of the vocabulary. This serves to depict the distributional character of word *meanings* as composed of a number of operationally defined components whose varying contributions can be identified with numerical values of the respective membership functions as derived from and specified by the differing usage regularities that the corresponding lexical items have produced in discourse. This translates the WITTGENSTEINian notion of *meaning* into an algorithmic operation that may be applied empirically to any corpus of *pragmatically homogeneous* texts (i.e. a *language game*).

4.3 Structural *lexical knowledge* is sofar represented as a relational data structure whose linguistically labeled elements (*meaning points*) and their mutual distances (*meaning differences*) form a system of *prototypes*. Accordingly, the *meaning* of a lexical item may be described either as a fuzzy subset of the vocabulary, as a meaning point vector, or as a meaning point's topological environment. The latter is determined by those points which are found to be most adjacent and hence will delimit the central point's meaning indirectly as its *stereotype* (*Tab.* 1).

WIRTSCHAFT/economy		0.000					
AUSLAND	3.785	BRITAIN	5.094	ENTWICKL	5.893	FOLGe	6.112
VERWALT	6.428	RAUM	6.903	EINSATZ	9.307	KONTAKT	9.934
HERRSCHEn	10.163	GESCHFT	10.931	KRANK	11.732	VERKEHR	11.984
VERANTWORT	12.298	SPRACH	12.429	MGLICH	13.257	WEG	13.285
NEU	13.871	ZENTRAL	14.831	LEHR	15.131	JUNG	15.550
ALLGEMEIN	15.796	MODE	15.850	AUFTRAG	15.952	MASCHINE	16.210
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Topological environment $E(z_i, r)$ of $i =$ WIRTSCHAFT/economy listing points situated within the hypersphere of radius r in the *semantic hyperspace* $\langle S, \delta_2 \rangle$ as computed from a text sample of the 1964 editions of the German daily DIE WELT (175 articles of approx. 7000 word tokens and 365 word types).

5 Procedural Structuring

5.1 Following a *semiotic* notion of *understanding* and *meaning constitution*, the *SHS*-structure may be considered the core of a two-level conceptual knowledge representation system [12]. Essentially, it separates the format of a basic (stereotype) word meaning representation from its latent (dependency) relational concept organization. Whereas the former is a rather static, topologically structured (associative) memory, the latter can be characterized as a collection of dynamic and flexible structuring procedures to reorganize the memory data by semiotic principles under various aspects and perspectives. Following *Spreading Activation Theory* [13], to understand faster spread of activation of related concepts in cases where these have previously been *primed*, this theory's heuristics can also be employed to signify a process which induces relevance relations between concepts on the basis of their similarity, allowing for *priming* and *activation* procedures alike.

5.2 *SHS* being a distance-relational data structure, well-known algorithmic search strategies cannot immediately be made to work. They are mostly based upon some non--symmetric relational structure as e.g. directed graphs in traditional meaning and knowledge representation formats. To convert the *SHS*-format into such a node-pointer-type structure, the *SHS*-model has to be considered as conceptual raw data or associative base structure which particular procedures may operate on to reorganize it. Thus, the *distributed* representational format of *SHS* which had appeared to be disadvantageous first, proved to be superior over more traditional formats of *symbolic* representation. Other than in these pre-defined semantic network structures of predicative knowledge, non-predicative meaning relations of lexical *relevance* and semantic *dispositions* depend heavily on con- and cotextual constraints which will more adequately be defined procedurally, i.e. by generative algorithms that induce them on changing data only and whenever necessary. This is achieved by a recursively defined procedure that produces hierarchies of meaning points, tree-structured under given aspects according to and in dependence of their meanings' relevancy.

6 Dispositional Dependencies

6.1 Unlike conceptual representations that link nodes to one another according to what cognitive scientists supposedly know about the way conceptual information is structured in memory [14], an algorithm has been devised which operates on the *SHS*-data to induce *dispositional dependency structures* (*DDS*) between its elements, i.e. among subsets of meaning points conceptually related. The recursively defined procedure detects fragments from *SHS* according to the meaning point it is started with and according to the constraints of semantic similarity it encounters during operation.

6.2 This is tantamount to a numerical assessment (*criteriality*) and a hierarchical restructuring (*tree-graph*) of elements under a head point's aspect and the induction of a reflexive, non-symmetric dependency relation between descendant points along which activation might spread in case of head point stimulation. Stop-conditions may deliberately be formulated either qualitatively (i.e. naming a target point) or quantitatively (i.e. number of points, realm of distance or criteriality to be processed).

6.3 Applied to the *SHS*-data, the *Dispositional Dependency Structures* (*DDS*) of WIRTSCHAFT/economy is given in *Fig.* 1 as generated by the procedure described. For a wide range of purposes in processing *DDS*-trees, differing criterialities of nodes can be used to estimate which paths are more likely being taken against others being followed less likely under priming activated by certain meaning points.

7 Semantic Inferencing

7.1 Exploiting the *syntagmatic/paradigmatic* constraints of linguistic string formation without parsing of their *syntactic* structures, the *dispositional dependencies* appear to be a prerequisite not only to source-oriented, contents-driven *search* and *retrieval* procedures which may thus be performed effectively and fast on any *SHS*-structure. Due to its procedural definition, *DDS* also allow to detect varying dependencies of nodes under different perspectival aspects which might change dynamically and could therefore be employed in conceptual, pre-predicative, and *semantic* inferencing as opposed to propositional, predicative, and *logic* deduction.

7.2 For this purpose a procedure was designed to operate simultaneously on two (or more) *DDS*-trees by way of (emulated) parallel processing. The algorithm is started by two (or more) meaning points which may be considered to represent conceptual or semantic *premises*. Their *DDS* can be generated while the actual inferencing procedure begins to work its way (breadth-first, depth-first, or according to highest criteriality) through both (or more) trees, tagging each encountered node. When the first node is met that has previously been tagged by activation from another premise, the search procedure stops to activate the dependency paths from this *concluding* common node back to the *premises*, listing the intermediate nodes to mediate (as illustrated in *Tab.* 2) the conceptual inference structure.

UNTERNEHM/enterprise	0.0/1.000	⇐ *Premises* ⇒	0.0/1.000	WIRTSCHAFT/economy
STADT/city	5.57/.428			
GEBIET/area	4.05/.239		6.43/.421	VERWALT/administrat
VERBAND/union/league	3.78/.144		5.62/.244	EINSATZeffort/supply
ALLGEMEIN/general	6.78/.076			
Conclusion ⇒	6.39/.046	VERANTWORT	6.37/.151	⇐ *Conclusion*

Table 2: *Inference paths* from UNTERNEHM/enterprise and WIRTSCHAFT/economy to VERANTWORT/responsibility.

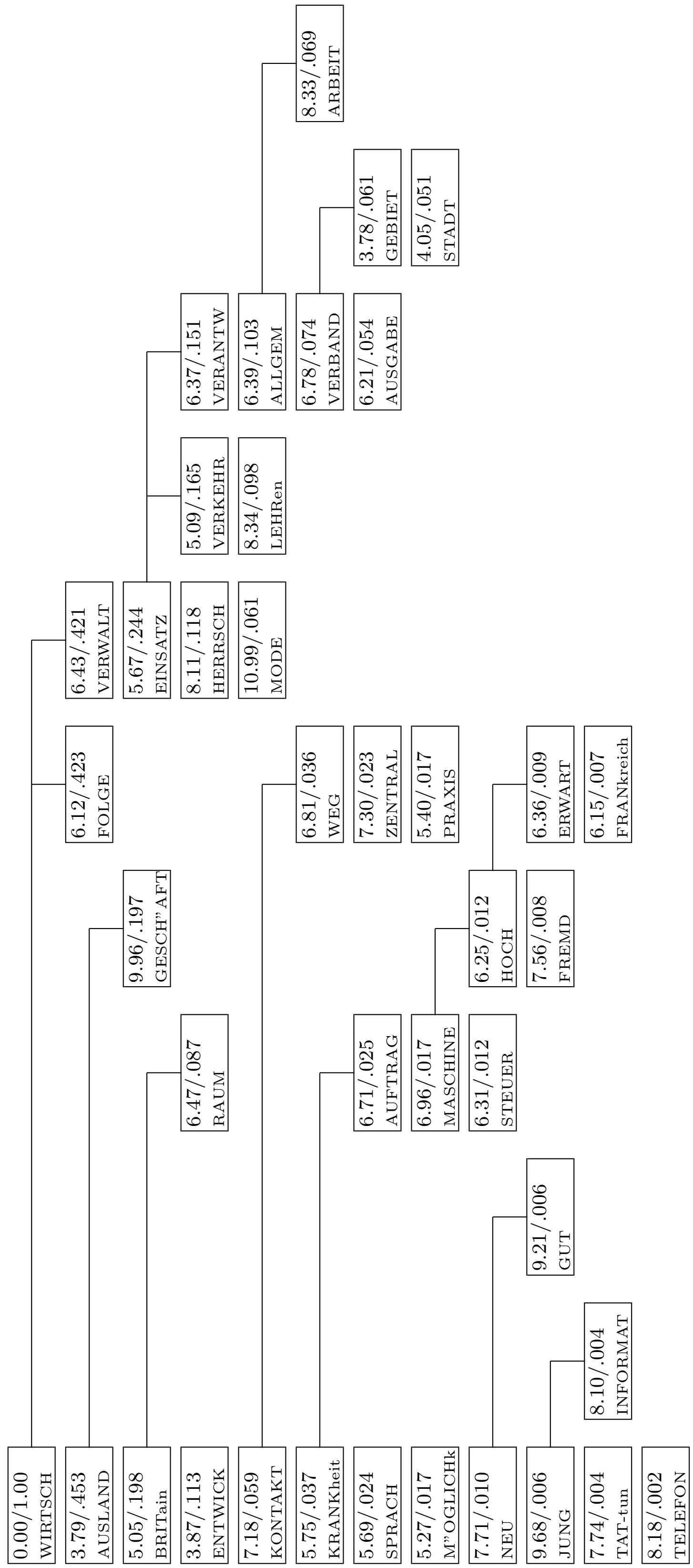

Figure 1: $DDS\langle z_i \rangle$-tree of start and head node i =WIRTSCHAFT/economy with *criterialities* (1st value) and *distances* (2nd value) of descendant nodes as calculated from the newspaper corpus of DIE WELT.

8 Conclusion

It is hoped that our system will prove to provide a flexible, source-oriented, contents-driven method for the *multi-perspective* induction of dynamic conceptual dependencies among stereotypically represented concepts which — being linguistically conveyed by natural language discourse on specified subject domains — may empirically be detected, formally be presented, and continuously be modified in order to promote *meaning learning and understanding*-systems (*MLU*) for machine intelligence.

References

* This study was supported by a Sabbatical Grant of The German Marshall Fund of the United States.

[1] Kitagawa, T. (1980): The Role of Computerized Information Systems in Knowledge Societies. In: Pfab, R./Stachelsky, F./Tonnemacher, J.(Eds): Technische Kommunikation und gesellschaftlicher Wandel, Berlin (Spiess), pp.24–73

[2] Rieger, B. B. (1984): The Baseline Understanding Model. A Fuzzy Word Meaning Analysis and Representation System for Machine Comprehension of Natural Language. In: O'Shea, T.(Ed): Proceedings of the 6th European Conference on Artificial Intelligence (ECAI 84), New York/Amsterdam (Elsevier Science), pp.748–749

[3] Rieger, B. B./Thiopoulos, C. (1989): Situations, Topoi, and Dispositions. On the phenomenological modelling of meaning. in: Retti, J./Leidlmair, K. (Hrsg): 5th Austrian Artificial Intelligence Conference. (ÖGAI 89) Innsbruck; (KI–Informatik-Fachberichte Bd.208) Berlin/Heidelberg/New York (Springer), pp.365–375

[4] Rieger, B. B. (1991): On Distributed Representation in Word Semantics. (ICSI Report TR–91–12), International Computer Science Institute, Berkeley, CA

[5] Rieger, B. B. (1988): TESKI – A natural language TExtSKImmer for shallow understanding and conceptual structuring of textually conveyed knowledge. LDV/CLReport 10/88, Dept. of Computational Linguistics, University of Trier

[6] Pfeifer, R./Schreter, Z./Fogelman-Soulie, F./Steels, L. (Eds.) (1989): Connectionism in Perspective. Amsterdam/New York (NorthHolland/Elsevier Science)

[7] Winograd,T./Flores, F. (1986): Understanding Computers and Cognition: A New Foundation for Design. Norwood, New Jersey (Ablex)

[8] Rieger, B. B. (1985): Lexical Relevance and Semantic Disposition. On stereotype word meaning representation in procedural semantics. In: Hoppenbrouwes, G./Seuren, P./Weijters, T. (Eds.): Meaning and the Lexicon. Dordrecht (Foris), pp.387–400

[9] Wittgenstein, L. (1969): Über Gewißheit - On Certainty. New York/San Francisco/London (Harper & Row), [No.61–65], p.10e

[10] Rieger, B. B. (1981): Feasible Fuzzy Semantics. In: Eikmeyer, H. J./Rieser, H.(Eds): Words, Worlds, and Contexts. New Approaches in Word Semantics. Berlin/New York (de Gruyter), pp.193–209

[11] Zadeh, L. A. (1965): Fuzzy sets. Information and Control 8(1965) 338–353

[12] Rieger, B. B. (1989): Unscharfe Semantik. Bern/Frankfurt/New York (Lang)

[13] Collins, A. M./Loftus, E. F. (1975): A spreading activation theory of semantic processing. Psychological Review 6(1975) 407–428

[14] Schank, R. C. (1982): Dynamic Memory. A Theory of Reminding and Learning in Computers and People. Cambridge/London/New York (Cambridge UP)