

Generating Dependency Structures of Fuzzy Word Meanings in Semantic Space*

Burghard B. Rieger

Technical University of Aachen, Germany

This paper will report on one of the central objectives of a project in computational semantics which is supported by the Northrhine-Westphalia Ministry of Science and Research under grant IV A 2-FA 8600.

1. Modelling system structures of word meanings and/or world knowledge is to face the problem of their mutual and complex relatedness. Under the notion of semantic relevance and knowledge disposition this interdependency may empirically be reconstructable from natural language discourse although most approaches in linguistic semantics and artificial intelligence do not address these issues. Instead, linguists as well as experts engaged in word meaning and/or world knowledge representation still provide the necessary semantic or external world data introspectively by exploring their own competence and memory capacities to depict their findings in some semantic or conceptual structures (lists, arrays, networks, etc.). They do so with the understanding that their models may have a more or less ad hoc character and tend to lack — beyond their limited operational performance — intersubjective control. Other than these introspective explorations, the present approach strives to derive directly via automatic analysis of natural language discourse some basic data whose relational structure will not be declared but procedurally be defined by algorithms which induce it.

2. Based upon statistical means for the empirical analysis of discourse and for the formal representation of vague word meanings in natural language texts, procedures have been devised which 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 concerned [?]. The coefficients applied will map lexical items onto fuzzy subsets of the vocabulary according to the numerically specified regularities these items have been used with in the discourse analysed. The resulting system of sets of fuzzy subsets is a datastructure which may be interpreted topologically as a hyperspace with a natural metric. Its linguistically labeled elements (representing meaning points) and their mutual distances (representing meaning differences) form discernable clouds and clusters which determine

*Published in: Hattori, S./Inoue, K. (Eds.): Proceedings of the XIIIth International Congress of Linguists 1982, Tokyo (CIPL) 1983, pp. 543–548.

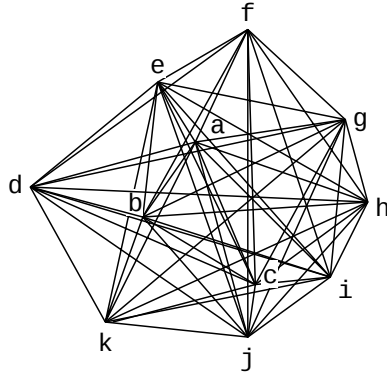


Fig. 1

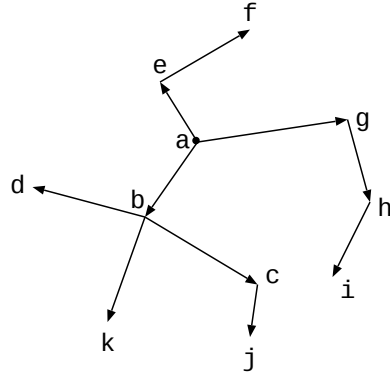


Fig. 2

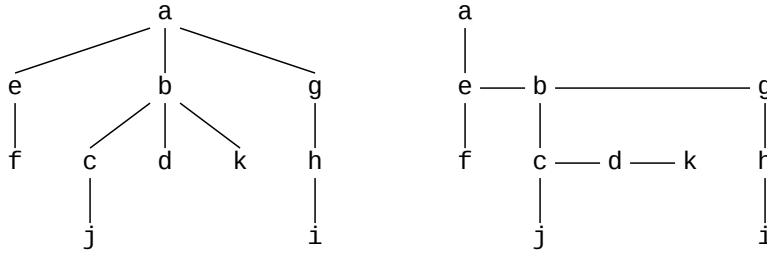


Fig. 3

the labels' associative meaning relations. Thus, the analysing algorithm takes natural language texts from a certain subject domain as input and produces as output the distance-like datastructure (semantic space) of linguistically labeled elements (meaning points) whose positions represent essential properties of the conceptual prototypes according to which their labels have been employed in the texts analysed [?]. Their varying dependencies which constitute a (latent) associative relational structure [?] may procedurally be defined and modelled on the semantic space data to allow not only for search and retrieval operations being executed but also for inferential processes being performed on that data structure under different aspects of semantic contents and relevance.

3. Taking up ideas from the theory of semantic memory and spreading activation in cognitive psychology [?], a new algorithm is presented which operates on the semantic space data to generate — other than the CDS-procedure [?] — associative dependency structures (ADS) in the format of general (n-ary) trees. Given one meaning point's position being primed, the algorithm will first start to list all neighbouring points by increasing distances. Then, the algorithm's generic procedure will take the first on the list, determine its most adjacent point among those already primed, and identify it as its mother-node before deleting the new daughternode'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 distances

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 ADS-algorithm transforms prevailing similarities of meanings as represented by adjacent points to establish a binary, non-symmetric, and transitive relation between them. This relation allows for the hierarchical reorganisation of meaning points as nodes under a primed head in an n-ary ADS-tree.

The process of detection and identification which the algorithm performs may be illustrated in view of a two-dimensional space configuration of 11 points $\langle \delta\{a, b, c, d, e, f, g, h, i, j, k\} \rangle$ (Fig. 1).

Submitted to the search procedure of least distances under initial priming of the point a the algorithm will identify the distances concerned as in Fig. 2 and produce the equivalent tree representations as shown in Fig. 3. For the effective use in procedural meaning representation and semantic processing, the ADS-trees may additionally be evaluated by associative criterialities, not given here. The criteriality is a numerical expression of the degree or intensity by which any ADS-node is dependent on its mother-node, calculated as a function of both, the involved meaning points' topology and its relative distances leading to the initially primed point in the semantic space.

Examples of associative dependency trees are given below where the upper fragments of the ADS's of ARBEIT/labour (Fig. 4) and INDUSTRIE/industry (Fig. 5) are shown as computed from the semantic space structure derived of a sample of German newspaper texts from the 1964 daily editions of 'Die Welt'.

4. The ADS-trees' properties permit different though related model-bound interpretations which can only be indicated here:

- identifying stored meaning representations with distorted and/or modified instantiations of them is no longer a problem. The procedural semantic approach replaces the storage of fixed and ready set relational (semantic) networks by source-oriented induction of a meaning point's associative dependency structure which is generated only when needed. Triggered by any identifiable label, the ADS generated may be identified with that label's associative meaning as instantiated according to the semantic space data, its subject domain, actual status, and possible changes;
- for the notion of semantic relevance and dispositional knowledge the ADS-procedure offers an empirically based approach to and a contents dependent representation of possible semantic/factual default connotations which become accessible under the aspect of a certain meaning point being primed;
- for models of memory and semantic structure the ADS-procedure provides a flexible, aspect-driven means for the detection of possible paths of spreading activation which branch across semantic space, submitting relevant portions of it to associatively guided semantic search strategies, retrieval operations, and processes of analogical reasoning, as opposed to logical deduction [?].

A R B E I T

8.332	ALLGEMEIN	8.756	ANGEBOT
6.786	STADT	6.931	AUSGAB
4.047	GEBIET	7.781	VERANTWORT
3.776	VERBAND	5.566	UNTERNEHM
		7.883	STELLE
		4.212	ORGANISATI
		4.513	UNTERRICHT
		4.060	BITTE
		5.776	SCHULE
		1.942	SCHREIB
		2.466	BERUF
		9.209	NEU
		9.673	JUNG
		8.226	TELEFON
		2.029	ELEKTRON
		2.106	INDUSTRIE
		2.051	SUCH
		1.779	LEIT
		1.530	COMPUTER
		4.835	TECHNIK
		2.318	FÄHIG
		1.294	ERFAHR
		2.619	SYSTEM
		1.571	DIPLOM
		2.541	KENNTNIS
		8.336	VERKEHR
		9.130	HERRSCH
		10.995	MODE
		8.567	GESCHÄFT
		10.345	LEHRE
		10.899	WERBUNG
		6.122	FOLGE
		3.785	AUSLAND
		6.597	RAUM
		5.046	BRITAIN
		14.084	GUT
		3.872	ENTWICKEL
		7.180	KONTAKT
		11.119	INFORMATI
		8.100	ZENTRAL

Fig. 4

2.051	—	2.106	
SUCH		ELEKTRON	
1.779	—	2.405	— 2.029
LEIT		SCHUL	BERUF
1.530		1.942	5.596
COMPUTER		SCHREIB	UNTERRICHT
1.294	—	1.772	— 2.065
ERFAHR		FÄHIG	SYSTEM
1.571		2.541	
DIPLOM		KENNTNIS	
		2.795	
		TECHNIK	
		4.060	— 4.513
		BITTE	ORGANISATI
		4.212	— 4.875
		STELLE	WUNSCH
		8.246	
		PERSON	
		5.867	— 4.465
		AUSGAB	ANGEBOT
		10.899	
		WERBUNG	
		19.748	
		BUCH	
		3.776	
		VERBAND	
		6.395	
		ALLGEMEIN	
		6.395	— 8.332
		VERANTWORT	ARBEIT
		7.781	— 6.746
		VERWALT	VERKEHR
		6.010	— 8.336
		WIRTSCHAFT	EINSATZ LEHR
		3.785	— 6.122
		AUSLAND	FOLGE
		5.046	
		BRITAIN	
		6.471	— 3.872
		RAUM	ENTWICKEL
		14.084	
		GUT	7.'80
			KONTAKT
		13.107	— 9.209
		INFORMATI	NEU
		8.100	
		JUNG	

5