Meaning Acquisition by SCIPS[*]

Burghard B. Rieger
FB II: Department of Computational Linguistics – University of Trier
`rieger@ldv01.Uni-Trier.de`

Abstract

The emergence of semantic structure as a self-organizing process is studied in *Semiotic Cognitive Information Processing Systems* on the basis of word usage regularities in natural language discourse whose linearly agglomerative (*syntagmatic*) and whose selectively interchangeable (*paradigmatic*) constraints are exploited by text analysing algorithms. They accept natural language discourse as input and produce a vector space structure as output which may be interpreted as an internal (*endo*) representation of the *SCIP* system's states of adaptation to the external (*exo*) structures of its environment as mediated by the discourse processed. In order to evaluate the sytem's *endo*-representation against the *exo*-view of its environment as described by the natural language discourse processed, a corpus of texts – composed of correct and true sentences with well-defined referential meanings – was generated according to a (very simple) phrase structure grammar and a fuzzy referential semantics which interpret simple composite predicates of *cores* (like: *on the left, in front* etc.) and *hedges* (like: *extremely nearby, very faraway* etc.). Processed during the system's training phase, the corpus reveals structural constraints which the system's hidden structures or internal meaning representations apparently reflect. The system's architecture is a two-level consecutive mapping of distributed representations of systems of (fuzzy) linguistic entities whose states acquire symbolic functions that can be equaled to (basal) referencial predicates. Test results from an experimental setting with varying fuzzy interpretations of hedges are produced to illustrate the *SCIP* system's miniature (cognitive) language understanding and meaning acquisition capacity without any initial explicit syntactic and semantic knowledge.

1 Language and cognition

Perception, identification, and interpretation of (external or internal) structures may be conceived as some form of *information processing* which (natural or artificial) cognitive systems—due to their own structuredness—are able to perform. Under this unifying paradigm for *cognition*, research programs in *cognitive linguistics* and *cognitive language processing* can roughly be characterized to consist of subtle forms in confronting models of *competence theory* of language with observable phenomena of communicative *language performance* to explore the structure of mental activities believed to underlie language learning and understanding by way of modelling these activities procedurally to enable algorithmic implementation and testing by machine simulation.

Whereas traditional approaches in artificial intelligence research (*AI*) or computational linguistics (*CL*) model cognitive tasks or natural language understanding in information processing systems according to the *realistic* view of semantics, it is argued here that *meaning* need not be introduced as a presupposition of *semantics* but may instead be derived as a result of procedural modelling[1] as soon as a *semiotic* line of approaches to cognition will be followed [3].

1.1 Understanding: situations

The present approach is based upon a phenomenological (re-)interpretation of the formal concept of *situation* [1] and the analytical notion of *language game*. The combination of both lends itself easily to operational extensions in empirical analysis and procedural simulation of associative meaning constitution which will grasp essential parts of the process of *understanding*.

According to *Situation Semantics* any language ex-

---

[*] Proceedings of ISUMA-NAFIPS'95 ©1995 IEEE Computer Society Press (ISBN 0-8186-7126-2/95)

[1] Procedural models denote a class of models whose interpretation is not (yet) tied to the semantics provided by an underlying theory of the objects (or its expressions) but consist (sofar) in the procedures and their algorithmic implementations whose instantiations as processes (and their results) by way of computer programs provide the only means for their testing and evaluation. The lack of an abstract (theoretical) level of representation for these processes (and their results) apart from the formal notation of the underlying algorithms is one of the reasons why *fuzzy set* and *possibility* theory [15] [16] and their logical derivates were wellcome to provide an open and new procedural format for computational approaches to natural language semantics without obligation neither to reject nor to accept traditional formal and modeltheoretic concepts.

pression is tied to reality in two ways: by the *discourse situation* allowing an expression's meaning being *interpreted* and by the *described situation* allowing its interpretation being *evaluated* truth-functionally. Within this relational model of semantics, *meaning* may be considered the derivative of information processing which (natural or artificial) systems—due to their own structuredness—perform by recognizing similarities or invariants between situations that structure their surrounding realities (or fragments thereof).

By ascertaining these invariants and by mapping them as *uniformities* across *situations*, cognitive systems properly *attuned* to them are able to identify and understand those bits of information which appear to be essential to form these systems' particular views of reality: a flow of *types of situations* related by *uniformities* like e.g. individuals, relations, and time-space-locations. These uniformities constrain a system's external world to become its *view of reality* as a specific fragment of persistent (and remembered) *courses of events* whose expectability renders them interpretable or even *objective*.

In semiotic sign systems like natural languages, such uniformities appear to be signalled also by *word-types* whose employment as *word-tokens* in texts exhibit a special form of *structurally conditioned* constraints. Not only allows their use the speakers/hearers to convey/understand meanings differently in different discourse situations (*efficiency*), but at the same time the discourses' total vocabulary and word usages also provide an empirically accessible basis for the analysis of *structural* (as opposed to *referencial*) aspects of *event-types* and how these are related by virtue of word uniformities accross phrases, sentences, and texts uttered. Thus, as a means for the *intensional* (as opposed to the *extensional*) description of (abstract, real, and actual) *situations*, the regularities of word-usages may serve as an access to and a representational format for those elastic constraints which underly and condition any word-type's *meaning*, the *interpretations* it allows within possible contexts of use, and the *information* its actual word-token employment on a particular occasion may convey.

1.2 Communicating: language games

The notion of *language games* [14] "complete in themselves, as *complete systems* of human communication" is primarily concerned with the way of how signs are used "simpler than those in which we use the signs of our highly complicated everyday language". Operationalizing this notion and analysing a great number of texts for *usage regularities* of terms can reveal essential parts of the concepts and hence the meanings conveyed by them. This approach [3] has also produced some evidence that an analytical procedure appropriately chosen could well be identified also with solving the representational task if based upon the universal constraints known to be valid for all natural languages.

The philosophical concept of *language game* can be combined with the formal notion of *situations* allowing not only for the identification of an cognitve system's (*internal*) structure with the (*external*) structure of that system's environment. Being tied to the observables of actual language performance enacted by communicative language useage opens up an empirical approach to procedural semantics. Whatever can formally be analysed as *uniformities* in BARWISEian *discourse situations* may eventually be specified by word-type regularities as determined by co-occurring word-tokens in pragmatically homogeneous samples of *language games*. Going back to the fundamentals of structuralistic descriptions of regularities of *syntagmatic* linearity and *paradigmatic* selectivity of language items, the correlational analyses of discourse will allow for a multi-level word meaning and world knowledge representation whose dynamism is a direct function of elastic constraints established and/or modified in language communication.

As has been outlined in some detail elsewhere [4] [6] [8] [12] the meaning function's range may be computed and simulated as a result of exactly those (semiotic) procedures by way of which (representational) structures emerge and their (interpreting) actualisation is produced from observing and analyzing the domain's regular constraints as imposed on the linear ordering (*syntagmatics*) and the selective combination (*paradigmatics*) of natural language items in communicative language performance. For natural language semantics this is tantamount to (re)present a term's meaning potential by a fuzzy *distributional pattern* of the modelled system's state changes rather than a *single symbol* whose structural relations are to represent the system's interpretation of its environment. Whereas the latter has to *exclude*, the former will automatically *include* the (linguistically) structured, pragmatic components which the system will both, embody and employ as its (linguistic) import to identify and to interpret its environmental structures by means of its own structuredness.

2 Knowledge and representation

In knowledge based cognitive linguistics and semantics, researchers get the necessary lexical, semantic, or external world information by exploring (or making test-persons explore) their own linguistic or cognitive capacities and memory structures in order to depict their findings in (or let hypotheses about them be tested on the bases of) traditional forms of knowledge representation. Being based upon this pre-defined and rather static concept of *knowledge*, these representations are confined not only to predicative and propositional expressions which can be mapped in well established (concept-hierarchical, logically deductive) formats, but they will also lack the flexibility and dynamics of *re-constructive* model structures more reminiscent of language understanding and better suited for automatic analysis and representation of meanings from texts. Such devices have been recognized to be essential [13] for any simulative modelling capable to set up and modify a system's own knowledge structure, however shallow and vague its semantic knowledge and inferencing capacity may appear compared to human understanding. The *semiotic* approach argued for here appears to be a feasible alternative [5] focussing on the dynamic structures which the speakers'/hearers' communicative use of language in discourse will both, constitute and modify, and whose reconstruction may provide a paradigm of cognition and a model for the emergence of meaning. In [9] [10] a corresponding meaning representation formalism has been defined and tested whose parameters may automatically be detected from natural language texts and whose non-symbolic and distributional format of a vector space notation allows for a wide range of useful interpretations.

2.1 Quantitative text analysis

Based upon the fundamental distinction of natural language items' agglomerative or *syntagmatic* and selective or *paradigmatic* relatedness, the core of the representational formalism can be characterized as a two-level process of abstraction. The first (called α-abstraction) on the set of *fuzzy* subsets of the vocabulary provides the word-types' usage regularities or *corpus points*, the second (called δ-abstraction) on this set of *fuzzy* subsets of corpus points provides the corresponding *meaning points* as a function of word-types which are being instantiated by word-tokens as employed in *pragmatically homogeneous* corpora of natural language texts.

The basically descriptive statistics used to grasp these relations on the level of *words* in discourse are centred around a correlational measure (*Eqn.* 1) to specify intensities of co-occurring lexical items in texts, and a measure of similarity (or rather, dissimilarity) (*Eqn.* 4) to specify these correlational value distributions' differences. Simultaneously, these measures may also be interpreted semiotically as set theoretical constraints or formal mappings (*Eqns.* 2 and 5) which model the meanings of words as a function of differences of usage regularities.

$\alpha_{i,j}$ allows to express pairwise relatedness of word-types $(x_i, x_j) \in V \times V$ in numerical values ranging from -1 to $+1$ by calculating co-occurring word-token frequencies in the following way

$$\alpha(x_i, x_j) = \frac{\sum_{t=1}^{T}(h_{it} - e_{it})(h_{jt} - e_{jt})}{\left(\sum_{t=1}^{T}(h_{it} - e_{it})^2 \sum_{t=1}^{T}(h_{jt} - e_{jt})^2\right)^{\frac{1}{2}}}; \quad (1)$$

$$-1 \leq \alpha(x_i, x_j) \leq +1$$

where $e_{it} = \frac{H_i}{L} l_t$ and $e_{jt} = \frac{H_j}{L} l_t$, with the textcorpus $K = \{k_t\}; t = 1, \ldots, T$ having an overall length $L = \sum_{t=1}^{T} l_t; 1 \leq l_t \leq L$ measured by the number of word-tokens per text, and a vocabulary $V = \{x_n\}; n = 1, \ldots, i, j, \ldots, N$ whose frequencies are denoted by $H_i = \sum_{t=1}^{T} h_{it}; 0 \leq h_{it} \leq H_i$.
Evidently, pairs of items which frequently either cooccur in, or are both absent from, 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*.
As a fuzzy binary relation, $\tilde{\alpha} : V \times V \to I$ can be conditioned on $x_n \in V$ which yields a crisp mapping

$$\tilde{\alpha} \mid x_n : V \to C; \{y_n\} =: C \quad (2)$$

where the tupels $\langle (x_{n,1}, \tilde{\alpha}(n,1)), \ldots, (x_{n,N}, \tilde{\alpha}(n,N)) \rangle$ represent the numerically specified, *syntagmatic* usage regularities that have been observed for each word-type x_i against all other $x_n \in V$. *α-abstraction* over one of the components in each ordered pair defines

$$x_i(\tilde{\alpha}(i,1), \ldots, \tilde{\alpha}(i,N)) =: y_i \in C \quad (3)$$

Hence, the regularities of usage of any lexical item will be determined by the tupel of its *affinity/repugnancy*-values towards each other item of the vocabulary which—interpreted as coordinates— can be represented by points in a vector space C spanned by the number of axes each of which corresponds to an entry in the vocabulary.

2.2 Distributed meaning representation

Considering C as representational structure of abstract entities constituted by *syntagmatic* regularities of word-token occurrences in *pragmatically homogeneous* discourse, then the similarities and/or dissimilarities of these entities will capture their corresponding word-types' *paradigmatic* regularities. These may be calculated by a distance measure δ of, say, Euclidian metric

$$\delta(y_i, y_j) = \left(\sum_{n=1}^{N}(\alpha(x_i, x_n) - \alpha(x_j, x_n))^2\right)^{\frac{1}{2}}; \quad (4)$$

$$0 \leq \delta(y_i, y_j) \leq 2\sqrt{n}$$

$$\begin{tikzcd}
& C \arrow[loop above, "\tilde{\delta}"] \arrow[dr, "\tilde{\delta}\,|\,y_j"] & \\
V \arrow[loop left, "\tilde{\alpha}"] \arrow[ur, "\tilde{\alpha}\,|\,x_i"] \arrow[rr, "\tilde{\delta}\,|\,y_n \,\circ\, \tilde{\alpha}\,|\,x_n"'] & & S \arrow[loop right, "\zeta"]
\end{tikzcd}$$

Figure 1: Fuzzy mapping relations $\tilde{\alpha}$ and $\tilde{\delta}$ between the structured sets of vocabulary items $x_n \in V$, of corpus points $y_n \in C$, and of meaning points $z_n \in S$.

Thus, δ may serve as a *second* mapping function to represent any item's differences of usage regularities measured against those of all other items. As a fuzzy binary relation, $\tilde{\delta} : C \times C \to I$ can be conditioned on $y_n \in C$ which again yields a crisp mapping

$$\tilde{\delta} \mid y_n : C \to S; \{z_n\} =: S \qquad (5)$$

where the tupels $\langle(y_{n,1}, \tilde{\delta}(n,1)), \ldots, (y_{n,N}\tilde{\delta}(n,N))\rangle$ represents the numerically specified *paradigmatic* structure that has been derived for each abstract *syntagmatic* usage regularity y_j against all other $y_n \in C$. The distance values can therefore be abstracted analogous to *Eqn.* 3, this time, however, over the other of the components in each ordered pair, thus defining an element $z_j \in S$ called *meaning point* by

$$y_j(\tilde{\delta}(j,1), \ldots, \tilde{\delta}(j,N)) =: z_j \in S \qquad (6)$$

Identifying $z_n \in S$ with the numerically specified elements of potential paradigms, the set of possible combinations $S \times S$ may structurally be constrained and evaluated without (direct or indirect) recourse to any pre-existent external world. Introducing a EUCLIDian metric

$$\zeta : S \times S \to I \qquad (7)$$

the hyperstructure $\langle S, \zeta \rangle$ or *semantic hyper space* (*SHS*) is declared constituting the system of *meaning points* as an empirically founded and functionally derived representation of a lexically labelled knowledge structure (*Tab.* 1).

As a result of the two-stage *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) each lexical item shows against all others in the discourse analysed. 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 items in discourse resulting from actual or intended acts of communication (*communicative performance*), text understanding is modelled procedurally the process to construct and identify the topological positions of any meaning point $z_i \in \langle S, \zeta \rangle$ corresponding to the vocabulary items $x_i \in V$ which can formally be stated as composition of the two restricted relations $\tilde{\delta} \mid y$ and $\tilde{\alpha} \mid x$ (*Fig.* 1).

Processing natural language texts the way these algorithms do would appear to grasp some interesting portions of the ability to recognize and represent and to employ and modify the structural information available to and accessible under such performance. A *semiotic cognitive information processing system (SCIPS)* endowed with this ability and able to perform likewise would consequently be said to have constituted some text *understanding*. The problem is, however, whether (and if so, how) the contents of what such a system is said to have acquired can be tested, i.e. made accessible other than by the language texts in question and/or without committing to a presupposed semantics determining possible interpretations.

$$\begin{aligned}
\textit{SCIP-S} :&\ \{\mathcal{O}, \mathcal{B}, \mathcal{W}, \mathcal{F}, \mathcal{K}\}\\
\textit{Orientation} :&\ \mathcal{O} := \{\vec{N} = (0,1), \vec{O} = (1,0),\\
&\qquad \vec{S} = (0,-1), \vec{W} = (-1,0)\}\\
\textit{Mobility} :&\ \mathcal{B} := \{k(0,1), k(1,1), k(1,0), k(1,-1),\\
&\qquad k(0,-1), k(-1,-1), k(-1,0), k(-1,1)\\
&\qquad : k = 1\}\\
\textit{Perception} :&\ \mathcal{W} := \{K := \{k_t\}, L := \textstyle\sum_{t=1}^{T} l_t, V := \{x_i\},\\
&\qquad H_i := \textstyle\sum_{t=1}^{T} h_{it} : i = 1, \ldots, j, \ldots, N\}\\
\textit{Processing} :&\ \mathcal{F} := \{\alpha,\ \delta,\ \zeta,\ \ldots\};\\
&\ \mathcal{K} := \{\tilde{\alpha} \mid x,\ \tilde{\delta} \mid y,\ \ldots\}\\
\textit{Semantics} :&\ \textbf{none}\\
\textit{Syntax} :&\ \textbf{none}
\end{aligned}$$

Table 2: Collection of *SCIP-systemic* properties.

$$\begin{aligned}
\textit{SCIP-E} :&\ \{\mathcal{R}_{\mathrm{E}}, \mathcal{R}_{\mathrm{O}}, \mathcal{R}_{\mathrm{R}}, \mathcal{D}, \ell_{\mathcal{R}}\}\\
\textit{Ref-plane} :&\ \mathcal{R}_{\mathrm{E}} := \{P_{n,m} : \exists R_{n,m} \in \mathcal{R}_{\mathrm{R}}(n_0, m_0, g),\\
&\qquad P_{n,m} \in R_{n,m}\}\\
\textit{Ref-objects} :&\ \mathcal{R}_{\mathrm{O}} := \{\square,\ \triangle,\ \bigcirc,\ \ldots\ \}\\
\textit{Ref-grid} :&\ \mathcal{R}_{\mathrm{R}}(n_0, m_0, g) := \{\mathrm{R}_{n,m} =\\
&\qquad [(n-1)g, ng] \times [(m-1)g, mg]\\
&\qquad 1 \le n \le n_0\,, 1 \le m \le m_0\,, g > 0\}\\
\textit{Directions} :&\ \mathcal{D} := \{\vec{N} := (0,1), \vec{O} := (1,0),\\
&\qquad \vec{S} := (0,-1), \vec{W} := (-1,0)\}\\
\textit{Obj-location} :&\ \ell_{\mathcal{R}}\ :\ \mathcal{R}_{\mathrm{O}} \longrightarrow \mathcal{R}_{\mathrm{E}}
\end{aligned}$$

Table 3: Collection of *SCIP-environmental* properties.

$$\begin{array}{ccccc}
V \times V & \alpha\text{-abstraction} & C \times C & \delta\text{-abstraction} & S \times S \\
\begin{array}{c|ccc} \tilde{\alpha} & x_1 & \dots & x_N \\ \hline x_1 & \alpha_{11} & \dots & \alpha_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & \alpha_{N1} & \dots & \alpha_{NN} \end{array} & \begin{array}{c} \Downarrow \\ \xrightarrow{\tilde{\alpha} \mid x_i} \\ \Uparrow \end{array} & \begin{array}{c|ccc} \tilde{\tilde{\delta}} & y_1 & \dots & y_N \\ \hline y_1 & \delta_{11} & \dots & \delta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ y_N & \delta_{N1} & \dots & \delta_{NN} \end{array} & \begin{array}{c} \Downarrow \\ \xrightarrow{\tilde{\delta} \mid y_j} \\ \Uparrow \end{array} & \begin{array}{c|ccc} \zeta & z_1 & \dots & z_N \\ \hline z_1 & \zeta_{11} & \dots & \zeta_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ z_N & \zeta_{N1} & \dots & \zeta_{NN} \end{array} \; \mathrm{P} \\
 & \textit{Syntagmatic} & \textit{Constraints} & \textit{Paradigmatic} &
\end{array}$$

Table 1: Formalizing (*syntagmatic/paradigmatic*) constraints by consecutive (α- and δ-) abstractions over usage regularities of items x_i, y_j respectively.

Word:	the sign-object identified as vocabulary element (type) whose occurrences in (linear) sets of sign-objects (tokens) are countable
Sentence:	the (non-empty, linear) set of words to form a correct expression of a true proposition denoting a relation of system-position and object-location
Text:	the (non-empty, linear) set of sentences with identical pairs of core-predicates denoting system-object-relations resulting from linear movement and directly adjacent system-positions
Corpus:	the (non-empty) set of texts comprising descriptions of (any or all) factually possible system-object relations within a specified systemic and environmental setting

Table 4: *SCIP*-Restrictions on concepts of language material entities.

3 The experimental setting

To enable an intersubjective scrutiny, the (unknown) results of an abstract system's (well known) acquisition process is compared against the (well known) traditional interpretations of the (unknown) processes of natural language meaning constitution[2]. To achieve this, it had t be guaranteed

- ▷ that the three main components of the experimental setting, the *system*, the *environment*, and the *discourse* are specified by sets of conditioning properties. These define the *SCIP* system by way of a set of procedural entities like *orientation, mobility, perception, processing* (*Tab.* 2), the *SCIP*-environment is defined as a set of formal entities like *plane, objects, grid, direction, location* (*Tab.* 3), and the *SCIP*-discourse material mediating between system and environment is structured first by a number of part-whole related entities like *word, sentence, text, corpus* (*Tab.* 4) of which *sentence* and *text* require further formal restrictions to be specified by a formal *syntax* (*Tab.* 5) and a referential *semantics* (*Tab.* 6).
- ▷ that the system's environmental data consists in a corpus of (natural language) texts of correct expressions of true propositions denoting system-object-relations described according to the formally specified syntax and semantics (representing the *exo*-view or *described situations*), and
- ▷ that the system's internal picture of its surroundigs (representing the *endo*-view or *discourse situations*) is to be derived from this textual language environment other than by way of propositional reconstruction, i.e. without syntactic parsing and semantic interpretation of sentence and text structures.

$$\begin{array}{rl}
\mathrm{T(ext)} := & \{S_i \mid S_i \longrightarrow S_{i+1} : \mathcal{B} \wedge \{KP_1, KP_2\} \in S_i \\
& \wedge \{KP_1, KP_2\} \in S_{i+1} \\
& \wedge \forall\, KP_j \in S_i \\
& \cup S_{i+1}; j = 1,2; \quad i = 1, \dots, I\} \\
\mathcal{B} := & \{k(0,1), k(1,1), k(1,0), k(1,-1), \\
& k(0,-1), k(-1,-1), k(-1,0), \\
& k(-1,1) : k = 1\} \\
\mathrm{S}_i \longrightarrow & \mathrm{NP} \quad \mathrm{VP} \\
\mathrm{NP} \longrightarrow & \mathrm{N} \\
\mathrm{VP} \longrightarrow & \mathrm{V} \quad \mathrm{PP} \\
\mathrm{PP} \longrightarrow & \mathrm{HP} \quad \mathrm{KP} \\
\mathrm{N} \longrightarrow & \textit{A} \ \langle \textit{ triangle} \mid \textit{square} \mid \textit{circle } \rangle \\
\mathrm{V} \longrightarrow & \textit{lies} \\
\mathrm{HP} \longrightarrow & \langle \textit{ extremely} \mid \textit{very} \mid \textit{rather } \rangle \\
& \langle \textit{ near by} \mid \textit{far away } \rangle \\
\mathrm{KP} \longrightarrow & \langle \textit{ on the left} \mid \textit{on the right } \rangle \\
& \mid \langle \textit{ in front} \mid \textit{behind } \rangle
\end{array}$$

Table 5: Syntax of textgrammar for the generation of strings of correct descriptions of possible system-*position* and object-*location* relations.

3.1 Positions and locations

The experimental setting consists of a two dimensional environment with some objects at certain places (*Fig.* 2) that a *SCIP*-system will have to identify on

---

[2] The concept of *knowledge* underlying this use here may be understood to refer to *known* as having well established (scientific, however controversial, but at least inter-subjective) *models* to deal with, whereas *unknown* refers to the lack of such models.

Core-predicates (KP)

in relations of system-positions x, y and object-locations n, m (with 0-coordinates down left) for all orientations N, O, S, W of the system

NORTH x, y	*in front*	*behind*
on the left	>m, <n	>m, >n
on the right	<m, <n	>m, <n

EAST x, y	*in front*	*behind*
on the left	<m, <n	>m, <n
on the right	<m, >n	>m, >n

SOUTH x, y	*in front*	*behind*
on the left	<m, >n	<m, <n
on the right	>m, >n	<m, >n

WEST x, y	*in front*	*behind*
on the left	>m, >n	<m, >n
on the right	>m, <n	<m, <n

Hedge-predicates (HP)

as distances of sytem-position/object-location (*crisp*- and *fuzzy*- interpretation): in numbers of grid-points $\mid x - \mathrm{n} \mid$ and $\mid y - \mathrm{m} \mid$)

Crisp 1.0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
extremely nearby	1	1	0	0	0	0	0	0	0	0
very nearby	0	0	1	1	0	0	0	0	0	0
rather nearby	0	0	0	0	1	0	0	0	0	0
rather faraway	0	0	0	0	0	1	0	0	0	0
very faraway	0	0	0	0	0	0	1	1	0	0
extremely faraway	0	0	0	0	0	0	0	0	1	1

Fuzzy 1.1	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
extremely nearby	1	1	.7	.2	0	0	0	0	0	0
very nearby	.2	.7	1	1	.7	.2	0	0	0	0
rather nearby	0	0	.2	.7	1	.7	.2	0	0	0
rather faraway	0	0	0	.2	.7	1	.7	.2	0	0
very faraway	0	0	0	0	.2	.7	1	1	.7	.2
extremely faraway	0	0	0	0	0	0	.2	.7	1	1

Table 6: Semantics to identify true core- and hedge-predicates (under *crisp* and *fuzzy*) interpretation) in correct sentences being generated for fixed (unchanged) object-*locations* and varying (changed) system-*positions*.

the grounds of natural language descriptions of system-position and object-location relations it is exposed to. Although the system's perception is limited to its (formal) language processing and as its ability to act (and react) is restricted to pacewise linear movement, what makes it *semiotic* is that—whatever the system might gather from its environment—it will not apply any coded knowledge available prior to that process, but will instead only be confined to the system's own (co- and contextually restricted) susceptibility and processing capabilities to (re-)organize the environmental data and to (re-)present the results in some dynamic structure which determines the system's knowledge (susceptibility), learning (change) and understanding (representation). It is based on the assumption that some deeper representational level or core structure might be identified as a common base for different notions of meaning developped sofar in theories of *referential* and *situational* semantics as well as some *structural* or *stereotype* semantics.

For the purpose of testing *semiotic* processes, their situational complexity has to be reduced by abstracting away irrelevant constituents, hopefully without oversimplifying the issue and trivializing the problem. Therefore, the propositional form of natural language predication, will be used here only to control the format of the natural language training material, not, however, to determine the way it is processed to model *understanding*.

3.2 Process and result

The strict separation between the process and its result on the system's side now corresponds to the sharp distinction between the formal specification to control the propositional generation of referentially descriptive language material and its non-propositional processing within the experimental *SCIP* setting.

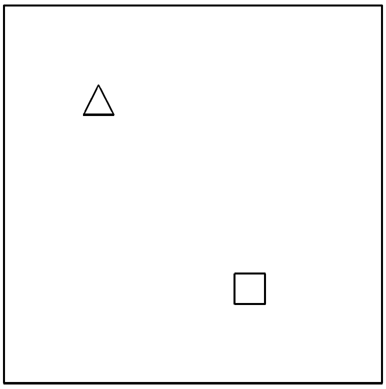

Figure 2: Reference plane with location of objects (△ and □) propositionally described by texts in the training corpus.

Illustrating an example situation, the reference plane (*Fig.* 2) shows two object-*locations*. These have (automatically) been described in a corpus of language

B E H I N D

	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	
	0	0	2	2	2	2	4	4	4	4		1	1	1	1	1	1	0	0	0	0	
	0	0	2	2	2	2	4	4	4	4		1	1	1	1	1	1	0	0	0	0	
	0	0	2	2	2	2	4	4	4	4		1	1	1	1	1	1	0	0	0	0	
	0	0	5	5	5	5	2	2	2	2		1	1	1	1	1	1	0	0	0	0	
	0	0	5	5	5	5	2	2	2	2		1	1	1	1	1	1	0	0	0	0	
	0	0	5	5	5	5	2	2	2	2		1	1	1	1	1	1	0	0	0	0	
R	0	0	4	4	4	4	2	2	2	2		1	1	1	1	1	1	0	0	0	0	L
I	0	0	4	4	4	4	2	2	2	2		1	1	1	1	1	1	0	0	0	0	
G											▽											E
H	0	0	1	1	1	1	1	1	1	1		6	6	3	3	3	3	0	0	0	0	F
T	0	0	1	1	1	1	1	1	1	1		6	6	3	3	3	3	0	0	0	0	T
	0	0	1	1	1	1	1	1	1	1		3	3	7	7	5	5	0	0	0	0	
	0	0	1	1	1	1	1	1	1	1		3	3	7	7	5	5	0	0	0	0	
	0	0	1	1	1	1	1	1	1	1		3	3	4	4	4	4	0	0	0	0	
	0	0	1	1	1	1	1	1	1	1		3	3	4	4	4	4	0	0	0	0	
	0	0	1	1	1	1	1	1	1	1		3	3	4	4	4	4	0	0	0	0	
	0	0	1	1	1	1	1	1	1	1		3	3	4	4	4	4	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	

I N F R O N T

Table 7: *Endo1*$_{i,j}$ showing regional object locations from system position (oriented south ▽) by sums of grid point marks received according to pairs of (*crisp*ly interpreted) hedged core predicate adjacencies.

N O R T H

	226	240	251	232	213	194	164	141	118	95	
	240	260	274	257	240	223	192	168	144	120	
	251	274	295	284	271	258	226	201	176	151	
W	237	262	289	285	277	269	238	216	194	172	*E*
E	223	250	280	280	276	272	242	223	204	185	*A*
S	209	238	271	275	275	275	246	230	214	198	*S*
T	191	222	258	269	276	283	258	243	228	213	*T*
	173	206	245	263	277	291	270	256	242	228	
	144	176	214	236	254	272	256	244	232	220	
	119	150	187	212	233	254	242	232	222	212	

S O U T H

Table 8: *Endo2*$_{m,n}$ showing regions of object location likelyhood computed for each gridpoint *m,n* by superimposing locality patterns from *Endo1*$_{i,j}$ value.

expressions comprising some 12 432 word tokens of 26 word types in 2 483 sentences and 684 texts generated according to the formal syntax and semantics specified for all possible system-*positions* and *orientations*. The training set of language material was then exposed to the *SCIP* system which perceived it as environmental data to be processed according to its system faculties as specified. It is worthwhile noting here again, that this processing is neither based on, nor does it involve any knowledge of *syntax* or *semantics* on the system's side.

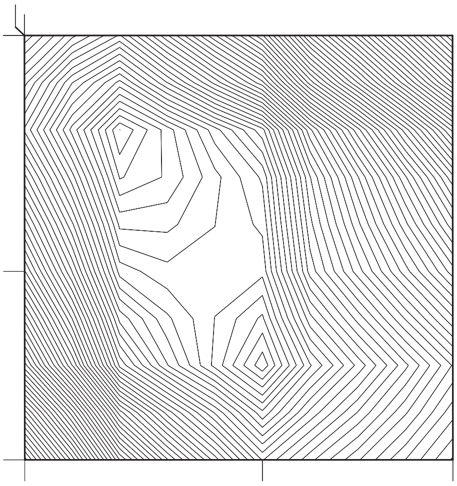

Figure 3: External *2-dim*-image of the *SCIP* system's *endo*-view showing regions of potential object locations under *crisp* hedge interpretation.

In the course of processing, the two-level consecutive mappings (*Tab.* 1, *Fig.* 1) result in the *semantic hyper space (SHS)* whose intrinsic structure reveal some properties which can be made visible in a three stage process:

- ▷ first, applying methods of KOHONEN-maps [2] or—with comparable results—average linkage cluster analysis [7] allows to identify structurally adjacent word-types (like object label and predicate label candidates) [11],
- ▷ second, their *numerical* hedge interpretation yields the distance values, and their *directional* core interpretations determines the regions of object locations relative to a centrally positioned system (*Tab.* 7), producing an intermediate representation of the system's own oriented view which can be transformed to
- ▷ third, a mapping that images an orientation indepedent representation of the system's *endo*-view of its environment (*Tab.* 8). It can be visualized in another format as
- ▷ fourth, a holistic representation of the referencial plane structured by a pattern of polygons which connect regions of denotational likelihood or *isoreferentials* (*Fig.* 3).

The *Endo1*$_{i,j}$ data (*Tab.* 7) serves as base for the following third step of a line- and column-wise transform which results in a new mapping *Endo2*$_{m,n}$ (*Tab.* 8) according to the summation equation

$$Endo2_{m,n} = \sum_{i=m}^{m+10} \sum_{j=n}^{n+10} Endo1_{i,j} \qquad (8)$$

The matrix $Endo2_{m,n}$ (*Tab.* 8) contains the data for an external *observer's* image of the system's *endo*-view as computed from the described object locations relative to system positions. The (two-dimensional) *2-dim*-scattergram of *Endo2* (*Fig.* 3) gives an overall picture of even referential likelihood by *isoreferentials* denoting potential object locations quite clearly, however *fuzzy*.

References

[1] J. Barwise and J. Perry. *Situations and Attitudes.* (MIT Press), Cambridge, MA, 1983.

[2] T. Kohonen. *Self-Organization and Associative Memory.* Springer Series in Information Sciences. Springer, Berlin/ Heidelberg/ New York/ London, 3rd edition, 1989.

[3] B. Rieger. *Unscharfe Semantik. Die empirische Analyse, quantitative Beschreibung, formale Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten.* Peter Lang, Frankfurt a.Main/ Bern/ Paris, 1989.

[4] B. B. Rieger. 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. In M. Nagao and K. Fuchi, editors, *COLING-80 8th International Conference on Computational Linguistics*, pages 76–84, Tokyo, 1980. International Committee on Computational Linguistics, ICCL.

[5] B. B. Rieger. Feasible Fuzzy Semantics. On some problems of how to handle word meaning empirically. In H. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts. New Approaches in Word Semantics*, Research in Text Theory 6, pages 193–209. de Gruyter, Berlin/ NewYork, 1981.

[6] B. B. Rieger. Fuzzy Representation Systems in Linguistic Semantics. In R. Trappl, N. Findler, and W. Horn, editors, *Progress in Cybernetics and Systems Research*, Vol. XI, pages 249–256. McGraw-Hill Intern., Washington/ NewYork/ London, 1982.

[7] B. B. Rieger. Clusters in Semantic Space. In L. Delatte, editor, *Actes du Congrès International Informatique et Science Humaines*, pages 805–814, Liège, 1983. Laboratoire d'Analyse Statistique des Langues Anciennes, LASLA.

[8] B. B. Rieger. Semantic Relevance and Aspect Dependancy in a Given Subject Domain. In D. Walker, editor, *COLING-84 10th International Conference on Computational Linguistics*, pages 298–301, Stanford, 1984. International Committee of Computational Linguistics, ACL-ICCL.

[9] B. B. Rieger. Lexical Relevance and Semantic Disposition. On stereotype word meaning representation in procedural semantics. In G. Hoppenbrouwes, P. Seuren, and T. Weijters, editors, *Meaning and the Lexicon*, pages 387–400. Foris Publications, Dordrecht, 1985.

[10] B. B. Rieger. Stereotype representation and dynamic structuring of fuzzy word meanings for contents-driven semantic processing. In J. Agrawal and P. Zunde, editors, *Empirical Foundations of Information and Software Sience*, pages 273–291. Plenum Press, NewYork/ London, 1985.

[11] B. B. Rieger. Situation semantics and computational linguistics: towards informational ecology. a semiotic perspective for cognitive information processing systems. In K. Kornwachs and K. Jacoby, editors, *Information. New Questions to a Multidisciplinary Concept*, pages 285–315, Berlin, 1995. Akademie.

[12] B. B. Rieger and C. Thiopoulos. Semiotic Dynamics: a self-organizing lexical system in hypertext. In R. Köhler and B. Rieger, editors, *Contributions to Quantitative Linguistics. Proceedings of the 1st Quantitative Linguistics Conference – QUALICO-91*, pages 67–78. Kluwer Academic Publishers, Dordrecht, 1993.

[13] T. Winograd. *Language as a Cognitive Process. Vol. I: Syntax.* Addison-Wesley, Reading, MA, 1983.

[14] L. Wittgenstein. *The Blue and Brown Books. (Ed. by R.Rhees).* Blackwell, Oxford, 1958.

[15] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning I, II & III. *Information Science*, (8 & 9):338–353; 301–357 & 43–80, 1975.

[16] L. Zadeh. Test-Score Semantics for Natural Languages and Meaning Representation via PRUF. In B. Rieger, editor, *Empirical Semantics I. A Collection of New Approaches in the Field*, Quantitative Linguistics 12, pages 281–349. Brockmeyer, Bochum, 1981.